ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE ENGINEERING AND TECHNOLOGY

EXPLORING THE CONFORMATIONAL TRANSITION BETWEEN CLOSED AND OPEN STATES OF THE SARS-CoV-2 SPIKE GLYCOPROTEIN USING MOLECULAR DYNAMICS SIMULATIONS

M.Sc. THESIS

Ceren KILINÇ

Department of Molecular Biology-Genetics and Biotechnology

Molecular Biology-Genetics and Biotechnology Programme

JULY 2020

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE ENGINEERING AND TECHNOLOGY

EXPLORING THE CONFORMATIONAL TRANSITION BETWEEN CLOSED AND OPEN STATES OF THE SARS-CoV-2 SPIKE GLYCOPROTEIN USING MOLECULAR DYNAMICS SIMULATIONS

M.Sc. THESIS

Ceren KILINÇ
(521181104)

Department of Molecular Biology-Genetics and Biotechnology

Molecular Biology-Genetics and Biotechnology Programme

Thesis Advisor: Assist. Prof. Dr. Mert GÜR

JULY 2020

ISTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

SARS-CoV-2 SPİKE GLİKOPROTEİNİNİN KAPALI VE AÇIK HALLERİ ARASINDAKİ KONFORMASYONEL GEÇİŞİN MOLEKÜLER DİNAMİK SİMÜLASYONLARI KULLANILARAK ARAŞTIRILMASI

YÜKSEK LİSANS TEZİ

Ceren KILINÇ
(521181104)

Moleküler Biyoloji-Genetik ve Biyoteknoloji AnaBilim Dalı

Moleküler Biyoloji-Genetik ve Biyoteknoloji AnaBilim Dalı Programı

Tez Danışmanı: Dr. Öğr. Üyesi Mert GÜR

TEMMUZ 2020

Ceren KILINÇ, a M.Sc. student of ITU Graduate School of Science Engineering and Technology student ID 521181104, successfully defended the thesis/dissertation entitled “EXPLORING THE CONFORMATIONAL TRANSITION BETWEEN CLOSED AND OPEN STATES OF THE SARS-CoV-2 SPIKE GLYCOPROTEIN USING MOLECULAR DYNAMICS SIMULATIONS”, which she prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor : **Assist. Prof. Dr. Mert GUR** ....
Istanbul Technical University

Jury Members : **Prof. Dr. Gizem DINLER DOGANAY** ....
Istanbul Technical University

Prof. Dr. Serdar DURDAGI ....
Bahçeşehir University

Date of Submission : 15 June 2020
Date of Defense : 24 July 2020

^.^

FOREWORD

I would like to extend my sincere gratitude to my advisor Assist. Prof. Mert Gür for his advice and guidance throughout my master's programme and contirbution to my academic aspect. I am also very thankful to my ITU Biomolecular Engineering Laboratory colleagues Elhan Taka, Sema Zeynep Yılmaz, Mert Gölcük and Umut Aktaş for their exquisite skill and help.

I would like to thank TÜBİTAK (The Scientific and Technical Research Council of Turkey) for their financial support within Trainee Researcher (TÜBİTAK-STAR) Scholarship Program. I would also gratefully acknowledge the support of the National Center of High Performance Computing (UHeM) at Istanbul Technical University, the COVID-19 High-Performance Computing (HPC) Consortium (Grant number: TG-MCB200070), and Extreme Science and Engineering Discovery Environment (XSEDE).

Above all, my sincere thanks to my parents, sister and friends for being supportive and understanding all the time.

June 2020 Ceren KILINÇ

TABLE OF CONTENTS

Page

FOREWORD .... **ix**
TABLE OF CONTENTS .... **xi**
ABBREVIATIONS .... **xiii**
SYMBOLS .... **xv**
LIST OF TABLES .... **xvii**
LIST OF FIGURES .... **xix**
SUMMARY .... **xxiii**
ÖZET .... **xxvii**
1. INTRODUCTION **1**
  1.1 Purpose of Thesis 1
  1.2 Coronaviruses 2
  1.3 Severe Acute Respiratory Syndrome Coronavirus 2 4
  1.4 SARS-CoV-2 Interaction With ACE2 8
  1.5 Hypothesis 10
2. METHODS **13**
  2.1 Theory of Molecular Dynamics 13
  2.2 Molecular Dynamics System Preparation 17
  2.3 Molecular Dynamics Simulations 18
  2.4 Steered Molecular Dynamics Simulations 19
  2.5 Root Mean Square Deviation 21
  2.6 Root Mean Square Fluctuation 22
  2.7 Principal Component Analysis 22
3. RESULTS **25**
  3.1 Characteristics of the SARS-CoV-2 S Protein In The Closed and Open State 25
    3.1.1 Mobility and interdomain interactions of the S protein 25
    3.1.2 Solvent accessibility of the binding surface of RBD 30
    3.1.3 Energy landscape based on MD simulations of S protein 30
    3.1.4 Transition mechanism between down and up states of S protein protomers 33
6. CONCLUSIONS AND RECOMMENDATIONS **35**
REFERENCES **37**
CURRICULUM VITAE **45**

YK

ABBREVIATIONS

ACE2 : Angiotensin-converting enzyme 2
CLD : C-terminal collectrin-like domain
CD : Connector domain
COVID-19 : Coronavirus disease 2019
CoVs : Coronaviruses
Cryo-EM : Cryogenic electron microscopy
CT : Cytoplasmic tail
Cα : Alpha carbon
E : Envelope
FP : Fusion peptide
H-bond : Hydrogen bond
HR1 : Heptad repeat 1
HR2 : Heptad repeat 2
M : Membrane
MD : Molecular dynamics
MERS-CoV : Middle East respiratory syndrome coronavirus
N : Nucleocapsid
NAMD : Nanoscale molecular dynamics
NMR : Nuclear magnetic resonance
NTD : N-terminal domain
ORF : Open reading frame
PBC : Periodic boundary conditions
PC : Principal component
PCA : Principal component analysis
PD : Peptidase domain
PDB : Protein Data Bank
PME : Particle mesh Ewald
RBD : Receptor binding domain
RBM : Receptor binding motif
RMSD : Root mean square deviation

RMSF : Root mean square fluctuation
RNA : Ribonucleic acid
S : Spike
SARS-CoV : Severe acute respiratory syndrome coronavirus
SARS-CoV-2 : Severe acute respiratory syndrome coronavirus 2
SMD : Steered molecular dynamics
TM : Transmembrane domain
VMD : Visual Molecular Dynamics

SYMBOLS

Å : Ångström
α : Alpha
a : Acceleration
C : Covariance matrix
E : Energy
F : Force
fs : Femtosecond
kcal : Kilocalorie
k : Spring constant
K : Kelvin
m : Mass
mM : Milimolar
N : Number of particles
n : Direction of pulling
ns : Nanosecond
N : Total number of atoms
P : Pressure
p : Principal component
ps : Picosecond
V : Potential energy
V : Volume
R : Coordinate
T : Temperature
t : Time
U : Guiding potential energy
v : Pulling velocity
ξ : Reaction coordinate
σ : Variance
µ : Cehmical potential

LIST OF TABLES

Page

Table 2.1 : Starting conformations and lengths of the performed MD simulations of SARS-CoV-2 S protein. **21**

YK

LIST OF FIGURES

Page

Figure 1.1: (A) The taxonomy tree of the order *Nidovirales*. (B) Schematic representation of the structure of a coronaviruses. **3**

Figure 1.2 : (A) Crystal structure of the SARS-CoV-2 S protein with its one RBD in up position. S protein sequence was taken from NCBI having RefSeq: YP_009724390 (O'Leary et al., 2016). Full length S protein structure was modeled by using the S protein crystal structure having PDB ID 6VYB and comprising %77 of the sequence as a template. Missing regions located in the residue stretch A27-P1140 were modeled using homology modeling web server SWISS-MODEL (Waterhouse et al., 2018) and missing structural models for residue stretches M1-Q26 and L1141-T1273 were modeled based on the fully-glycosylated full-length SARS-CoV-2 S protein model of Woo et. al. (Woo, H. et al., 2020) and Casalino et al. (Casalino et al., 2020), respectively. Protomers of the S protein are shown in straw yellow, silver, and orange. The coordinates and structures of the glycans were derived from the study of Casalino et al. (Casalino et al., 2020). Glycans are colored using the Symbol Nomenclature for Glycans (SNFG) (Nomenclature). S protein structure was manually inserted into the membrane by its TM domain using VMD (Humphrey et al., 1996). Location of TM domain selected for S protein based on UniProt ID P0DTC2 (Consortium, 2019). Phospholipid head groups and tails of the viral membrane are shown in gray color. (B) S protein protomer domain structures in its up conformation are shown. Domains of the S protomer; SP, NTD, RBD, FP, HR1, CD, HR2, TM, and CP are shown in purple, orange, dark blue, pink, mustard, green, red, cyan and brown, respectively. (C) Representative scheme of the functional domains and cleavage sites of the SARS-CoV-2 S protein..... **6**

Figure 1.3 : Down and up positions of RBD of the SARS-CoV-2 S protein in the closed and open states. (A) Side view of the trimeric S protein conformation is shown. Closed and open state conformations sampled from MD simulations structures are aligned based on their secondary structures in the S2 subunit and only the RBD is shown for the open state. RBD of closed and open states are shown in cyan and dark blue colors, respectively. Protomers A, B, and C are shown in straw yellow, silver, and orange. S protein trimer is shown from the top view for the (A) closed and (B) open states. **7**

Figure 1.4 : Schematic illustration of the binding and fusion mechanism of SARS-CoV-2 S protein. The virus membrane and host cell membrane are shown in light and dark gray, respectively. **8**

Figure 1.5 : SARS-CoV-2 attachment to the host cell by virtue of S protein RBD binding to ACE2 PD. Trimer S protein and ACE2 dimer (PDB ID: 6M17) are shown embedded in the viral and host cell membranes, respectively. S protein and ACE2 structures were manually inserted into the membrane by its transmembrane domains using VMD (Humphrey et al., 1996). Location of transmembrane domains selected for S protein and ACE2 based on UniProt IDs P0DTC2 and Q9BYF1, respectively (Consortium, 2019).S protein protomers are shown in straw yellow, silver, and orange. ACE2 protomers are shown in bright lilac and purple. Phospholipid head groups and tails are shown in grey color. Glycans are not shown.. **10**

Figure 2.1 : Schematic representation of PBCs. The molecules in the unit cell (light green) of the infinite system are shown with bright orange and blue in the middle of the scheme while the rest of the representative infinite system is shown with faded colors.. .... **18**

Figure 2.2 : The prepared systems for (A) closed and (B) open states of S protein in the presence of water and ions. In the figure, protomers, ions, and glycans and waters are shown in cartoon, van der Waals, and quick surface representation, respectively. Protomer A, protomer B, protomer C, glycans, sodium, and chloride ions are shown in straw yellow, silver, orange, purple, tan, and pink, respectively. RBD conformations in closed and open states shown in cyan and dark blue, respectively, to show the positions of the RBDs in different states. **18**

Figure 2.3 : SMD pulling vector determined for SARS-CoV-2 S protein SMD simulations. The pulling direction is selected by (A) a vector pointing from the center of mass of SMD atoms on the RBD in down position pointing the center of mass of SMD atoms on the RBD in up position and (B) vice versa. The pulling direction is shown with a magenta and green arrow. C_α atoms of SMD atoms on RBD and fixed atoms on neighboring protomers are shown with yellow and red beads, respectively.. .... **20**

Figure 3.1 : Residue fluctuations of SARS-CoV-2 RBD residues in the down (cyan) and up (blue) position in the closed and open states, respectively. RMSFs of RBD in both conformations were calculated using coordinates of the RBD obtained from the MD simulations. All protomer conformations were aligned using the C_α atoms of the S2 domain secondary structures (helices and beta sheets) of the up protomer for the open state crystal structure. RMSF values are shown for the structured regions only. Schematic representations of the secondary structures are indicated below with bars for α helices and arrows for β sheets. **26**

Figure 3.2 : Distance distribution between RBD and S2 subunit observed in MD simulations shown for (A) the closed state and (B) open state. The interdomain distance was calculated by evaluating the average of the distance between the Cα atoms of residue pairs K378(RBD)-E988(S2) and K386(RBD)-D985(S2). Distances calculated from the down (PDB ID: 6VXX) and up (PBD ID: 6VYB) crystal structures are marked with black and magenta stars. **26**

Figure 3.3 : Observed salt bridges between RBD and neighboring protomers for RBD in a down position in the closed state in MD simulations. **27**

Figure 3.4 : Interactions between RBD in down conformation and the remaining parts of the SARS-CoV-2 S protein. Distance distribution of interdomain salt bridge (1-6) and hydrogen bond (7-12) forming residues of the RBD during MD simulations were shown. The frequency of salt bridge and hydrogen bonds as observed in MD simulations are indicated in each subplot. If a residue pair contains more than one hydrogen bond, then the observation frequency of each is given separately. Dashed lines represent the cutoff values used to determine the existence of a salt bridge or hydrogen bond. **28**

Figure 3.5 : Interactions between RBD in up conformation and the remaining S protein of the SARS-CoV-2 S protein. Two interaction networks were sampled in MD simulations. Observed salt bridges and hydrogen bonds for first (1-5) and second (6-9) interaction networks are shown for conformations of the RBD in up position sampled through MD simulations. Distance distributions between the salt bridge (blue) and hydrogen bond (green) forming residues of the RBD during MD simulations were shown. The frequency of salt bridge and hydrogen bonds as observed in MD simulations are indicated in each subplot. **29**

Figure 3.6 : Solvent and ACE2 accessibility of S protein RBD binding surface. (A) Close contact water molecules at the ACE2 binding surface of RBD of protomer B are shown for the closed state (upper panel) and open state (lower panel) in ice-blue colored licorice representation. (B) ACE2 binding pose is shown on RBD in the down and up conformation of protomer B. The steric clash zone is highlighted with circles. The amino acids involved in RBD-ACE2 interactions on RBD are shown with red beads. **31**

Figure 3.7 : The first two PC obtained from the MD simulations. PC1 and PC2 are shown with red and yellow arrows, respectively.................................. **31**

Figure 3.8 : PC1, PC2, and the free energy landscapes derived from MD simulations. (A) Energy landscape obtained from the closed and open state MD simulations trajectories for S protein protomers. Crystal structures of the closed (PDB ID:6VXX) and open (PDB ID:6VYB) are shown with magenta and gray squares. The initial conformers obtained from down to up and up to down SMD simulations indicated with gray and magenta dots, respectively. (B) Energy landscape obtained using MD and SMD simulations trajectories. The minimum free energy pathway is shown with dots. Magenta, orange, and gray dots correspond to down, semi-open, and up conformations. (C) The first two PC of the MD simulations superimposed on protomer B structure. (D) Down, semi-open, up, and extended S protein protomers are shown. NTD of protomer is not shown. **33**

Figure 3.9 : Time evolution of salt bridges and hydrogen bonds during the transition between the (A) down and (B) up conformations. Each data point shows the average distance... **34**

Figure 3.10 : Superposition of the ACE2 bound RBD structure onto an intermediate state conformation sampled from MD simulations. RBD amino acids involved in ACE2 interaction are shown with red beads.. .... **36**

EXPLORING THE CONFORMATIONAL TRANSITION BETWEEN CLOSED AND OPEN STATES OF THE SARS-CoV-2 SPIKE GLYCOPROTEIN USING MOLECULAR DYNAMICS SIMULATIONS

SUMMARY

Coronaviruses (CoVs) are classified as a genus under the *Coronaviridae* family within the order *Nidovirales* which includes *Mesoniviridae*, *Arteriviridae*, *Roniviridae*, and *Coronaviridae* families. The *Coronaviridae* family includes the *Coronavirinae* subfamily, which is subdivided into four genera; alphacoronaviruses, betacoronaviruses, gammacoronaviruses, and deltacoronaviruses. Coronaviruses are responsible for a wide variety of diseases in humans and animals; alpha- and betacoronaviruses infect mammals, gammacoronaviruses infect mostly birds and few mammals and deltacoronaviruses infect both mammals and birds. Although coronaviruses have been associated with many diseases, historically, coronaviruses had been associated with 15-30% of self-limiting respiratory infections each year in humans. This situation was accepted as such until a member of betacoronavirus named the severe acute respiratory syndrome coronavirus (SARS-CoV) epidemic occurred in 2002. Subsequently, another outbreak caused by another virus belonging to the betacoronavirus genus named the Middle East respiratory syndrome (MERS-CoV) following the SARS-CoV outbreak occurred between 2002-2003 and caused the coronaviruses to be seen as possible pandemic agents. Coronaviruses are able to adapt to new conditions through recombinations and mutations uncomplicatedly, thus, they can alter their host targets efficiently. As a result of these adaptation abilities, a new type of betacoronavirus, which can spread much faster and bind better than SARS-CoV and MERS-CoV, was detected in 2019.

A novel coronavirus called severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) that causes the infectious coronavirus disease 2019 (COVID-19) has resulted in a pandemic crisis since its first registration in December 2019. This virus belongs to the same large betacoronavirus family including SARS-CoV and MERS-CoV that caused other epidemic diseases in the past years. Similar to other coronaviruses, SARS-CoV-2 consists of a viral envelope that has a bilayer lipid structure and three structural proteins embedded in this viral envelope: envelope, membrane, and spike (S). Among these structural proteins, S proteins have a critical role on host cell infections since they are involved in the recognition of the host cells and fusion between viral and host cell membranes. S proteins are large trimeric glycoproteins and contain two functional subunits: the S1 subunit responsible for binding to host cell receptors and the S2 subunit responsible for fusion of the host cell and viral membrane. The S1 subunit contains the N-terminal domain (NTD) and the receptor-binding domain (RBD) which can bind directly to the receptor of the host cell. RBD undergoes rigid-body motion to hide or expose the receptor-binding surface. Depending on this movement, S protein can be in receptor inaccessible or receptor accessible states. In the receptor inaccessible state (closed state), all RBDs are in the down position whereas, in the receptor accessible state (open state), at least one RBD is in the up position to engage with the host cell receptor. In order to initiate the binding and fusion mechanism, the S protein switches from the closed state to the open state and bind to the host cell receptor. SARS-CoV-2 S proteins target the angiotensin-converting enzyme 2 (ACE2) receptors located on the membranes of human respiratory epithelial cells and S proteins can bind to ACE2 receptors due to the RBD in their structure. As a result of receptor binding, a series of structural changes occur in the S protein required for fusion mechanism. There are two main cleavage sites in the structure of S protein that are cleaved by the host cell enzymes: the S1/S2 and S2’ sites. Upon cleavage of the S1/S2 site, S1 and S2 subunits remain non-covalently bound with each other in the pre-fusion conformation. Cleavage of the S2’ site is a prerequisite for the separation of S1 subunit and viral anchored S2 subunit, and for the fusion of the viral and host cell membranes since fusion peptide (FP) at S2 subunit is exposed to the solvent environment after cleavage on this site. Thus, conformational changes that are essential for the fusion mechanism occur in critical fusion mechanism structures heptad repeat 1 (HR1) and heptad repeat 2 (HR2) structures. Conformational change of HR1 to an extended alpha-helix promotes FP insertion in the host cell membrane. This interaction triggers the rearrangement of HR2 to fold over HR1 and form a six-helix bundle. Consequently, viral and host cell membranes are pulled into proximity, allowing the fusion of these membranes and the release of the viral genome into the host cell. This pre-fusion to the post-fusion transition of the S protein can be blocked at any point during the process to prevent the release of the genome into the host cell. Since the recognition of the host cell and release of the viral genome into the host cell are the most crucial steps for pathogenesis and viral infectivity, exploration of the binding and fusion mechanism of S proteins as a potential therapeutic target for developing antiviral drugs has become prominent. In the literature, the binding and fusion process using the complete S protein structures has not been modeled at an all-atom level using molecular dynamics (MD) simulations. MD simulations provide effective insights into structural, dynamic, and energetic information at the atomic level which are difficult to access by experimental techniques. The main aim of the thesis is the modeling of the transition between closed and open states of the SARS-CoV-2 S protein RBD by performing MD simulations.

In this thesis, the switching mechanism of the RBD from its closed state to open state is modeled and analyzed using MD simulations and statistical thermodynamic methods. Initial conformations that are used in MD simulations were obtained from crystal structures having PDB ID 6VXX and 6VYB where the RBD of the SARS-CoV-2 S protein in closed and open states, respectively. These structures cover 76.4% of the protein sequence. The remaining parts of the protein sequence were completed using the homology modeling method. Glycan molecules attached to S protein in the crystal structures were conserved throughout the MD simulations. For both open and closed state structures, MD simulations were performed. Based on the results, S protein in the open state was found to be more mobile than the closed state. Salt bridge and hydrogen bond analyses between RBD structures in up and down positions showed that there were a different number of interactions between two positions; and the difference of these interactions between up and down positions might be the reason for the mobility difference between them. In addition, a steric clash between S protein and ACE2 was found in the closed state. This steric clash between ACE2 and the closed state S protein prevents ACE2 binding in the closed state. Therefore, an inhibitor that binds to the S protein in the closed state does not need to compete for the binding interface with the ACE2 receptor. This finding indicates that inhibitory molecules targeting the S protein in the closed state might be

developed to prevent RBD-ACE2 binding. *In silico* pulling experiments were performed to obtain the transition between the two states using steered MD (SMD) simulations. SMD simulations were initiated from closed and open state structures that are obtained with MD simulations and performed for two directions by pulling the RBD structure from down to up conformation to switch between closed to open state and up to down conformation to switch between open to the closed state. In this way, the switching path between closed and open states is obtained. Results of MD and SMD simulations were used for principal component analysis (PCA) to determine the most important and dominant movements taking part in the transition. PCA is an effective and proven method used to determine the most prominent movements of a protein. The covariance matrix of alpha carbon atom coordinates in the protein structure was constructed using the MD simulations data. The diagonal elements of this matrix give the variance value for each amino acid. Based on the simulation results, the first two principal components (PCs) correspond to 96.1% of the total variance, thereby identifying the two most prominent PCs of the protein. Using the determined PCs, the free energy surface was created, and based on the generated energy landscapes, the minimum free energy pathway representing the transition between the down and up state of the S protein protomers was constructed. Energy landscapes suggest the existence of a semi-open state between down and up states of the S protein protomers and several additional substates at various locations were determined. While RBD is in the semi-open state, RBD of one protomer is halfway between its down and up positions while the RBDs of the remaining two protomers are in the down position. The semi-open state shows a different network of interactions than the down and up states and does not show any steric clash to binding with the ACE2 receptor. These findings show the possibility of RBD-ACE2 binding while the S protein in the semi-open state.

In the thesis, investigations have been made to discover structural characteristics and transition pathway of the SARS-CoV-2 S protein RBD. This thesis provides an extensive insight into the interdomain interactions, dynamics, and solvent accessibility of the SARS-CoV-2 S protein RBD in its open and closed states and its transition pathway between the closed and open states.

SARS-CoV-2 SPİKE GLİKOPROTEİNİNİN KAPALI VE AÇIK HALLERİ ARASINDAKİ KONFORMASYONEL GEÇİŞİN MOLEKÜLER DİNAMİK SİMÜLASYONLARI KULLANILARAK ARAŞTIRILMASI

ÖZET

Koronavirüsler; *Mesoniviridae, Arteriviridae, Roniviridae* ve *Coronaviridae* ailelerini içeren *Nidovirales* takımı içinde *Coronaviridae* ailesi altında bulunan cinsler olarak sınıflandırılmaktadır. *Coronaviridae* ailesi içerisinde yer alan *Coronavirinae* alt ailesi ise kendi içerisinde dört türe ayrılmaktadır; alfakoronavirüsler, betakoronavirüsler, gammakoronavirüsler ve deltakoronavirüsler. Koronavirüsler, insanlarda ve hayvanlarda görülen birçok çeşitli hastalıktan sorumludurlar; alfa- ve betakoronavirüsler memelileri, gamakoronavirüsler çoğunlukla kuşları ve az sayıda memelileri, deltakoronavirüsler ise hem memelileri hem de kuşları enfekte etmektedir. Koronavirüsler birçok hastalıkla ilişkilendirilmelerine rağmen, tarihsel olarak, insanlarda görülen kendini sınırlayan solunum yolu enfeksiyonlarının %15-30'u ile ilişkilendirilmiştir. Bu durum 2002 yılında ortaya çıkan ve bir betakoronavirüs üyesi olan şiddetli akut solunum yolu sendromu koronavirüsü (SARS-CoV) salgınına kadar böyle kabul edilmekteydi. 2002-2003 yılları arasında görülen koronavirüs salgınını, Orta Doğu solunum yolu sendromu koronavirüsü (MERS-CoV) olarak adlandırılan başka bir betakoronavirüs ailesine ait virüsün neden olduğu diğer bir salgının izlemesi ile birlikte, koronavirüslerin insanları enfekte edebilen ve pandemiye neden olabilen olası virüsler olarak görülmesine neden olmuştur. Koronavirüsler sahip oldukları genetik çeşitlilik sayesinde ve rekombinasyon ve mutasyon ile yeni koşullara adapte olabilmektedirler ve böylece, hedef konakçı hücre çeşitliliklerini arttırabilmektedirler. Bu adaptasyon yeteneklerinin bir sonucu olarak, SARS-CoV ve MERS-CoV'dan çok daha hızlı yayılabilen ve iyi bağlanabilen yeni bir betakoronavirüs çeşidi daha 2019 yılında tespit edilmiştir.

Koronavirüs hastalığı 2019'a (COVID-19) neden olan şiddetli akut solunum yolu sendromu koronavirüs 2 (SARS-CoV-2) adlı yeni bir koronavirüs, 2019 yılının aralık ayında ilk olarak kayıtlara geçtiğinden bu yana bir pandemi olarak sonuçlanmıştır. Bu virüs, geçtiğimiz yıllarda başka salgın hastalıklara neden olan SARS-CoV ve MERS-CoV gibi virüslerin de ait olduğu, insanları enfekte edebilen geniş bir koronavirüs ailesine mensuptur. Diğer koronavirüslerde olduğu gibi, SARS-CoV-2 iki katlı lipit tabaka yapısındaki viral zarftan ve bu viral zarf içine gömülü üç yapısal proteinden oluşmaktadır: membran, zarf ve spike (S). Yapısal proteinlerden birisi olan S proteinleri, konakçı hücreyi tanıma ve viral membran füzyonunu gerçekleştirmede rol aldıkları için konakçı hücrelerin virüs tarafından enfekte edilmesinde kritik bir rol oynamaktadır. S proteinleri, büyük trimerik glikoproteinlerdir ve iki fonksiyonel alt birim içermektedirler: konakçı hücre reseptörlerine bağlanılmasından sorumlu olan S1 alt birimi ve konakçı hücre membranı ile viral membranın füzyonunun gerçekleştirilmesinden sorumlu S2 alt birimi. S1 alt birimi, N-ucu bölgesinden (NTD) ve doğrudan konakçı hücrenin reseptörüne bağlanabilen reseptör bağlanma bölgesinden (RBD) oluşmaktadır. RBD,

reseptör bağlanma yüzeyini geçici olarak gizlemek veya ortaya çıkarmak için katı cisim benzeri konformasyonel bir hareket yapmaktadır. Bu hareket sonucundaki pozisyonuna göre, S proteini hedef reseptörle etkileşime girebildiği veya iki yapı arasında çakışma oluştuğu için etkileşime giremediği pozisyonlarda olabilmektedir. Reseptörle etkileşime giremediği halde (kapalı hal), S proteininin yapısındaki tüm RBD'ler aşağı pozisyonda iken, reseptörle etkileşime girebildiği halde (açık hal), en azından tek bir RBD yukarı pozisyonda bulunmaktadır. Bağlanma ve füzyon mekanizmasının başlayabilmesi için, S proteininin, füzyon öncesi konformasyonunda bulunduğu kapalı halden açık hale geçmesi ve konakçı hücrenin reseptörüne bağlanması gerekmektedir. SARS-CoV-2 S proteinleri, insanlardaki solunum yolu epitel hücrelerinin membranları üzerinde bulunan anjiyotensin-dönüştürücü enzim 2 (ACE2) reseptörlerini hedeflemektedirler ve yapılarında bulunan RBD sayesinde konakçı hücre üzerinde bulunan ACE2 reseptörlerine bağlanabilmektedirler. Bu bağlanma sayesinde S proteininde, füzyon için gereken bir dizi yapısal değişiklik gerçekleşmektedir. S proteininin yapısında konakçı hücre enzimleri tarafından kesilmesini sağlayan iki temel kesilme bölgesi bulunmaktadır: S1/S2 ve S2' bölgesi. S1/S2 bölgesinde kesilme gerçekleşmesiyle birlikte, S1 ve S2 alt birimleri birbirleriyle kovalent olmayan bağlar kuracak şekilde etkileşimde kalmaktadırlar. S2' bölgesindeki kesilme ile birlikte, S1 alt birimi ile viral membrana bağlı şekilde bulunan S2 alt birimini birbirinden ayırılmaktadır ve proteinin konakçı hücre membranı ile etkileşmesi için gerekli olan füzyon peptidi (FP) ortamla etkileşimde olacak şekilde ortaya çıkmaktadır. Bu sayede, füzyon mekanizmasında kilit rol oynayan heptad tekrarı 1 (HR1) ve heptad tekrarı 2 (HR2) yapılarında füzyon mekanizması için gerekli olan konformasyonel değişiklikler meydana gelmektedir. HR1'in konformasyonel değişikliğe uğrayarak uzatılmış bir alfa sarmalı şeklinde yeniden düzenlenmesi ile FP konakçı hücre membranına yerleşebilmektedir. Bu yerleşim, S2 alt biriminin yapısında ve HR1 bölgesinin devamında bulunan HR2'nin, HR1'in üstüne katlanarak altılı sarmal bir demet oluşturmasını tetiklemektedir. Böylece S proteini, füzyon öncesi halden oldukça kararlı bir füzyon sonrası hale geçişini tamamlamaktadır. Sonuç olarak konakçı hücre membranı ile viral membran, bu membranların füzyonuna ve viral genomun konakçı hücreye salınmasına izin verecek şekilde yakın konuma gelmektedirler ve membran füzyonu gerçekleşmektedir. Viral genomun konakçı hücreye salınmasını önlemek için, S proteininin füzyon öncesi-sonrası geçişi sırasında gerçekleşen mekanizması herhangi bir noktada bloke edilmelidir. Konakçı hücrenin tanınması ve virüs genomunun konakçı hücreye girişi, patogenez ve viral enfektivite için en kritik adımlar olduğundan, S proteinlerinin bağlanma ve füzyon mekanizmasının çözülmesi, antiviral ilaçların geliştirilebilmesi açısından potansiyel terapötik hedef olarak öne çıkmaktadır. Literatürde, S proteininin bağlanma ve füzyon mekanizması, moleküler dinamik (MD) simülasyonları kullanılarak atomik düzeyde henüz modellenmemiş ve incelenmemiştir. MD simülasyonları, deneysel tekniklerle ulaşılması zor olan veya ulaşılamayan atomik seviyedeki yapısal, dinamik ve enerjetik bilgilere etkin bir şekilde ulaşılmasını sağlamaktadır. Tez çalışmasının temel amacı, S proteininin füzyon öncesi-sonrası geçiş mekanizmasınının adımları atomik seviyede MD simülasyonları gerçekleştirilerek modellenmesidir.

Bu tez çalışmasında, S proteininin RBD yapısının kapalı halden açık hale geçiş mekanızması MD simülasyonları ve istatistiksel termodinamik yöntemleri kullanılarak modellenmiştir. MD simülasyonlarında başlangıç konformasyonları olarak, SARS-CoV-2 S proteininin RBD yapısının, sırasıya, kapalı ve açık halde bulunduğu 6VXX ve 6VYB PDB kodlu kristal yapılar kullanılmıştır. Bu yapılar protein sekansının %76.4’ünü kapsamaktadırlar. Protein sekansının eksik kalan kısımları homoloji modelleme yöntemi kullanılarak tamamlanmıştır. Kristal yapılarda bulunan ve S proteini bağlı halde bulunan glikan molekülleri simülasyonlar boyunca korunmuştur. Öncelikle açık ve kapalı haldeki yapılar için konvansiyonel MD simülasyonları gerçekleştirilmiştir. Bu simlüasyonlar sonucunda, açık haldeki S proteinin kapalı hale göre daha hareketli olduğu görülmüştür. Kapalı ve açık hale, iki farklı pozisyonda bulunan RBD yapıları arasındaki tuz köprüsü ve hidrojen bağı analizleri, iki pozisyon arasında farklı sayıda etkileşimler olduğunu göstermiş ve iki pozisyon arasındaki hareketliliğin bu etkileşimlerin farkından kaynaklandığı görülmüştür. Buna ek olarak, iki farklı hallerdeki RBD yapılarınnın ACE2 reseptörüne bağlanma pozisyonları incelenmiş ve kapalı haldeki S proteininin ACE2 yapısı ile çakışma gösterdiği bulunmuştur. Bu çakışma, kapalı halde bulunan S proteininin ACE2 reseptörüne bağlanmasını engellemektedir. Bu nedenle, kapalı hale bağlanan bir inhibitörün bağlanma arayüzü için ACE2 reseptörü ile rekabet etmesi gerekmemektedir. Bu bulgu, RBD-ACE2 bağlanmasını önlemek için kapalı haldeki S proteinini hedef alan inhibitör moleküllerin geliştirilebileceğini göstermektedir. İki hal arasındaki geçişin elde edilebilmesi için yönlendirilmiş MD (SMD) simülasyonları uygulanarak *in silico* çekme deneyleri gerçekleştirilmiştir. SMD simülasyonları, serbest şekilde simüle edilen iki yapıdan da başlatılmıştır ve hem kapalı halde aşağıda bulunan RBD yapısını yukarı doğru çekerek açık hale getirecek şekilde hem de açık halde yukarıda bulunan RBD yapısını aşağı çekerek proteini kapalı hale getirecek şekilde gerçekleştirilmiştir. Bu şekilde kapalı ve açık hal arasındaki geçiş yolu elde edilmiştir. Konvansiyonel MD ve SMD simülasyonları sonucunda elde edilen veriler kullanılarak RBD’nin kapalı ve açık haller arasındaki geçişinde rol alan hareketleri arasındaki en önemli ve baskın olan hareketlerin belirlenebilmesi için temel bileşenler analizi (PCA) yapılmıştır. PCA, bir proteinin en belirgin hareketlerini incelemek için kullanılan etkili ve kanıtlanmış bir yöntemdir. Analizin gerçekleştirilebilmesi için kovaryans matrisi oluşturulmuştur ve bu matris protein yapısındaki alfa karbon atom koordinatları kullanılarak elde edilmiştir. Bu matrisin diyagonal elemanları her amino asit için olan varyans değerini vermektedir. Elde edilen verilere dayanılarak, ilk iki temel bileşenin toplam varyansın %96.1’ine denk geldiği bulunmuştur ve böylelikle proteinin en belirgin iki temel bileşeni belirlenmiştir. Belirlenen temel bileşenler kullanılarak serbest enerji yüzeyi oluşturulmuştur ve oluşturulan serbest enerji yüzeyinden yararlanılarak S proteininin protomerlerinin aşağı ve yukarı hali arasındaki geçişi temsil eden minimum serbest enerji yolu belirlenmiştir. Serbest enerji yüzeyi, iki hal arasında bulunan yarı-açık bir halin bulunduğunu önermektedir ve çeşitli yerlerde birkaç ilave alt hal olduğunu ortaya çıkarmıştır. RBD yarı-açık halde iken, bir protomerin RBD yapısı aşağı ve yukarı haller arasındaki geçişin ortasında bulunmaktadır, kalan RBD yapıları ise aşağı halde bulunmaktadır. Bulunan yarı-açık hal, açık ve kapalı halden farklı bir etkileşim ağı göstermektedir ve ACE2 reseptörü ile bağlanma sırasında herhangi bir çakışma göstermemektedir. Bu bulgu, RBD-ACE2 bağlanmasının yarı-açık haldeki S proteini için de gerçekleşme olasılığını göstermektedir.

Bu tez çalışmasında, SARS-CoV-2 S proteininin RBD yapısının yapısal özellikleri ve geçiş yolu araştırılmıştır. Bu tez çalışması, SARS-CoV-2 S proteininin RBD yapısının kapalı ve açık halde bulunurkenki bölgeler arası etkileşimleri, dinamiği ve çözücü erişilebilirliği hakkında bilgiler sağlamaktadır.

1. INTRODUCTION

1.1 Purpose of Thesis

In the 21st century, a wide range of veterinary and human diseases has been caused by many different coronaviruses. Recently, in December 2019, a novel coronavirus named SARS-CoV-2 caused a global health crisis worldwide and resulted in a pandemic. Therefore, outbreaks caused by these viruses have a chance to affect humans and animals in the future. Since the recognition of the host cell and the entry of the virus into the host cell are the most critical steps for pathogenesis and viral infectivity, understanding the binding and fusion mechanism of viruses has crucial importance for the development of antiviral drugs. This thesis presents information on conformational and structural mechanics of the transition pathway of S protein, which is the main structure that mediates host cell attachment and fusion, and interactional details of its receptor binding site in an attempt to aid determination of potential drug-binding points on SARS-CoV-2 S protein for potential therapeutic drug designs. The purpose of the thesis is to explore the transition mechanism of SARS-CoV-2 S protein RBD. To this aim, molecular dynamics (MD) of the SARS-CoV-2 S protein performed in order to provide atomic-level insight in the S protein transition mechanism from its receptor-inaccessible to an -accessible structure, which will benefit the development of therapeutic strategies to treat COVID-19. Thermodynamic analyses are able to achieve an in-depth understanding of biologically important characteristics, including conformational and structural dynamics, interaction properties, energy conversion, and such. Computational methods such as MD simulations are favorable methods since they can provide insight into atomic levels efficiently. Thus, the usage of computational methods to perform thermodynamic analyses is a convenient way to examine biological systems.

1.2 Coronaviruses

Coronaviruses (CoVs) are classified as a genus under the *Coronaviridae* family within the order *Nidovirales* which includes *Mesoniviridae, Arteriviridae, Roniviridae* and *Coronaviridae* families. *Coronaviridae* family comprises two subfamilies being *Coronavirinae* and *Torovirinae* and further, subfamily *Coronavirinae* is subdivided into four genera; alphacoronaviruses, betacoronaviruses, gammacoronaviruses, and deltacoronaviruses (Figure 1.1) (Fehr and Perlman, 2015). CoVs are large spherical enveloped viruses that contain positive single-stranded ribonucleic acid (RNA) (Delmas and Laude, 1990) and their genome is the second-largest genome ranging from 27 to 32 kilobases amongst all RNA viruses (Saberi et al., 2018). Genome organizations of coronaviruses are similar for the coding regions comprise of open reading frames (ORFs), spike (S), envelope (E), membrane (M), nucleocapsid (N) with 5'-3' gene order (Woo, P. C. et al., 2010). Among various numbers of ORFs, the first ORF called ORF1ab contains approximately two-thirds of the coronavirus genome, and the large part of the remaining one-third genome encodes four main structural proteins; S, E, M, and N (Perlman and Netland, 2009). The N protein is the only structural protein that interacts with the RNA and forms a helical capsid to genome packaging inside of the capsid. This nucleocapsid is surrounded by a viral envelope containing E and M proteins embedded and S protein present on the surface of the virus exposing on the outside of the viral membrane of coronaviruses and acquire them a crown-like characteristic (Figure 1.1). The E and M proteins mediate virus assembly whereas S protein participates in virus attachment and entry into the host cells (Li, F., 2016). S protein binding to the different types of host cell receptors as an initial attachment of the virus to the host cell is the primary determinant to infect the host cells by coronaviruses. Following the receptor binding, a set of structural rearrangements in the S protein is facilitated generally by proteolytic cleavage on different sites. In consequence of the rearrangements, viral and host cell membranes mix, resulting in fusion and release of the viral genome to the host cell cytoplasm. After the viral genome released into the host cell, a susceptible environment for RNA synthesis, translation, and assembly of the viral replicases is created. Following these, the structural proteins are inserted and assembled viruses are transported in vesicles to the cell surface and released (Fehr and Perlman, 2015).

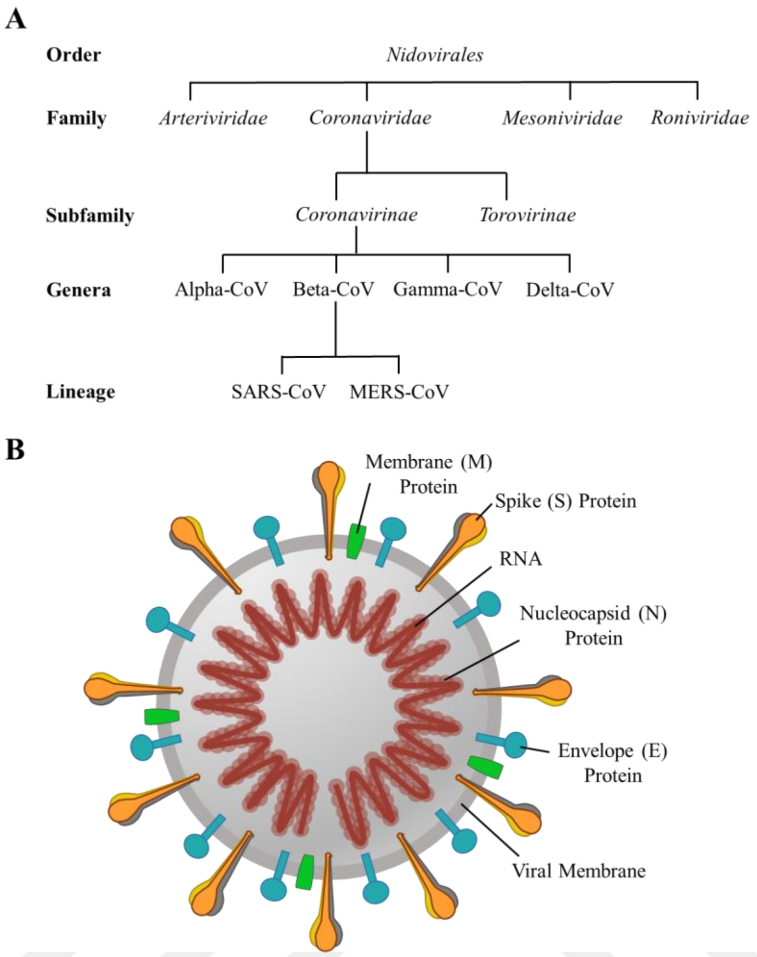
A
Order
Nidovirales
Family
Arteriviridae
Coronaviridae
Mesoniviridae
Roniviridae
Subfamily
Coronavirinae
Torovirinae
Genera
Alpha-CoV
Beta-CoV
Gamma-CoV
Delta-CoV
Lineage
SARS-CoV
MERS-CoV
B
Membrane (M) Protein
Spike (S) Protein
RNA
Nucleocapsid (N) Protein
Envelope (E) Protein
Viral Membrane

Figure 1.1 : (A) The taxonomy tree of the order *Nidovirales*. (B) Schematic representation of the structure of a coronaviruses.

CoVs are responsible for multifarious diseases in humans and companion and livestock animals. Alpha- and betacoronaviruses are composed of mammalian CoVs, gammacoronaviruses are composed of mostly avian and few mammalian CoVs, and deltacoronaviruses are composed of both mammalian and avian CoVs (Woo, P. C. et al., 2010; Woo, P. C. et al., 2012). Historically, CoVs had been associated with 15-30% of self-limiting respiratory infections each year in humans prior to the unprecedented severe acute respiratory syndrome coronavirus (SARS-CoV) outbreak and this outbreak raised the specter of CoVs to possible pandemic agents. A member of betacoronavirus genus, SARS-CoV, is identified as the cause of the severe acute respiratory syndrome (SARS) outbreak that emerged in Guangdong, China in 2002–2003 and infected 8,098 and killed 774 people with a ~9% mortality rate (Drosten et al., 2003; Fehr and Perlman, 2015; Walls et al., 2020). Although SARS-CoV epidemic was quickly controlled, another betacoronaviruses member CoV, named Middle-East respiratory syndrome coronavirus (MERS-CoV), emerged in the

Arabian Peninsula in 2012 resulting in infecting ~2519 and killing 858 people with ~40% high mortality rate (Fehr and Perlman, 2015; Memish et al., 2020; Walls et al., 2020). Despite its high mortality rate, the transmission of MERS-CoV from human to human is much more difficult than SARS-CoV except for very close contact cases, and accordingly, the outbreak did not accelerate in 2013 (Killerby et al., 2020). These two CoVs are originated from the same precursor, bats, however, the intermediate hosts of MERS-CoV is found as dromedary camels while it is raccoon dogs and palm civets for SARS-CoV (Guan et al., 2003; Haagmans et al., 2014; Kan et al., 2005; Memish et al., 2013; Wang et al., 2005). Likely, the precursors of betacoronaviruses capable of crossing the species barrier and successfully adapting transmission from animals to humans as happened in 2002 and 2012 (Peiris et al., 2004). These properties elevated CoVs as a major public health threat globally since there was a chance for other outbreaks derived from CoVs and more capable of transmission from human to human. The genetic diversity they have ensures favorable utilization of the unique and conserved mechanisms of binding and fusion mechanism to infect different host cells. CoVs are able to adapt to new conditions through recombination and mutation uncomplicatedly, thus, they can alter the tissue tropism and range of their host targets efficiently (Graham and Baric, 2010; Li, W. et al., 2006).

1.3 Severe Acute Respiratory Syndrome Coronavirus 2

In late December 2019, a disease named coronavirus disease 2019 (COVID-19) originating from a novel human pathogen CoV named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was initially identified in the city of Wuhan, Hubei province of China. Since then, COVID-19 is associated with an outbreak that has affected more than 6,900,000 people and killed over 400,000 people of whom affected in at least 200 countries as of June 2020. Thus, the World Health Organization (World Health Organization) declared in January 2020 that SARS-CoV-2 is a public health emergency of international concern and in March 2020 as a pandemic (World Health Organization). Based on its genomic sequence, SARS-CoV-2 is classified as a member of betacoronaviruses and originated from bats like SARS-CoV and MERS-CoV (Lu, R. et al., 2020; Wu et al., 2020; Zhou et al., 2020b). Thus, genetic analyses have revealed that SARS-CoV-2 has an 82%

sequence identity to SARS-CoV and only 51.8% sequence identity to MERS-CoV (Han, Yu and Yang, 2020; Lu, R. et al., 2020; Wu et al., 2020; Zhou et al., 2020b). Correlatively, SARS-CoV-2 transmission from human to human has been confirmed similarly to SARS-CoV and unlikely to MERS-CoV (Chan et al., 2020). Transmission of SARS-CoV-2 among humans occurs more rapidly and faster than SARS-CoV and MERS-CoV (Li, W. et al., 2003; Paraskevis et al., 2020; Wrapp et al., 2020; Zhu et al., 2020). Host cell recognition and virus entry into the host cell, mediated by S proteins in all CoVs, have vital importance on viral pathogenicity and infectivity and SARS-CoV-2 has notable variations in its S protein structure distinct from other CoVs that might explain the rapid infectiousness (Gallagher and Buchmeier, 2001).

S protein is a densely glycosylated homotrimeric glycoprotein classified under class I fusion proteins (Bosch et al., 2003). During its biosynthesis, S protein is first synthesized as a monomer and as a result of multiple steps of conformational changes and glycosylations, assembled as a trimer to form different surface spikes of CoVs (Delmas and Laude, 1990). Each protomer of SARS-CoV-2 S protein contains two functional subunits: S1 subunit is liable for receptor binding while the S2 subunit mediates the membrane fusion (Kirchdoerfer et al., 2018; Walls et al., 2017). S1 subunit consists of the signal peptide (SP), N-terminal domain (NTD) and receptor-binding domain (RBD) and the latter two domains are functional and responsible for receptor binding (Li, W. et al., 2003). S2 subunit comprises fusion peptide (FP), heptad repeat 1 (HR1), connector domain (CD), heptad repeat 2 (HR2), transmembrane domain (TM), and a cytoplasmic tail (CT) which FP, HR1, and HR2 are the functional domains responsible for fusion machinery (Figure 1.2) (Jiang et al., 2020; Wrapp et al., 2020). Additional to functional domains, each S protein protomer has two cleavage sites in their structure: S1/S2 cleavage site and S2’ cleavage site. The S1/S2 cleavage site is localized between S1 and S2 subunits and 103 residues downstream from the S2’ cleavage site. S2’ cleavage site, on the other hand, is localized upstream of the FP (Bosch et al., 2003; Kirchdoerfer et al., 2018; Park, J.-E. et al., 2016). Besides, SARS-CoV-2 S protein structure contains a furin recognition site insertion in the S1/S2 site which means that the SARS-CoV-2 has a bigger potential to attack multiple organs than other CoVs in the same family since furin is found in immense tissues in humans (Coutard et al., 2020; Mallapaty, 2020).

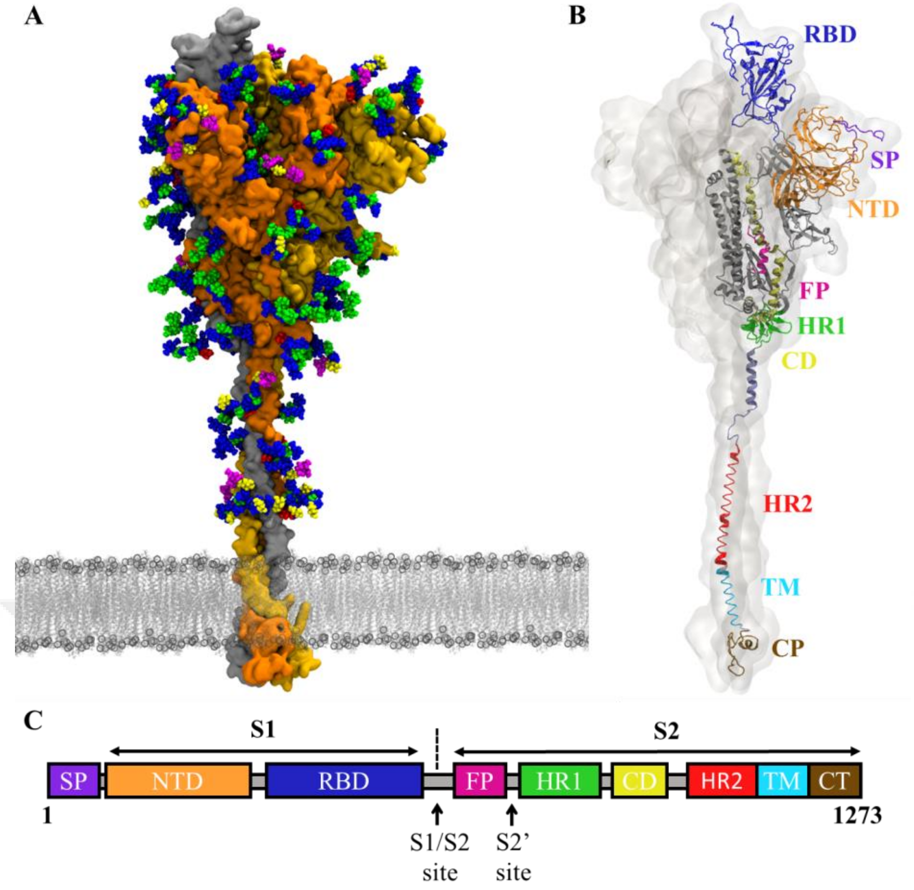

Figure 1.2 : (A) Crystal structure of the SARS-CoV-2 S protein with its one RBD in up position. S protein sequence was taken from NCBI having RefSeq: YP_009724390 (O'Leary et al., 2016). Full length S protein structure was modeled by using the S protein crystal structure having PDB ID 6VYB and comprising %77 of the sequence as a template. Missing regions located in the residue stretch A27-P1140 were modeled using homology modeling web server SWISS-MODEL (Waterhouse et al., 2018) and missing structural models for residue stretches M1-Q26 and L1141-T1273 were modeled based on the fully-glycosylated full-length SARS-CoV-2 S protein model of Woo et. al. (Woo, H. et al., 2020) and Casalino et al. (Casalino et al., 2020), respectively. Protomers of the S protein are shown in straw yellow, silver, and orange. The coordinates and structures of the glycans were derived from the study of Casalino et al. (Casalino et al., 2020). Glycans are colored bsed on the Symbol Nomenclature for Glycans (SNFG) (Nomenclature). S protein structure was manually inserted into the membrane by its TM domain using VMD (Humphrey et al., 1996). Location of TM domain selected for S protein based on UniProt ID P0DTC2 (Consortium, 2019). Phospholipid head groups and tails of the viral membrane are shown in gray color. (B) S protein protomer domain structures in its up conformation are shown. Domains of the S protomer; SP, NTD, RBD, FP, HR1, CD, HR2, TM, and CP are shown in purple, orange, dark blue, pink, mustard, green, red, cyan and brown, respectively. (C) Representative scheme of the functional domains and cleavage sites of the SARS-CoV-2 S protein.

In an attempt to trigger conformational changes in the S protein structure and facilitate the fusion process, binding to the host cell receptor should take place initially. SARS-CoV-2 S proteins attach to the respiratory epithelial cells in humans by binding to the angiotensin-converting enzyme 2 (ACE2) receptors of the host cells with RBD structures located in S1 subunit (Letko et al., 2020; Wrapp et al., 2020). To be able to bind host cell receptors, S protein RBD switches between two positions, up and down position. Related with the RBD position, S protein possesses two states: the closed (receptor-inaccessible) and open (receptor-accessible) state. While S protein is in the closed state, all RBDs get in the down position covering up the S2 subunit whereas, in the open state, a single RBD at least gets in the up position and outwardly rotates from S2 subunit (Figure 1.3). The binding surface of the RBD is exposed in the open state and provides a favorable conformation for binding, moreover, the conformational switch from down to up position is a prerequisite for S protein-ACE2 binding (Li, W. et al., 2003; Wong et al., 2004; Zhou et al., 2020a).

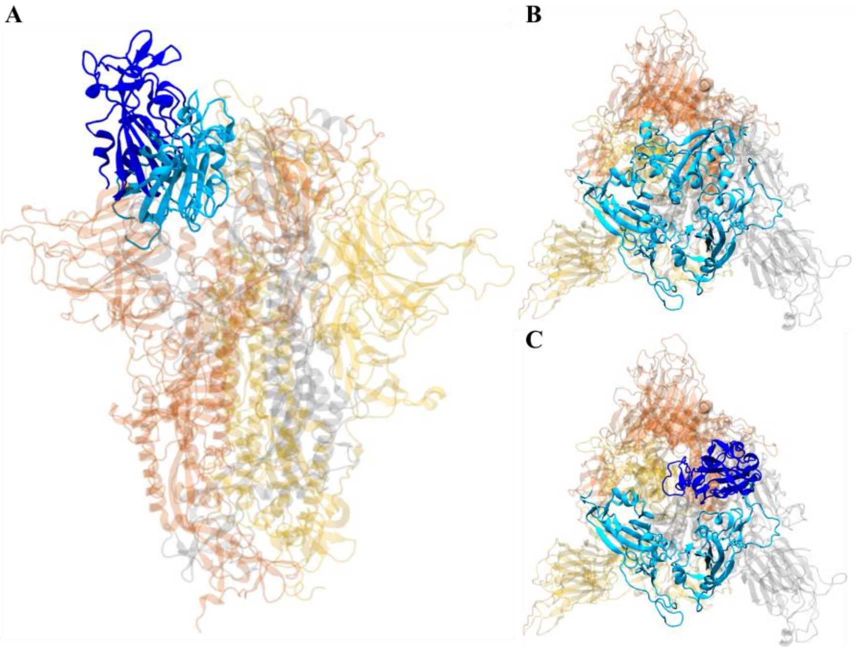

Figure 1.3 : Down and up positions of RBD of the SARS-CoV-2 S protein in the closed and open states. (A) Side view of the trimeric S protein conformation is shown. Closed and open state conformations sampled from MD simulations structures are aligned based on their secondary structures in the S2 subunit and only the RBD is shown for the open state. RBDs in the closed and open states are shown in cyan and dark blue colors, respectively. Protomers A, B, and C are shown in straw yellow, silver, and orange. S protein trimer is shown from the top view for the (A) closed and (B) open states.

Upon S protein binding to ACE2 and proteolytic cleavage of S protein into S1 and S2 subunits, large-scale conformational changes in the S protein mediate interactions with the host cell membrane and facilitate the fusion process. S1/S2 site cleavage divides S protein into S1 and S2 subunits which are remained as non-covalently bound. However, there is no consensus on whether cleavage on the S1/S2 site is a requirement for pre- to post-fusion state transition of S protein. On the other hand, S2’ site cleavage is a prerequisite for the transition of S protein from pre- to postfusion state and membrane fusion, since this cleavage divides S protein into two subunits and exposes the FP in S2 subunit to the environment (Bosch et al., 2003; Kirchdoerfer et al., 2018; Park, J.-E. et al., 2016). In order to insert the FP into the host cell membrane, S2’ site cleavage promotes structural changes on HR1 to form an extended α-helix. Consequently, HR2 bends on the HR1 and these two domains form a six-helix bundle structure (Bosch et al., 2003; Bullough et al., 1994; Walls et al., 2017). This completes the S protein transition from its pre- to highly stable postfusion state and results in getting viral and host cell membranes into proximity. Thus, the membrane fusion occurs allowing the viral genome releasing into the host cell (Figure 1.4) (Walls et al., 2017).

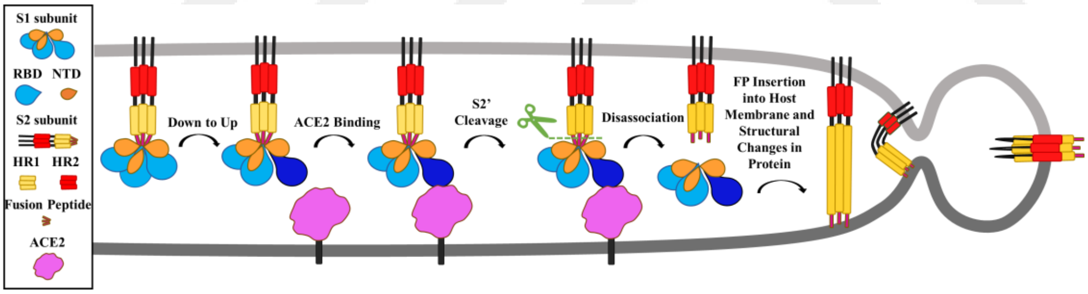

Figure 1.4 : Schematic illustration of the binding and fusion mechanism of SARS-CoV-2 S protein. The virus membrane and host cell membrane are shown in light and dark gray, respectively.

1.4 SARS-CoV-2 Interaction With ACE2

For viruses, receptor recognition is the fundamental first step to mediate viral infections and CoVs have a complicated receptor recognition pattern to recognize a different set of host cell receptors (Baranowski et al., 2001; Li, F., 2015). Each target receptors for CoVs are functional proteins that possess different physiological functions in the cells. For SARS-CoV-2, ACE2 proteins on the host cells are the functional receptors and it attaches to the human respiratory epithelial cells upon binding to the ACE2 by RBD of S protein (Letko et al., 2020; Wrapp et al., 2020). ACE2 is an integral membrane protein consisting of two homodimers and each protomer of the homodimer has 805 amino acids in its structure (Donoghue et al., 2000; Tipnis et al., 2000). ACE2 is a zinc-dependent peptidase utilizing zinc as a cofactor and each of its protomers can bind with one zinc ion and since chloride also regulates ACE2, it can bind with one chloride ion (Burrell et al., 2004; Guang et al., 2012; Kuba et al., 2010). Conversion of angiotensin is catalyzed by ACE2 cleaving a residue at the angiotensin peptide C-terminus which is a peptide hormone governs the regulation of blood pressure in the epithelial cells at the different organs such as kidneys, heart, intestines, and lungs (Donoghue et al., 2000; Towler et al., 2004; Yan et al., 2020). Moreover, ACE2 protects the human body against severe acute lung failure and downregulation of ACE2 caused by SARS-CoV results with lung injury (Imai et al., 2005; Kuba et al., 2005). Each protomer of ACE2 contains a peptidase domain (PD) at N-terminal and Collectrin-like domain (CLD) at C-terminal conjunction with TM and intercellular part. Two PD domains generate a structure looking like a claw in the homodimeric form of the ACE2 and the bottom of the claw structure cavity has a catalytic activity that is liable for binding and cleavage of angiotensin (Prabakaran et al., 2004). ACE2 has Loop3 and 4, and α1- and α2-helices, which are located on the PD at a distance from the catalytic site, interacting with S protein RBD via β3- and β4-sheets on receptor-binding motif (RBM) of RBD according to crystal structures of RBD-ACE2 complex (Burrell et al., 2004; Lan et al., 2020; Li, F., 2008; Li, F. et al., 2005; Yan et al., 2020). RBM is one of two entities that SARS-CoV-2 have and the remaining part is composed of the core. The core comprises five antiparallel β strands (β1, 2, 3, 4, and 7) that form twisted β-sheet structure, loops, and short helices while strands β5 and β6, helices α4 and α5, and other loops are located on the RBM (Lan et al., 2020; Shang et al., 2020). The majority of the ACE2 contacting residues of SARS-CoV-2 S protein RBD are found on the RBM (Figure 1.5).

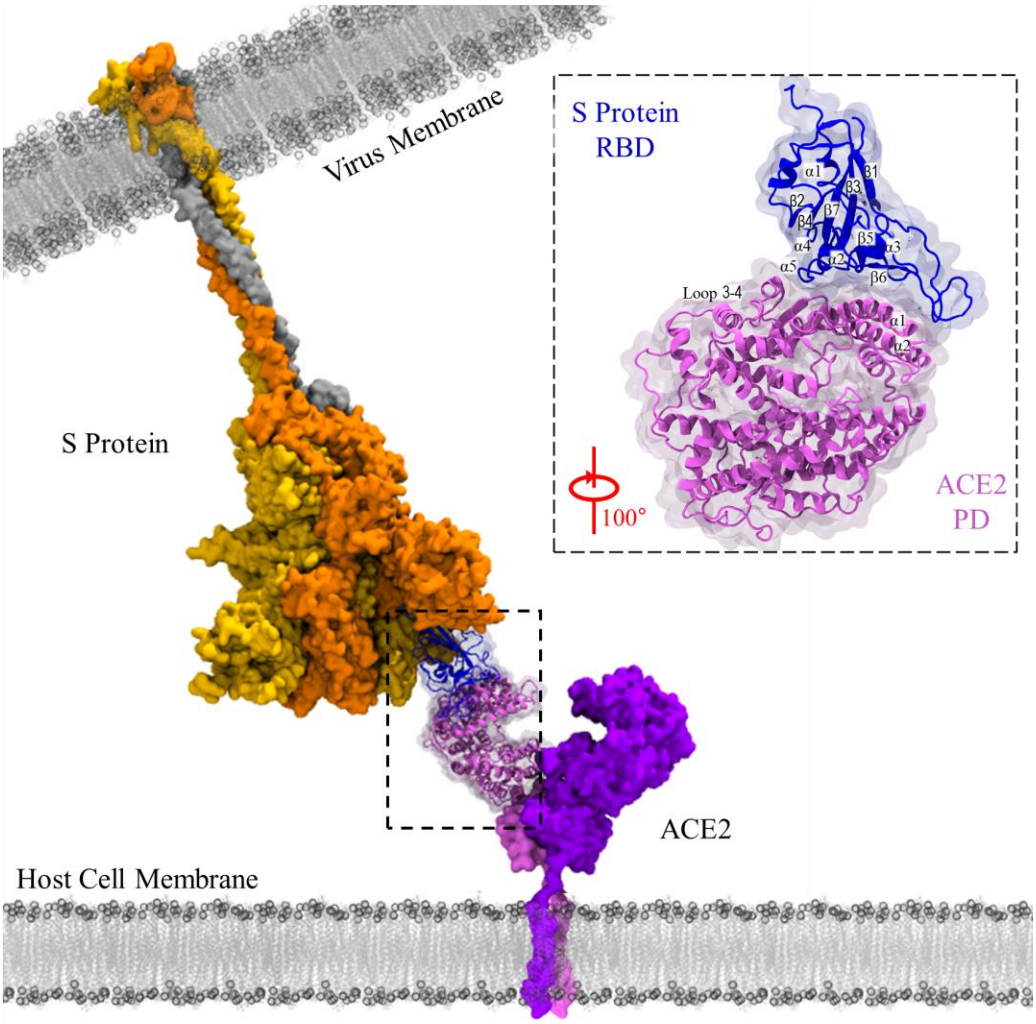
Virus Membrane
S Protein
S Protein
RBD
α1
β1
β3
β2
β7
β4
α4
β5
α3
α5
α2
β6
Loop 3-4
α1
α2
100°
ACE2
PD
ACE2
Host Cell Membrane

Figure 1.5 : SARS-CoV-2 attachment to the host cell by virtue of S protein RBD binding to ACE2 PD. Trimer S protein and ACE2 dimer (PDB ID: 6M17) are shown embedded in the viral and host cell membranes, respectively. S protein and ACE2 structures were manually inserted into the membrane by its transmembrane domains using VMD (Humphrey et al., 1996). Location of transmembrane domains selected for S protein and ACE2 based on UniProt IDs P0DTC2 and Q9BYF1, respectively (Consortium, 2019). S protein protomers are shown in straw yellow, silver, and orange. ACE2 protomers are shown in bright lilac and purple. Phospholipid head groups and tails are shown in grey color. Glycans are not shown.

1.5 Hypothesis

The novel human pathogenic virus, SARS-CoV-2, have spread all around the world and its significant effects pose a health emergency globally. Although there is an urgent need for vaccine and drug development, currently, no proven treatment has been found for COVID-19. Since the virus attachment and entry to the host cell have a critical role in viral infections, S protein has become a popular target for potential therapeutics. To develop suitable therapeutics and curb SARS-CoV-2, the molecular mechanism and key points of the binding mechanism should be discovered. To this aim, the fundamental step of the viral infection is investigated in this thesis using computational and thermodynamic methods; the receptor binding mechanism of the SARS-CoV-2 S protein. SARS-CoV-2 S protein structure and dynamics are investigated via all-atom MD simulations to provide crucial information at the atomic-level on the transition and binding mechanism of S protein. Since S protein is highly glycosylated, it is important to determine potential target points on it and since it is hard to determine these points experimentally, MD simulations provide a visual aspect to explore movement, dynamics and structure of the S protein. Thus, S protein switches between two states and the characteristics of this switching mechanism such as when it occurs, which movements are taking part in this movement, which interactions are the key ones to facilitate it and how its solvent accessibility changes provide the fundamental and strongly benefit information to develop therapeutic strategies against COVID-19. Moreover, there is an outstanding effort to design inhibitors, for instance, peptide inhibitors based on ACE2 structure and small molecules, targeting the protein-protein interaction surface of the ACE2-RBD complex (Han, Yanxiao and Král, 2020; Smith and Smith, 2020; Varshney et al., 2020; Zhang et al., 2020). Determining the binding positions and dynamics of the SARS-CoV-2 S protein RBD might provide essential information on RBD-ACE2. Since three outbreaks originated from CoVs had occurred in the 21st century, health threats from these viruses would be constant in the long-term. The energetic point of view using all-atom MD simulation techniques and thermodynamic analysis would contribute not only to the SARS-CoV-2 S protein characteristics, also for the understanding of the general characteristics of all CoVs.

2. METHODS

2.1 Theory of Molecular Dynamics

Molecular dynamics (MD) is a computational simulation method that integrates the classical (Newtonian) equations based on Newton's second law or equations of motion to investigate the motion of a set of particles. Newton's second law can be written as for each atom as seen in equation 2.1, where $\vec{\boldsymbol{F}}_{\boldsymbol{i}}$ is the force applied on the particle, $\boldsymbol{m}_{\boldsymbol{i}}$ is the mass of the particle and $\vec{\boldsymbol{a}}_{\boldsymbol{i}}$ is the acceleration (Leach, A.R., 2001).

$$\vec{F}_i = m_i \, \vec{a}_i \qquad (2.1)\ (1)$$

The corresponding force $\vec{\boldsymbol{F}}_{\boldsymbol{i}}$ applied on the atoms are derived from potential energy as seen in equation 2.2, where $\boldsymbol{r}_{\boldsymbol{i}}$ is the coordinates of particles and $\boldsymbol{t}$ is the time;

$$\vec{\boldsymbol{F}}_{\boldsymbol{i}} = -\vec{\boldsymbol{\nabla}}_{i}\,\boldsymbol{V} = -\,\frac{\boldsymbol{dV}(\boldsymbol{r}_{\mathbf{1},\dots,}\boldsymbol{r}_{\boldsymbol{i}})}{\boldsymbol{dr}_{\boldsymbol{i}}} = \boldsymbol{m}_{\boldsymbol{i}}\,\frac{\boldsymbol{d}^{\mathbf{2}}\boldsymbol{r}_{\boldsymbol{i}}}{\boldsymbol{dt}^{\mathbf{2}}} = -\,\frac{\boldsymbol{dV}}{\boldsymbol{dr}_{\boldsymbol{i}}} \qquad (2.2)$$

The integration of Newtonian equations to a set of particles gives a result as a trajectory of the molecular system through time and gives information at the microscopic level such as atomic velocities and positions of each atom. Since most of the experiments can only measure the macroscopic properties of a system, the conversion of microscopic properties to observable macroscopic properties including energy, pressure, and heat capacities using statistical mechanics is required. The goal of statistical mechanics is to understand macroscopic phenomena from the features of individual atoms of the molecular systems and statistical averages independent from the time are often introduced in an attempt to make a connection between microscopic and macroscopic systems (Ouldridge, 2018). Since experiments usually contain a large number of molecule samples with tremendous conformations, statistical mechanics provides the usage of ensemble averages correspond to a large number of experimental examples. An ensemble is defined as a collection of a large number of possible copies of the system having distinct microscopic states but an identical macroscopic state. The thermodynamic state of a system is defined by a different set of parameters such as temperature (T), pressure (P), and the number of particles (N). In MD, different types of ensembles can be used (Frenkel and Smit, 2001; Wilde and Singh, 1998);

- Microcanonical Ensemble (NVE): In the microcanonical ensemble, the system is isolated and the number of atoms (N), volume (V), and the energy (E) of the system is kept constant.
- Canonical Ensemble (NVT): In the canonical ensemble, the system is in thermal equilibrium, and the number of atoms (N), volume (V), and the temperature (T) of the system is kept constant while the system has different total energies in its possible states.
- Isobaric-Isothermal Ensemble (NPT): In the isobaric-isothermal ensemble, the number of the atoms (N), pressure (P), and the temperature (T) of the system is kept constant.
- Grand Canonical Ensemble (μVT): In the grand canonical ensemble, the system is in thermal and chemical equilibrium and chemical potential (μ), volume (V) and the temperature (T) of the system is kept constant while the total energy and the number of particles of the system can differ.

The two most common ways to generate equilibrium ensembles are Monte Carlo and MD simulations. MD simulations can be used to investigate individual particle motions and dynamics properties of a molecular system as a function of time in numerous fields including molecular biology, biophysics, biochemistry, and biotechnology. All classic simulation methods rely on a mathematical calculation model that describes the energy of a molecule with a function called force field that depends on the atomic coordinates of the particle. Force fields are mainly subdivided into two parts: bonded and non-bonded interactions. Bonded interactions consist of covalent bond-stretching, angle bending, and torsion angle while non-bonded interactions comprise of van der Waals and Coulomb forces. Using the contribution of bonded and non-bonded interactions, the potential energy function of a molecular system can be evaluated (MacKerell Jr, 2004). The values of distances between atoms, torsion angles, bond lengths, and angles for potential energy evaluation are derived from model structures obtained from experimental nuclear magnetic resonance (NMR) or X-ray structures. The rest of the required parameters for potential energy function calculations are derived from reproduced experimentally known information or *ab initio* data. Various force fields are obtained for different molecules and the most common force fields are AMBER (Ponder and Case, 2003), CHARMM (MacKerell Jr et al., 2002), and GROMOS (Scott et al., 1999). In this thesis, the CHARMM36 force field is used and the energy function of the CHARMM36 force field is given in the equation 2.3, where the first three terms (sum of covalently bonded atoms, bending of bond angles, and torsions) in the equation above are associated with covalently bonded interactions while the last two terms (van der Waals interactions and electrostatic interactions) describe non-bonded interactions between atoms (Best et al., 2012; MacKerell Jr et al., 1998).

$$\begin{aligned} V_{total} = & \sum_{bonds} \frac{1}{2} k_b \left(b - b_{eq}\right)^2 + \sum_{angles} \frac{1}{2} k_\theta \left(\theta - \theta_{eq}\right)^2 \\ & + \sum_{dihedral} k^{\emptyset}[1 + cos(n\emptyset - \emptyset^0)] \\ & + \sum_{nonbonded} (\epsilon_{ij}[\left(\frac{Rmin_{ij}}{r_{ij}}\right)^{12} - \left(\frac{Rmin_{ij}}{r_{ij}}\right)^{6}]) \\ & + \frac{q_i q_j}{4\pi\epsilon r_{ij}} \end{aligned} \tag{2.3}$$

In MD simulations, different algorithms and parameters are used for precise calculations, and in general, simulations follow similar experimental procedures. Since experiments need material for research at the beginning, the initial crystal structures of the molecules (proteins, nucleic acids, etc.) for the MD simulations are derived from the experimental methods such as X-ray and cryogenic electron microscopy (cryo-EM) and obtained structures of the molecules (e.g. atomic coordinates) are collected in the protein data bank (PDB) with specific identification codes for each molecule (Berman et al., 2000). The selected molecule is solvated and ionized before starting the simulations to represent a more typical biological environment since biological activities arise in solvent environments and from interactions between molecules (protein-protein, protein-solvent, etc.). To this aim, solvents can be added to systems explicitly or implicitly. In the implicit solvent model, solvent molecules are modeled as a continuous medium whereas, in the explicit solvent model, solvent molecules are added to the molecular system explicitly. Since the explicit solvent allows specific solvent-molecule interactions during MD simulations, it provides a more realistic environment for biomolecular systems. In order to preserve the number of the particles of the system, periodic boundary conditions (PBCs) are used for approximating an infinite (large) system by using a small part (unit cell) of the system. With PBCs, molecules can pass through one side of the cell and appear on the opposite side of the cell without changing any property and if the size of the molecular system box is large enough, the molecules would not interact with their periodic copies (Figure 2.1) (Leach, A.R., 2001).

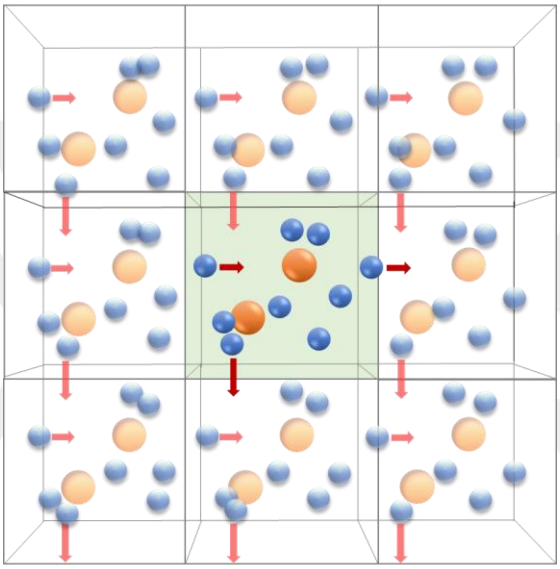

Figure 2.1 : Schematic representation of PBCs. The molecules in the unit cell (light green) of the infinite system are shown with bright orange and blue in the middle of the scheme while the rest of the representative infinite system is shown with faded colors.

Even though periodic copies of the molecules do not interact with each other, the non-bonded (long-range) interactions should be taken into consideration and summed over all neighbors in the infinite system. Cutoff values for long-range interactions can be used to avoid this problem but for Coulomb interactions, errors might occur as a result of sudden cutoffs. As a better solution, to overcome these problems, the particle mesh Ewald (PME) method containing a summation that splits short- and long-range interactions is used in MD simulations (Leach, A.R., 2001). Since non-bonded interactions are one of the demanding parts of the MD simulations, extending the time step contributes to the simulation performance. To avoid bond vibration errors when the time step is extended, bond constrains algorithms like SHAKE (Ryckaert et al., 1977) are applied to MD simulations. The initial structures that are obtained from experimental studies usually do not have proper contacts for simulations due to the difference between experimental and actual conditions. Thus, in order to bring the structures to desirable conditions, the minimization and equilibration steps are applied after preparing the simulation systems with solvation and ionization. After obtaining the suitable initiation structure, different parameters can be applied to MD simulations related to the research subject of interest.

2.2 Molecular Dynamics System Preparation

In this thesis, all visualization and system preparation steps were performed using the Visual Molecular Dynamics (VMD) program (Humphrey et al., 1996). For exploring transition mechanism between the closed and open state of SARS-CoV-2 S protein, crystal structures of SARS-CoV-2 S protein in the open (PDB ID: 6VYB (Walls et al., 2020)) and closed (PDB ID: 6VXX (Walls et al., 2020)) states were used as the template for starting structures in MD simulations. Both crystal structures cover 76.4% of the protein sequence including the RBD and NTD of S1 subunit and HR1 and FP of S2 subunit for each protomer of trimeric S protein. For each protomer, the protein sequence comprises residues from A27 to S1147. Out of these 1121 residues, structural information of 149 residues are missing in the closed state: V70-F79, Y144-N164, Q173-N185, R246-A262, V445-G446, L455-C488, G502, P621-S640, Q677-A688, L828-Q853. In the open state, structural information of 155, 172 and 161 residues are missing for A, B and C protomers, respectively: V70-N81, T114-Q115, Y144-N165, Q173-N185, A243-A262, S443-G447, E471-Y489, G502, P621-S640, Q677-S689, P812, L828-L854 of protomer A, A67-D80, L141-A163, Q173-N185, I197-G199, L212-R214, A243-A262, L455-L461, D467-F490, E516-P521, P621-S640, Q677-A688, P812, L828-Q853 of protomer B, and A67-D80, Y144-N165, Q173-N185, A243-A263, V445-G447, L455-L461, E471-F490, P621-S640, Q677-S689, P812, L828-F855 of protomer C. The three-dimensional information and model of aforementioned missing parts of the S protein were obtained for open and closed states using SWISS-MODEL web server (Waterhouse et al., 2018). For the FASTA sequence of S protein, the NCBI web server is used (RefSeq: YP_009724390 (O'Leary et al., 2016)). The mutations that were performed in the crystal structure to obtain a stable structure and abrogate cleavage sites were reversed prior to SWISS-MODEL submission. N-linked glycan molecules that are attached to

S protein in the crystal structures were conserved in all simulations. Open and closed state structures were separately solvated in water boxes having at least 12 Å water paddings in each direction. TIP3P water model was used for all water boxes used for solvation in each system. After solvation, in order to neutralize the systems, NaCl ions were added and ion concentration was set to 150 mM for each system to provide physiological environmental conditions. The sizes of the closed and open state systems were of ~410000 and ~415000 atoms after solvation and ionization, respectively (Figure 2.2).

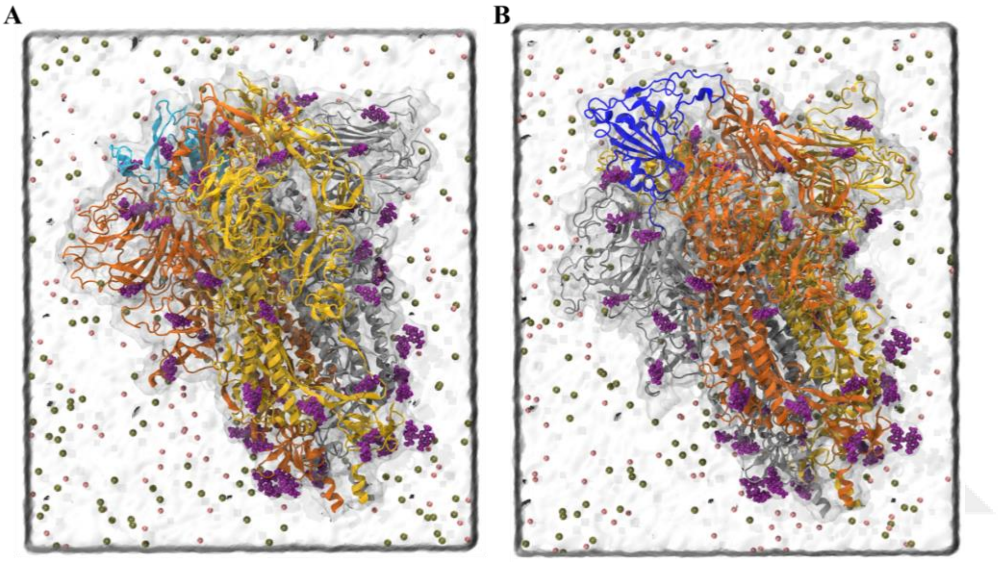

Figure 2.2 : The prepared systems for (A) closed and (B) open states of S protein in the presence of water and ions. In the figure, protomers, ions, and glycans and waters are shown in cartoon, van der Waals, and quick surface representation, respectively. Protomer A, protomer B, protomer C, glycans, sodium, and chloride ions are shown in straw yellow, silver, orange, purple, tan, and pink, respectively. RBD conformations in closed and open states shown in cyan and dark blue, respectively, to show the positions of the RBDs in different states.

2.3 Molecular Dynamics Simulations

In this thesis, all-atom MD simulations were performed using Nanoscale Molecular Dynamics (NAMD) 2.13 (Phillips, J. C. et al., 2005) and CHARMM36 (Best et al., 2012) force field under constant particle number (N), constant temperature (T) and constant pressure (P) ensemble conditions where temperature and pressure were kept constant at 310 K and 1 atm, respectively. The system temperature and pressure were maintained constant using Langevin dynamics with 1 ps^{-1} damping coefficient and Langevin Nosé-Hoover method with a 100 fs oscillation period and 50 fs damping time scale, respectively. 12 Å cutoff distance was used for van der Waals interactions and long-range electrostatic forces were computed using the particle-mesh Ewald method. A time step of 2 fs was used. For each system prepared in section 2.2, MD simulation details that are indicated in this section (section 2.3) were applied. The first 5 ns of each of MD simulations were discarded to remove the influence of the bias applied during SMD simulations. For each system, first, the system energy was minimized for 10000 steps without any constraints or restraints. Then, water molecules were equilibrated for 2 ns by keeping complete protein fixed. The second minimization was performed for 10000 steps after water equilibration. Afterwards, a harmonic constraint with 1 *kcal* mol^{-1} Å-^{2} spring constant was applied to each alpha carbon (Cα) atoms of the protein at the second equilibration for 20 ns. Subsequently, all constraints and restraints on the protein were released and 5 ns equilibration was performed. As a final step, production MD simulations were initiated. At every 10 ps, atomic coordinates ($\boldsymbol{R}$) of all atoms, energies, and pressures were recorded. For each open and closed states of the SARS-CoV-2 S protein, two sets of simulations were performed (Table 2.1).

2.4 Steered Molecular Dynamics Simulations

SMD (Isralewitz et al., 2001) is an advanced MD simulation technique that applies external time-dependent forces to pull the selected atom or atoms throughout a vector. The pulling process can be carried out at a constant speed or by applying a constant force (Phillips, J. C. et al., 2005). In the thesis, SMD simulations were performed using constant velocity; a dummy atom binds with a virtual spring to the center of mass of a group of atoms called steered atoms (SMD atoms) and is pulled along a selected pulling direction at a constant speed. Force, which is applied on SMD atoms along the vector, depends on instant coordinates of SMD atoms' center of mass ($\mathbf{R}$) as seen in equation 2.4 (Phillips, J. C. et al., 2005).

$$\mathbf{F} = -\nabla U \tag{2.4}$$

$$U = \frac{1}{2}k[\mathbf{v}t - (\mathbf{R} - \mathbf{R}_0) \cdot \mathbf{n}]^2$$

In equation 2.4, U is the guiding potential energy, k is the spring constant, $\mathbf{v}$ is the pulling velocity, t is the time instant, $\mathbf{R}$ and $\mathbf{R}_0$ are the coordinates of the center of mass of steered atoms at time *t* and 0, respectively and **n** is the direction of pulling (Phillips, J. C. et al., 2005). In SMD simulations, it is possible to explore biological processes such as molecule unbinding (Eskici and Gur, 2013), protein unfolding (Lu, H. et al., 1998) and domain motions (Izrailev et al., 1999) that are usually hard to achieve by conventional MD simulations. SMD simulations were performed to model the transition pathway between open and closed states of SARS-CoV-2 S protein. The conformations of the SMD simulations were selected from MD simulations of the open and closed states mentioned in section 2.3. Up and down RBD positions in open and closed states were superimposed and residues that have shown minimal deformations in simulations were selected as SMD atoms; C_α atoms of RBD residues P337-A344, F347-V350, W353-S371, S375-Y380, V390-I410, G416-Y423, T430-S438, G447-R454, L492-Q498, Q506-L517, T523-G526 of protomer B. Thus, residues D985 and K986 of protomer A, and K113, E988, D985 of protomer C have interacted with RBD of protomer B at the beginning of (first 20 ns) the closed state. Therefore, C_α atoms of these residues were kept fixed in SMD simulations. Up and down RBD conformations in the open and closed states were aligned based on the protomer B of S2 subunit and a vector, pointing from the center of the SMD atoms in the down confirmation to the up conformation, was constructed as pulling vector (Figure 2.3).

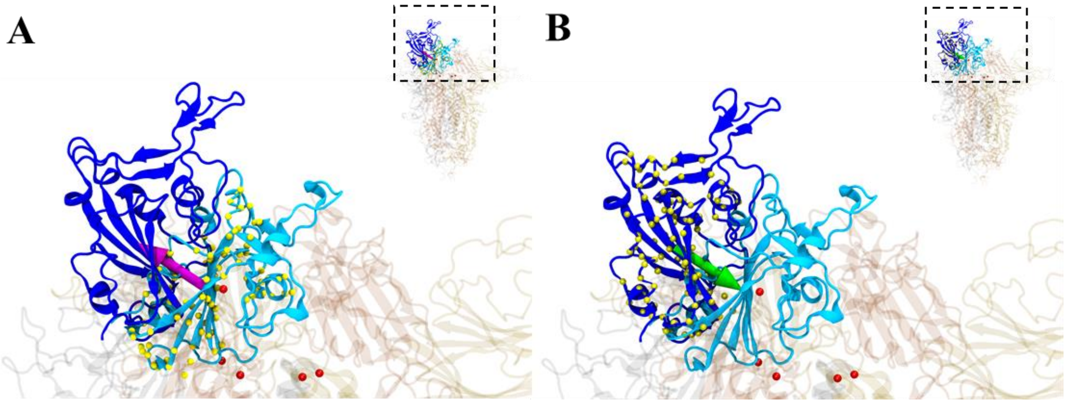

Figure 2.3 : SMD pulling vector determined for SARS-CoV-2 S protein SMD simulations. The pulling direction is selected by (A) a vector pointing from the center of mass of SMD atoms on the RBD in the down position pointing the center of mass of SMD atoms on the RBD in up position and (B) vice versa. The pulling direction is shown with a magenta and green arrow. C_α atoms of SMD atoms on RBD and fixed atoms on neighboring protomers are shown with yellow and red beads, respectively.

1 Å/ns and 50 $kcal/mol/Å^2$ were selected for pulling velocity and spring constant, respectively. It is important to ensure stiff-spring approximation (i.e. center of mass of the SMD atoms followed closely by dummy atom) and proper conditions (not too soft or stiff) for the spring constant to permit small deviations (Isralewitz et al., 2001) while determining the spring constant and pulling velocity in SMD simulations. The spring constant k of the guiding potential should be large enough for reaction coordinate ξ to follow the constraint center λ closely. Table 2.1 presents the complete list and length of each MD and SMD simulations performed for the SARS-CoV-2 S protein structure.

Table 2.1 : Starting conformations and lengths of the performed MD simulations of SARS-CoV-2 S protein.

MD ID	Starting Conformation	Simulation Type	Simulation Duration (ns)
1a-b-c	Closed state crystal structure (6VXX)	MD	(a) 150, (b) 150, (c) 150
2a-b-c	Open state crystal structure (6VYB)	MD	(a) 450, (b) 450, (c) 450
3a-b-c	Closed state conformers (MD)	SMD	(a) 9, (b) 9, (c) 9
4a-b-c	Final conformers of MD 3a-b-c	MD	(a) 100, (b) 100, (c) 100
5a-b-c	Final conformers of MD 3a-b-c	SMD	(a) 2, (b) 2, (c) 2
6a-b-c	Final conformers of MD 5a-b-c	MD	(a) 100, (b) 100, (c) 100
7a-b-c	Final conformers of MD 5a-b-c	SMD	(a) 2, (b) 2, (c) 2
8a-b-c	Final conformers of MD 7a-b-c	MD	(a) 100, (b) 100, (c) 100
9a-b-c	Final conformers of MD 7a-b-c	SMD	(a) 2, (b) 2, (c) 2
10a-b-c	Final conformers of MD 9a-b-c	MD	(a) 100, (b) 100, (c) 100
11a-b-c	Final conformers of MD 9a-b-c	SMD	(a) 2, (b) 2, (c) 2
12a-b-c	Final conformers of MD 11a-b-c	MD	(a) 100, (b) 100, (c) 100
13a-b-c	Open state conformers (MD)	SMD	(a) 3, (b) 3, (c) 3
14a-b-c	Final conformers of MD 13a-b-c	MD	(a) 100, (b) 100, (c) 100
15a-b-c	Final conformers of MD 13a-b-c	SMD	(a) 2, (b) 2, (c) 2
16a-b-c	Final conformers of MD 15a-b-c	MD	(a) 100, (b) 100, (c) 100
17a-b-c	Final conformers of MD 15a-b-c	SMD	(a) 2, (b) 2, (c) 2
18a-b-c	Final conformers of MD 17a-b-c	MD	(a) 100, (b) 100, (c) 100

2.5 Root Mean Square Deviation

In MD simulations, root mean square deviation (RMSD) is used to quantitatively measure the similarity between two molecules and represents the average distance measure between atoms of superimposed molecules. RMSD of a molecule with atomic coordinates r_i at time t and reference molecule’s atomic coordinates $r_i(0)$ is calculated as given in equation 2.5.

$$RMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(r_i(t) - r_i(0))^2} \qquad (2.5)$$

The appropriate RMSD can be calculated by the superposition of a protein conformation observed along with MD simulations to its crystal structure yielding a measure for conformational changes and allow to observe the equilibrated conformation of the molecule. On the other hand, the superposition of a protein conformation observed along with MD simulations to another protein conformation observed through MD simulations can be used to determine the conformational difference between two molecules (Kufareva and Abagyan, 2011).

2.6 Root Mean Square Fluctuation

Although the RMSD calculation can give information about conformational changes and general motions, it is difficult to obtain information about more specific motions like structural flexibility. To identify the local structural flexibility and thermal stability, root mean square fluctuations (RMSF) are used. RMSF can be calculated by temperature factors (B-factors) obtained in X-ray experiments and also, molecular dynamics simulations can be used for RMSF calculations using an equation as seen in equation 2.6, where, $\overline{\boldsymbol{R_i}}$ is the mean atomic coordinate of the ith C_α atom and $\boldsymbol{R_i}$ is its instantaneous coordinate.

$$\langle \Delta R_i^2 \rangle^{1/2} = \langle \left(R_i - \overline{R_i}\right)^2 \rangle^{1/2} \qquad (2.6)$$

2.7 Principal Component Analysis

PCA is a very efficient analysis for characterizing the most important and dominant structural and distinct changes among the internal movements and folding and non-folding features of the protein. This analysis is based on the largest fluctuations in biomolecules that occur through a subset of collective degrees of freedom. The dominance of the small subset of degrees of freedom in MD relies on the presence of internal constraints. The dominant motion that contributes mostly to the total fluctuations can be identified with PCA. PCA gives 3N motion modes for an MD system contains N atom and the most dominant motion among these modes can be identified. As a first step, protomer conformations obtained in MD simulations were aligned with the crystal structure of the open state B protomer based on the C_α atoms of the beta-sheets and helices in the S2 subunit. Thus, the highly mobile translation and rotation motions were excluded from PCA analysis to focus on the internal movements. Upon alignment, the covariance matrix $\mathbf{C}$ was constructed using the 3x459 dimensional configuration vector, $\mathbf{R}$, composed of the instantaneous C_α atom coordinates of the RBD and S2 domain helices and beta sheets as seen in equation 2.7.

$$\mathbf{C} = \langle (\mathbf{R} - \langle \mathbf{R} \rangle)(\mathbf{R} - \langle \mathbf{R} \rangle)^{\mathrm{T}} \rangle \tag{2.7}$$

Here, $\mathbf{R}$ is the 3N dimensional configuration vector and $\langle \mathbf{R} \rangle$ is the trajectory average of $\mathbf{R}$. Principal components (PCs) are derived by performing eigenvalue decomposition of $\mathbf{C}$ as seen in equation 2.8.

$$\mathbf{C} = \sum_{i=1}^{3N} \sigma_i \mathbf{p}_i \, \mathbf{p}_i^{T} \tag{2.8}$$

In equation 2.8, $\mathbf{p}_i$ is the i^{th} PC and σ_i is the corresponding variance. Thus, σ_i scales with the magnitude of motion along $\mathbf{p}_i$. PCs are ordered in descending order concerning their σ_i values. $\mathbf{p}_1$ (or PC1) has the largest variance, σ_1, thus represents the most dominant motion. Similarly, PC2 is the second most dominant motion observed in the MD trajectory. PCA was performed for conventional MD and SMD simulation trajectories of SARS-CoV-2 S protein.

3. RESULTS

3.1 Characteristics of the SARS-CoV-2 S Protein In The Closed and Open State

3.1.1 Mobility and interdomain interactions of the S protein

For each open and closed state system, two sets of MD simulations were performed starting from the crystal structures of the open and closed state. Since S protein has three identical protomers in its structure in the closed state, each of the closed state MD simulations generates three MD simulation trajectories for down conformation. As it is mentioned above and shown in Figure 1.3, in the closed state of the S protein, all three RBD structures are in the down position whereas one RBD is in the up position in the open state. Pursuant to the open and closed state crystal structures, protomers of the S protein will be termed as protomer A, protomer B, and protomer C, which protomer B represents the up conformation. All closed state protomers were superimposed based on C_α atoms of the helix and beta-sheet structures and the RMSD values were calculated. RMSD is a quantitative measure that is used for determining atomic coordinate similarities and it was found ~1 Å for closed state protomers. As shown in Figure 1.3, RBDs cover the S2 subunit of the S protein differently in different states. Distance between RBD and S2 subunit was observed along the MD trajectories of the closed and open states and the fluctuation of the distance between these two structures and the RMSF values of the RBD were investigated using the MD trajectories of the closed and open state with the intent of investigating RBD domain mobility. Substantially larger RMSF values were observed for RBD residues in the up position in the open state for RBD compared to its down position in the closed state (Figure 3.1). Additional to these results, the distance distribution between RBD and S2 indicated a significantly wider distribution while RBD is in up position compared to the down position (Figure 3.2). Standard deviations of the distance distributions of RBD-S2 were found as 2.2 Å and 0.6 Å for up and down positions of RBD, respectively. Consequently, while RBD in the up position in the open state, it has higher mobility than RBD in down position in the closed state. Yet, the superposition

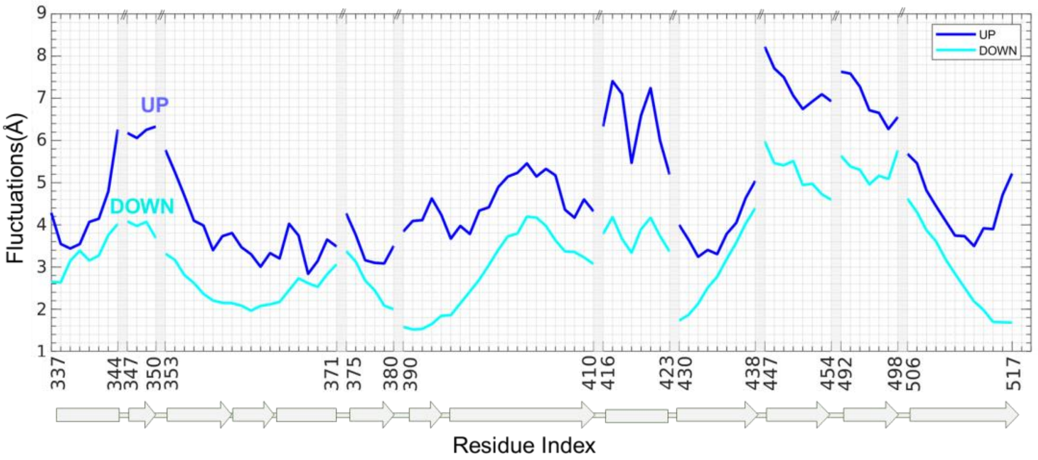

Figure 3.1 : Residue fluctuations of SARS-CoV-2 RBD residues in the down (cyan) and up (blue) position in the closed and open state, respectively. RMSFs of RBD in both conformations were calculated using coordinates of the RBD obtained from the MD simulations. All protomer conformations were aligned using the C_α atoms of the S2 domain secondary structures (helices and beta sheets) of the up protomer for the open state crystal structure. RMSF values are shown for the structured regions only. Schematic representations of the secondary structures are indicated below with bars for α helices and arrows for β sheets.

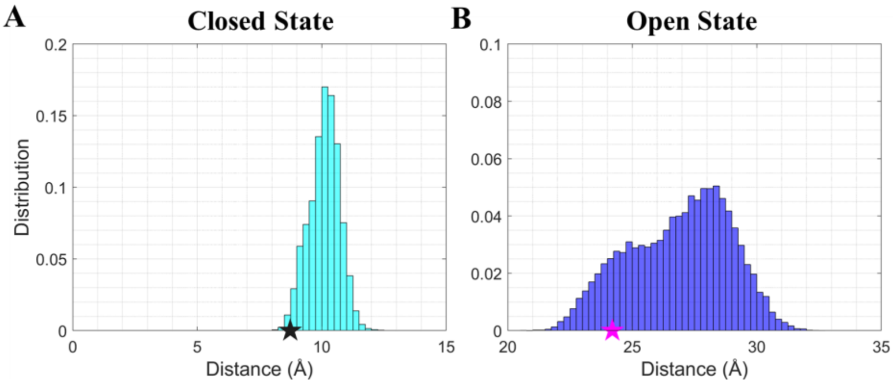

Figure 3.2 : Distance distribution between RBD and S2 subunit observed in MD simulations shown for (A) the closed state and (B) open state. The interdomain distance was calculated by evaluating the average of the distance between the Cα atoms of residue pairs K378(RBD)-E988(S2) and K386(RBD)-D985(S2). Distances calculated from the down (PDB ID: 6VXX) and up (PBD ID: 6VYB) crystal structures are marked with black and magenta stars.

of C_α atoms of the helix and beta-sheets structures of the up and down positioned RBDs shows 0.5 Å RMSD value with minimal structural differences. Thus, conformations of the RBD were conserved along with MD simulations and intramolecular changes might not cause increased mobility in the RBD structure. According to the simulation results and crystal structures, it can be estimated that RBD undergoes rigid-body motion and since no significant difference between RBD structures was found, interdomain interaction differences of RBD might cause RBD mobility discrepancy between open and closed states. The stability of proteins requires a balance between thermodynamics forces and different non-covalent interactions such as salt bridges and hydrogen bonds. To this aim, interdomain interactions which are salt bridges and hydrogen bonds of the RBD with remaining parts of the S protein were investigated for both states. For the cutoff distance of salt bridges, 6 Å was used between the basic nitrogens and acidic oxygens of salt bridge forming residues (Beckstein et al. 2009). For the hydrogen bonds, 3.5 Å distance cutoff together with 30° cutoff for the angle between the hydrogen atom was used between the donor heavy atom and the acceptor heavy atom of hydrogen bonds (Durrant and McCammon 2011). In the crystal structure of the closed state (Walls, Park et al. 2020), there are three salt bridges present between RBD and neighboring protomer: K378-E988 and K386-D985 with S2 subunit, and E516-K202 with NTD. Additional salt bridges were found in the closed state MD simulations between RBD and neighboring protomers: R408-D405 with RBD, K458-D985 with S2 domain, and K462-D198 with NTD (Figure 3.3).

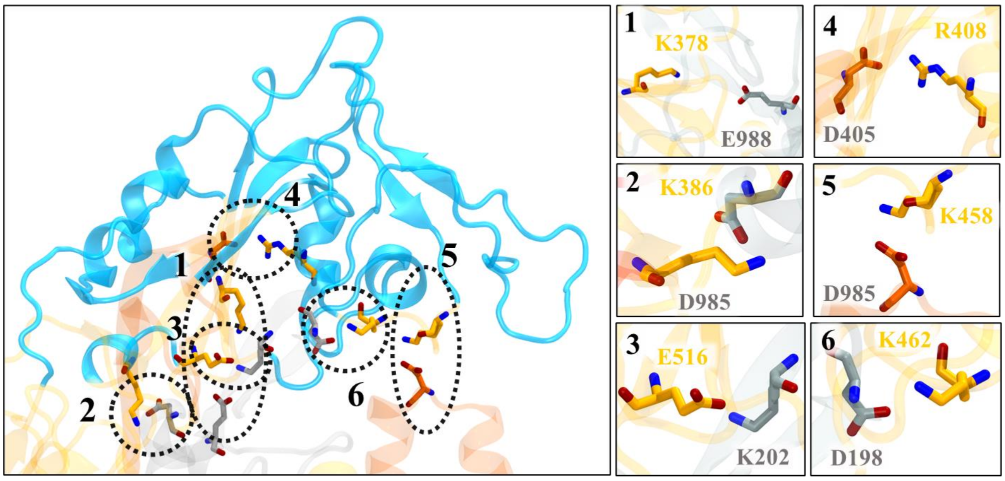

Figure 3.3 : Observed salt bridges between RBD and neighboring protomers for RBD in a down position in the closed state in MD simulations.

For obtaining distance distributions, distances between acidic oxygens and basic nitrogens of a total of six salt bridges and distances between the hydrogen bond donor and acceptor atoms of additional six hydrogen bonds of RBD were tracked throughout simulations (Figure 3.4). Thus, the frequency of occurrence of the hydrogen bonds observed in MD simulations were evaluated with HBonds plugin of VMD. While 50% and above frequencies are classified as high, frequencies less than 15% were classified as low for observation frequencies of interactions. The interaction network was consistently observed in all of the MD simulation sets. Based on the MD simulations starting from the closed state crystal structure, salt bridges between residue pairs K378-E988 and E516-K202 were observed merely occasionally and K386-D985 was observed with a moderate frequency of 42.6% in MD simulations. Each of the additional salt bridges identified in the MD simulations was observed with moderate frequency (29-33%). While hydrogen bonds between residue pairs T385-D985 and Y396-Y200 were observed with a mediocre frequency, S383-D985 and E516-Y200 were observed with a relatively high frequency. Despite a close distance distribution was obtained for residue pairs, the frequency of hydrogen bond formation was low since these residue pairs often failed to ensure the angle criteria to form a hydrogen bond. Altogether, a dynamic network was observed to be constructed by hydrogen bonds and salt bridges to stabilize RBD in the down

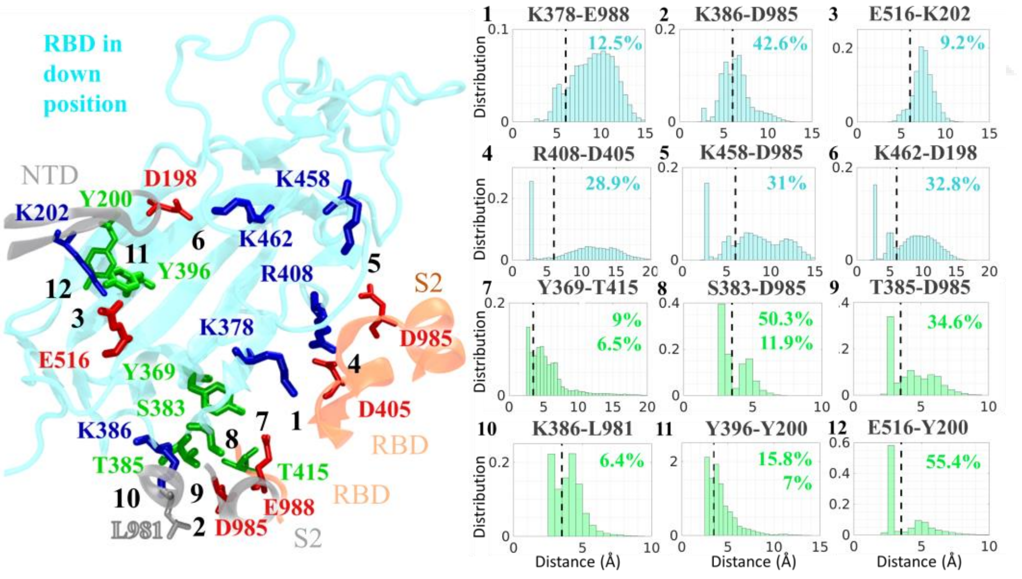

Figure 3.4 : Interactions between RBD in down position and the remaining parts of the SARS-CoV-2 S protein. Distance distribution of interdomain salt bridge (1-6) and hydrogen bond (7-12) forming residues of the RBD during MD simulations were shown. The frequency of salt bridge and hydrogen bonds as observed in MD simulations are indicated in each subplot. If a residue pair contains more than one hydrogen bond, then the observation frequency of each is given separately. Dashed lines represent the cutoff values used to determine the existence of a salt bridge or hydrogen bond.

position. In comparison to the closed state, no salt bridges were detected in the crystal structure of the open state of the S protein (Walls, Park et al. 2020) and fewer interdomain salt bridges and hydrogen bonds were observed from MD simulations starting from the open state. Based on MD simulations results, two different interaction networks were identified for the open state of the S protein (Figure 3.5). At one interaction network, the RBD interaction network was constructed by three salt bridges: R357-E169, D428-R403, and R466-E132; and two hydrogen bonds: N370-C480 and D427-Y505. All interactions in the network were observed at high frequencies, except N370-C480. At the other interaction network, the RBD interaction network was comprised of one salt bridge: R357-E132; and three hydrogen bonds: N370-G485, Y380-R403, and F429-Y505 and all these interactions were observed at moderate to high frequencies. The 5 common residues, R357 and N370 on RBD, and E132, R403, and Y505, observed on neighboring protomers in the interaction networks putting an emphasis on overall similarities of the interaction networks. Unlike these two observed interaction networks, interaction network stabilizing the up position of RBD was not observed in one set of MD simulations and the up position of the RBD was not conserved. These findings indicate that the

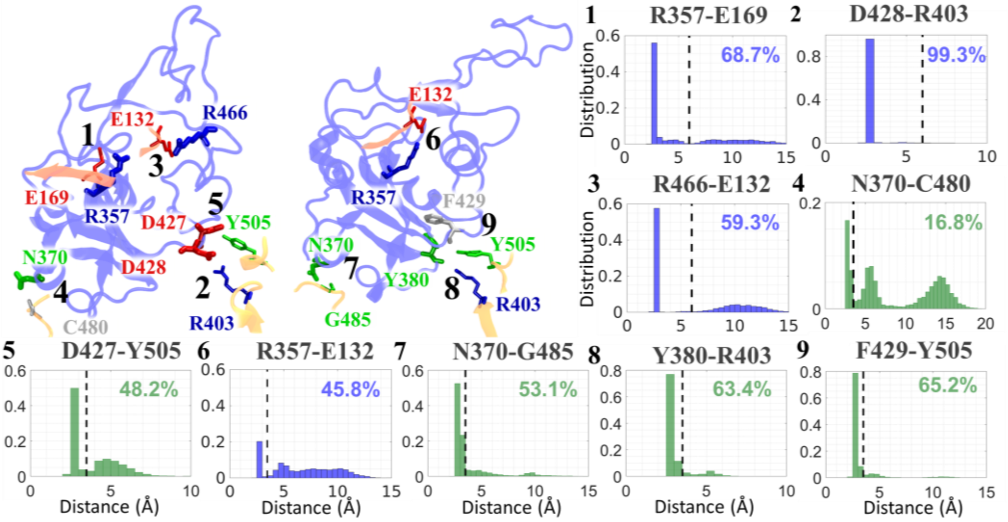

Figure 3.5 : Interactions between RBD in up conformation and the remaining S protein of the SARS-CoV-2 S protein. Two interaction networks were sampled in MD simulations. Observed salt bridges and hydrogen bonds for first (1-5) and second (6-9) interaction networks are shown for conformations of the RBD in up position sampled through MD simulations. Distance distributions between the salt bridge (blue) and hydrogen bond (green) forming residues of the RBD during MD simulations were shown. The frequency of salt bridge and hydrogen bonds as observed in MD simulations are indicated in each subplot.

variable components construct different interaction networks that stabilizes RBD in up position and variability of the interaction network shows the potential transient character of interaction network while RBD in its up position.

3.1.2 Solvent accessibility of the binding surface of RBD

Water distributions close to the receptor-binding sites are important since the water-protein interactions assist in maintaining the flexible conformation conditions required for recognition processes of protein. In the crystal structure of RBD bound with ACE2 (PDB ID: 6M0J (Lan et al., 2020)), amino acids K417, G446, Y449, Y453, L455, F456, A475, F486, N487, Y489, Q493, G496, Q498, T500, N501, G502, Y505 on RBD are found as interacting with ACE2. To examine the solvent accessibility of these residues in the closed and open states, close contact water molecules were calculated for both states. To this aim, water molecules within 5 Å and 10 Å of these residues were evaluated for each sampled conformation in MD simulations. In down and up positions, the average number of close contact water molecules did not differ substantially. Thus, in both the open and closed states, the ACE2 binding surface is fully exposed to water even S protein RBDs are closely packed in the closed state. As a next step, the RBD bound ACE2 structure was superpositioned onto the S protein in the closed state via the RBD and the ability of the S protein RBD binding to ACE2 in its closed state was investigated. A large steric clash between ACE2 and the RBD of the neighboring protomer was observed, thus, preventing RBD to bind ACE2 while in a closed state (Figure 3.6).

3.1.3 Energy landscape based on MD simulations of S protein

With an intent to understand the important and dominant motions of the protein in the open and closed states, principal component analysis (PCA) on all conformations of protomer A, protomer B and protomer C in the closed state simulations and all conformations of protomer B in the open state simulations was performed. PCA is a proven and efficient method used to reveal the large-scale dominant motions exhibiting by protein throughout its MD trajectory (Gur, Mert et al., 2018). PCA generates 3xN modes of motion for an MD trajectory with N atoms and among these motions, principal components 1 and 2 (PC1 and PC2) correspond to the first most prominent and second prominent motion observed in simulations. The PC1 and PC2 motions for both states are shown in Figure 3.7. The ratio for the variances between two PCs are found σ_1/σ_2= 33.2 and cumulatively, these two PCs liable for 96.1% of all motions in the trajectory (account for 96.1% of the total variance). PC1 is a motion

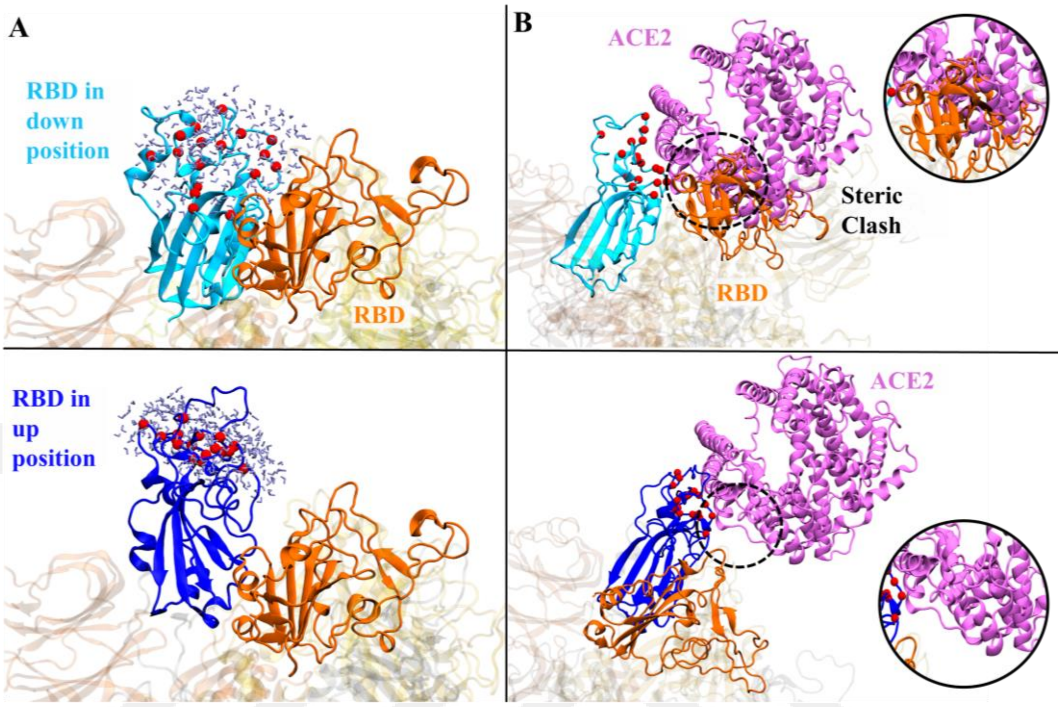
A
RBD in down position
RBD
RBD in up position
B
ACE2
Steric Clash
RBD
ACE2

Figure 3.6 : Solvent and ACE2 accessibility of S protein RBD binding surface. (A) Close contact water molecules at the ACE2 binding surface of RBD of protomer B are shown for the closed state (upper panel) and open state (lower panel) in ice-blue colored licorice representation. (B) ACE2 binding pose is shown on RBD in the down and up conformation of protomer B. The steric clash zone is highlighted with circles. The amino acids involved in RBD-ACE2 interactions on RBD are shown with red beads.

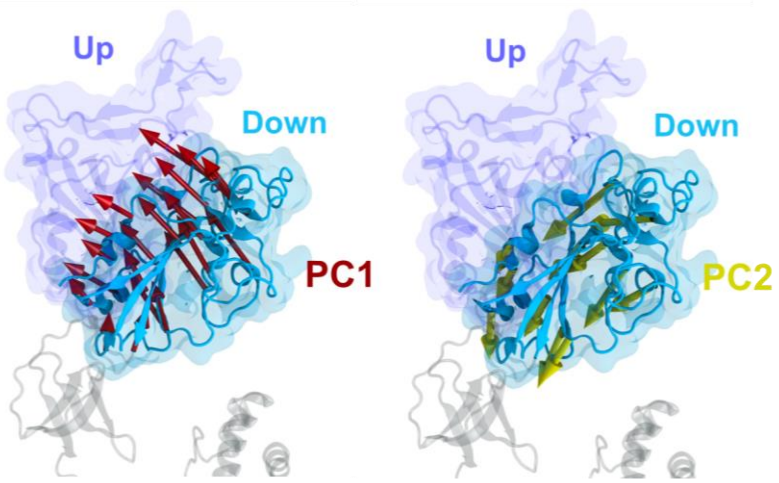
Up
Down
PC1
Up
Down
PC2

Figure 3.7 : The first two PC obtained from the MD simulations. PC1 and PC2 are shown with red and yellow arrows, respectively.

mode that moves RBD inward and outward with a rigid body-like motion and this motion is alone describes 93.3% of the structural transition between down and up positions. Also, PC2 is predominantly characterized by the RBD movement. Sampled conformations in simulations were projected onto PC1 and PC2. According to the occupancy of these grids, distributions for the projections $f(\mathbf{R})$ were calculated. Higher positive and lower negative values along PC1 demonstrates an increased level of RBD opening and closure, respectively. Using PC1 and PC2 distributions along with them, free energy surfaces were calculated as $A(\mathbf{R}) = -kTln\left(f(\mathbf{R})\right) + \text{constant}$ (Figure 3.8) (Gur, M et al., 2013). The free energy surface indicates three energy wells for up and down conformations and an additional intermediate state. The region between down and the intermediate states was not sampled by the MD simulations, probably due to the inability of MD simulations to simulate global transitions between these two states in the absence of any bias at the performed simulations lengths. To sample region between the down and up states on the energy surface and explore the minimum free energy pathway connecting these states, MD simulations starting from unpopulated regions between the down and up states were performed. SMD simulations were performed starting from conformations sampled in the closed state and SMD simulations were continued with a new set of MD simulations. RBD of protomer B was steered by performing SMD simulations towards its down-to-up position (down→up) or vice versa (up→down). Further, three different conformers sampled in the closed and open state MD simulations were used as starting conformers to perform down→up and up→down SMD simulations, respectively. Total of 24 MD simulations were performed initiating from S protein conformations sampled from SMD simulations at 9 Å, 11 Å, 13 Å, 15 Å, and 17 Å along the down→up SMD pulling direction and 3 Å, 5 Å and 7 Å along the up→down SMD pulling direction. To construct the free energy surface, MD simulations trajectory totaling of 5100 ns was projected onto PC1 and PC2 (Figure 3.8). Based on the free energy surface, a new semi-open state located halfway across the transition between the down and up positions of S protein RBD, and few additional substates at various locations were revealed. The semi-open intermediate state is separated from the down and up states by energy barriers of ~3kT and ~2kT, respectively. The energy barrier connecting the down and semi-open states provided a passage lowering the barrier to ~2kT. Besides, a small passage between the semi-open and up states without lowering the energy required to pass over the energy barrier was identified. Therefore, RBD might visit a long-lived semi-open intermediate state, before it transitions to an open state, and this intermediate state probably lowers the energy barrier between closed and open states.

3.1.4 Transition mechanism between down and up states of S protein protomers

The minimum free energy pathway was constructed by determining energy regions with low value on the landscape (Figure 3.8). For the transition between up and semi- open states, the crossing point is selected from the mild passage between the

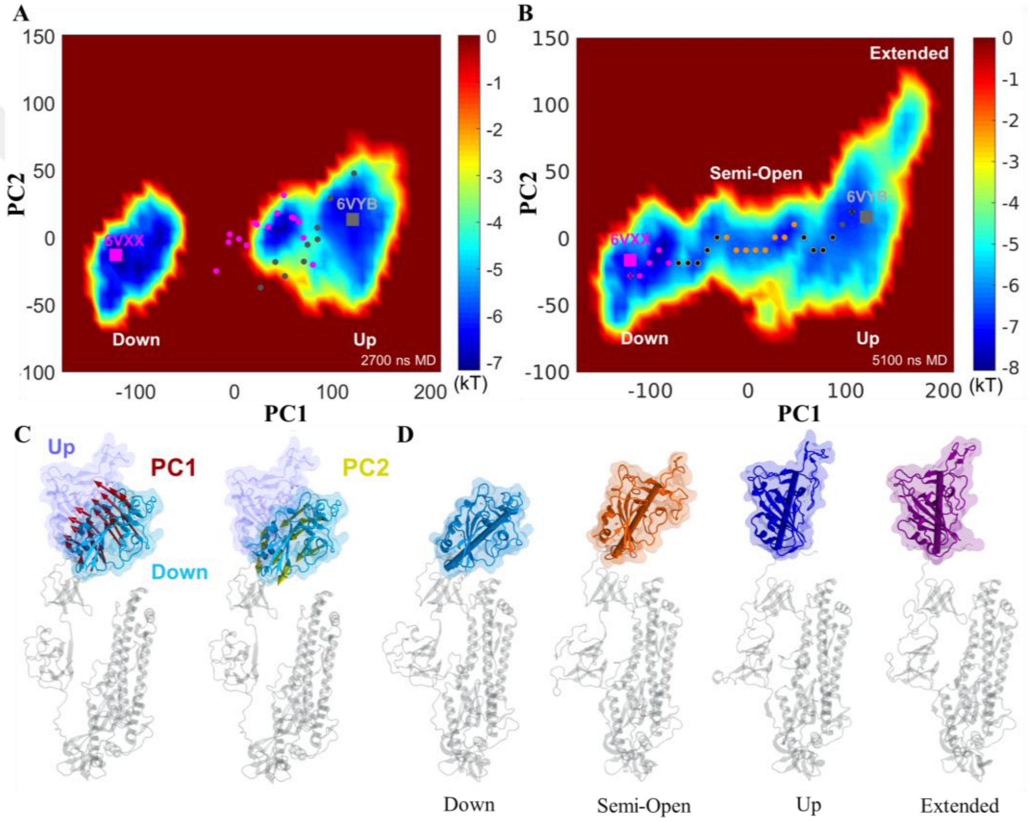

Figure 3.8 : The first two PCs and the free energy landscapes derived from MD simulations. (A) Energy landscape obtained from the closed and open state MD simulations trajectories for S protein protomers. Crystal structures of the closed (PDB ID:6VXX) and open (PDB ID:6VYB) are shown with magenta and gray squares. The initial conformers obtained from down to up and up to down SMD simulations indicated with gray and magenta dots, respectively. (B) Energy landscape obtained using MD and SMD simulations trajectories. The minimum free energy pathway is shown with dots. Magenta, orange, and gray dots correspond to down, semi-open, and up conformations. (C) The first two PCs of the MD simulations superimposed on protomer B structure. (D) Down, semi-open, up, and extended S protein protomers are shown. NTD of protomer is not shown.

up and semi-open states since it was the shortest path connecting the low energy regions of the semi-open and up states. To investigate salt bridge formation and breakage contribution to the transition between two states, the pathway divided into 24 bins and salt bridges localized inside each bin was calculated. For each bin, the distribution of the distance between the salt bridge forming residue pairs and hydrogen bond-forming residue pairs were evaluated (Figure 3.9). As can be seen in distributions of Figure 3.4 and 3.5, for the listed interactions the criterion of the salt bridge and hydrogen formation was not fully satisfied throughout the closed and open state trajectory, yet strong interactions were conserved at slightly higher distances and salt bridge and hydrogen bond formation was observed frequently. Thus, if conformers inside a bin are forming salt bridges frequently and otherwise strong interactions are preserved for the amino acid pairs, this bin is referred to as a salt bridge forming bin. Except for salt bridge R408-D405 and hydrogen bond E516-Y200, all interactions present in the down conformation of the S protein RBD were broken during the transition between down to the semi-open state. The formation of salt bridges and hydrogen bonds D428-R403 and D427-Y505 were only formed in the semi-open position toward the semi-open to open transition.

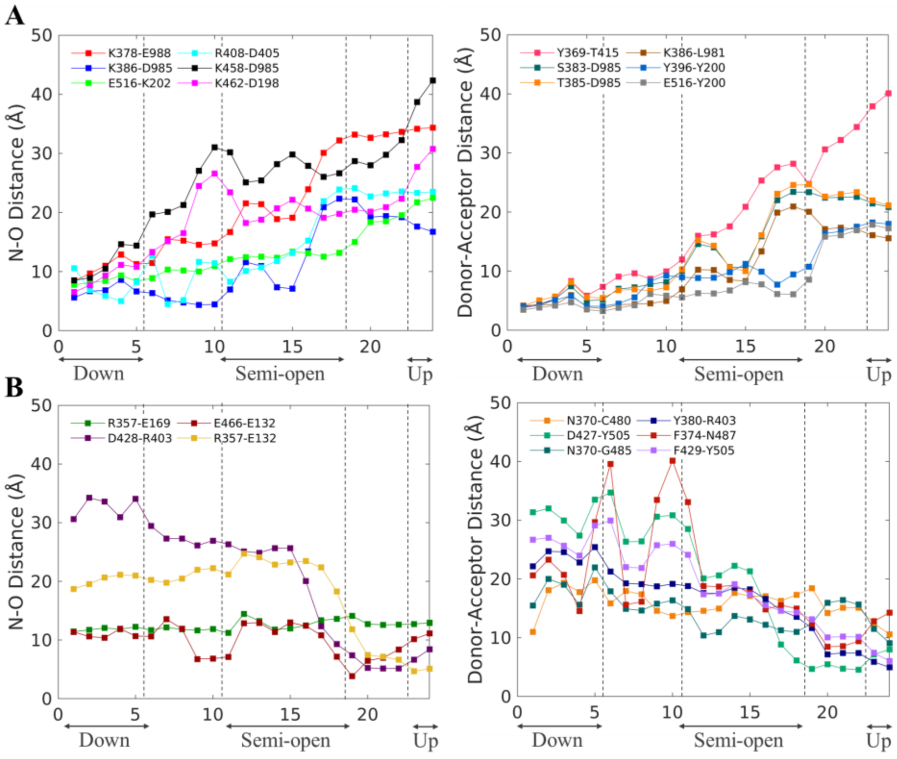

Figure 3.9 : Time evolution of salt bridges and hydrogen bonds during the transition between the (A) down and (B) up conformations. Each data point shows the average distance.

6. CONCLUSIONS AND RECOMMENDATIONS

Since its first appearance, more than 6,900,000 people infected and 400,000 killed because of the COVID-19 caused by SARS-CoV-2. An urgent need to develop vaccines and drugs has arisen and currently, there isn't any proven treatment to fight against COVID-19. As a promising target, the S proteins embedded on the viral membrane of SARS-CoV-2 emerged since they have an essential role in virus attachment and entry to the host cells. Crystal structures of S protein in its closed and open states were resolved in March 2020. These structures contain 77% of the protein sequence and almost the complete S protein structure. According to the crystal structures of the S protein, S protein RBD posesses different positions in the closed and open states, respectively. In this thesis, a broad range of MD simulations was performed to explore the binding mechanism of the S protein. Simulations demonstrate that while RBD is in its down position in the closed state, it has considerably lower mobility compared to RBD in its up position in the open state. The cause of decrease in the mobility has been investigated. Analysis of the interdomain interactions of the S protein RBD shows that fewer salt bridges and hydrogens bonds formed in RBD in its up position might be the reason behind the higher mobility of the RBD while its in up position. Therefore, stabilizing effect of higher number of interdomain interactions on RBD in down position might indicate that RBD in down position preserve its positions in the absence of ACE2. The free energy landscape has been generated and landscape has shown that there are multiple substates present beside the closed and open states. Energy landscapes based on the MD simulations showed the existence of an semi-open intermediate state for the S protein. This semi-open state shows a distinct interaction network from the down and up states. Superposition of the ACE2 bound RBD structure to the intermediate conformation shows no steric clash with the RBD of the neighboring protomer raising the possibility that ACE2 binding could take place in the semi-open intermediate state (Figure 3.10). The minimum energy pathway between down and up positions comprised a gradual salt bridge switching mechanism. Solvent accessbility of the ACE2 interaction surface of S protein RBD is conserved among all states. In terms of solvent accessibility, it would be possible to develop small molecules targeting RBD in down, semi-open intermediate and up states. Although the ACE2 interaction surface of it has solvent accessibility while it is in down position, superposition of the ACE2-bound RBD structures clearly show that it would be impossible for the S protein to bind ACE2 in its closed state. Thus, a potential inhibitor binding to the closed state would not need to compete with ACE2 for the binding interface. To conclude, this thesis provides a thermodynamical aspect of the dynamics and structure of the S protein RBD. Findings show that transition of the SARS-CoV-2 S protein RBD from the closed state to the semi-open intermediate state fluctuations could enable its binding to ACE2. Therefore, transition from down state to the semi-open state might be a critical step for the host cell recognition of SARS-CoV-2, and this semi-open intermediate state could be a potential target for drug binding.

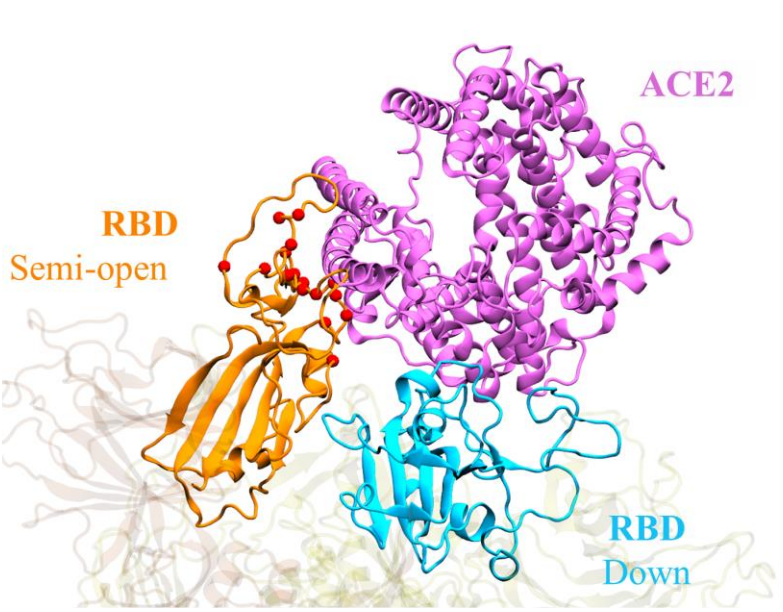

Figure 3.10 : Superposition of the ACE2 bound RBD structure onto an intermediate state conformation sampled from MD simulations. RBD amino acids involved in ACE2 interaction are shown with red beads.

REFERENCES

Aledo, J. C., Cantón, F. R., and Veredas, F. J. 2015. "Sulphur atoms from methionines interacting with aromatic residues are less prone to oxidation", Scientific reports, 5, 16955.

Baranowski, E., Ruiz-Jarabo, C. M., and Domingo, E. 2001. "Evolution of cell recognition by viruses", Science, 292(5519), 1102-1105.

Beckstein, O., Denning, E. J., Perilla, J. R. and Woolf, T. B. 2009. "Zipping and unzipping of adenylate kinase: atomistic insights into the ensemble of open↔ closed transitions." Journal of molecular biology 394(1), 160-176.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. 2000. "The protein data bank", Nucleic acids research, 28(1), 235-242.

Best, R. B., Zhu, X., Shim, J., Lopes, P. E., Mittal, J., Feig, M., and MacKerell Jr, A. D. 2012. "Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain $\chi 1$ and $\chi 2$ dihedral angles", Journal of chemical theory and computation, 8(9), 3257-3273.

Bosch, B. J., van der Zee, R., de Haan, C. A., and Rottier, P. J. 2003. "The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex", Journal of virology, 77(16), 8801-8811.

Bullough, P. A., Hughson, F. M., Skehel, J. J., and Wiley, D. C. 1994. "Structure of influenza haemagglutinin at the pH of membrane fusion", Nature, 371(6492), 37-43.

Burrell, L. M., Johnston, C. I., Tikellis, C., and Cooper, M. E. 2004. "ACE2, a new regulator of the renin–angiotensin system", Trends in Endocrinology & Metabolism, 15(4), 166-169.

Casalino, L., Gaieb, Z., Dommer, A. C., Harbison, A. M., Fogarty, C. A., Barros, E. P., Taylor, B. C., Fadda, E., and Amaro, R. E. 2020. "Shielding and Beyond: The Roles of Glycans in SARS-CoV-2 Spike Protein", bioRxiv.

Chan, J. F.-W., Yuan, S., Kok, K.-H., To, K. K.-W., Chu, H., Yang, J., Xing, F., Liu, J., Yip, C. C.-Y., and Poon, R. W.-S. 2020. "A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster", The Lancet, 395(10223), 514-523.

Consortium, U. 2019. "UniProt: a worldwide hub of protein knowledge", Nucleic acids research, 47(D1), D506-D515.

Coutard, B., Valle, C., de Lamballerie, X., Canard, B., Seidah, N., and Decroly, E. 2020. "The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade", Antiviral research, 176, 104742.

Delmas, B., and Laude, H. 1990. "Assembly of coronavirus spike protein into trimers and its role in epitope expression", Journal of virology, 64(11), 5367-5375.

Donoghue, M., Hsieh, F., Baronas, E., Godbout, K., Gosselin, M., Stagliano, N., Donovan, M., Woolf, B., Robison, K., and Jeyaseelan, R. 2000. "A novel angiotensin-converting enzyme–related carboxypeptidase (ACE2) converts angiotensin I to angiotensin 1-9", Circulation research, 87(5), e1-e9.

Drosten, C., Günther, S., Preiser, W., Van Der Werf, S., Brodt, H.-R., Becker, S., Rabenau, H., Panning, M., Kolesnikova, L., and Fouchier, R. A. 2003. "Identification of a novel coronavirus in patients with severe acute respiratory syndrome", New England Journal of Medicine, 348(20), 1967-1976.

Durrant, J. D. and McCammon, J. A. 2011. "HBonanza: a computer algorithm for molecular-dynamics-trajectory hydrogen-bond analysis." Journal of Molecular Graphics and Modelling 31, 5-9.

Eskici, G., and Gur, M. 2013. "Computational design of new peptide inhibitors for amyloid beta (Aβ) aggregation in Alzheimer's disease: application of a novel methodology", Plos one, 8(6).

Fehr, A. R., and Perlman, S. 2015. Coronaviruses: an overview of their replication and pathogenesis. In Coronaviruses (pp. 1-23): Springer.

Frenkel, D., and Smit, B. 2001. Understanding molecular simulation: from algorithms to applications (Vol. 1): Elsevier

Gallagher, T. M., and Buchmeier, M. J. 2001. "Coronavirus spike proteins in viral entry and pathogenesis", Virology, 279(2), 371-374.

Graham, R. L., and Baric, R. S. 2010. "Recombination, reservoirs, and the modular spike: mechanisms of coronavirus cross-species transmission", Journal of virology, 84(7), 3134-3146.

Guan, Y., Zheng, B., He, Y., Liu, X., Zhuang, Z., Cheung, C., Luo, S., Li, P., Zhang, L., and Guan, Y. 2003. "Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China", Science, 302(5643), 276-278.

Guang, C., Phillips, R. D., Jiang, B., and Milani, F. 2012. "Three key proteases–angiotensin-I-converting enzyme (ACE), ACE2 and renin–within and beyond the renin-angiotensin system", Archives of cardiovascular diseases, 105(6-7), 373-385.

Gur, M., Blackburn, E. A., Ning, J., Narayan, V., Ball, K. L., Walkinshaw, M. D., and Erman, B. 2018. "Molecular dynamics simulations of site point mutations in the TPR domain of cyclophilin 40 identify conformational states with distinct dynamic and enzymatic properties", The Journal of chemical physics, 148(14), 145101.

Gur, M., Zomot, E., and Bahar, I. 2013. "Global motions exhibited by proteins in micro-to milliseconds simulations concur with anisotropic network model predictions", The Journal of chemical physics, 139(12), 09B612_611.

Haagmans, B. L., Al Dhahiry, S. H., Reusken, C. B., Raj, V. S., Galiano, M., Myers, R., Godeke, G.-J., Jonges, M., Farag, E., and Diab, A. 2014. "Middle East respiratory syndrome coronavirus in dromedary camels: an outbreak investigation", The Lancet infectious diseases, 14(2), 140-145.

Haliloglu, T., Gul, A., and Erman, B. 2010. "Predicting important residues and interaction pathways in proteins using Gaussian Network Model: binding and stability of HLA proteins", PLoS computational biology, 6(7).

Haliloglu, T., Seyrek, E., and Erman, B. 2008. "Prediction of binding sites in receptor-ligand complexes with the Gaussian Network Model", Physical review letters, 100(22), 228102.

Han, Y., and Král, P. 2020. "Computational Design of ACE2-Based Peptide Inhibitors of SARS-CoV-2", ACS nano.

Han, Y., and Yang, H. 2020. "The transmission and diagnosis of 2019 novel coronavirus infection disease (COVID- 19): A Chinese perspective", Journal of Medical Virology.

Humphrey, W., Dalke, A., and Schulten, K. 1996. "VMD: visual molecular dynamics", Journal of molecular graphics, 14(1), 33-38.

Imai, Y., Kuba, K., Rao, S., Huan, Y., Guo, F., Guan, B., Yang, P., Sarao, R., Wada, T., and Leong-Poi, H. 2005. "Angiotensin-converting enzyme 2 protects from severe acute lung failure", nature, 436(7047), 112-116.

Isralewitz, B., Gao, M., and Schulten, K. 2001. "Steered molecular dynamics and mechanical functions of proteins", Current opinion in structural biology, 11(2), 224-230.

Izrailev, S., Crofts, A. R., Berry, E. A., and Schulten, K. 1999. "Steered molecular dynamics simulation of the Rieske subunit motion in the cytochrome bc1 complex", Biophysical journal, 77(4), 1753-1768.

Jarzynski, C. 1997a. "Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach", Physical Review E, 56(5), 5018.

Jarzynski, C. 1997b. "Nonequilibrium equality for free energy differences", Physical Review Letters, 78(14), 2690.

Jiang, S., Du, L., and Shi, Z. 2020. "An emerging coronavirus causing pneumonia outbreak in Wuhan, China: calling for developing therapeutic and prophylactic strategies", Emerging Microbes & Infections, 9(1), 275-277.

Kan, B., Wang, M., Jing, H., Xu, H., Jiang, X., Yan, M., Liang, W., Zheng, H., Wan, K., and Liu, Q. 2005. "Molecular evolution analysis and geographic investigation of severe acute respiratory syndrome coronavirus-like virus in palm civets at an animal market and on farms", Journal of virology, 79(18), 11892-11900.

Killerby, M. E., Biggs, H. M., Midgley, C. M., Gerber, S. I., and Watson, J. T. 2020. "Middle East respiratory syndrome coronavirus transmission", Emerging infectious diseases, 26(2), 191.

Kirchdoerfer, R. N., Wang, N., Pallesen, J., Wrapp, D., Turner, H. L., Cottrell, C. A., Corbett, K. S., Graham, B. S., McLellan, J. S., and Ward, A. B. 2018. "Stabilized coronavirus spikes are resistant to conformational changes induced by receptor recognition or proteolysis", Scientific reports, 8(1), 1-11.

Kuba, K., Imai, Y., Ohto-Nakanishi, T., and Penninger, J. M. 2010. "Trilogy of ACE2: A peptidase in the renin–angiotensin system, a SARS receptor, and a partner for amino acid transporters", Pharmacology & therapeutics, 128(1), 119-128.

Kuba, K., Imai, Y., Rao, S., Gao, H., Guo, F., Guan, B., Huan, Y., Yang, P., Zhang, Y., and Deng, W. 2005. "A crucial role of angiotensin converting enzyme 2 (ACE2) in SARS coronavirus–induced lung injury", Nature medicine, 11(8), 875-879.

Kufareva, I., and Abagyan, R. 2011. Methods of protein structure comparison. In Homology Modeling (pp. 231-257): Springer.

Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., Zhang, Q., Shi, X., Wang, Q., and Zhang, L. 2020. "Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor", Nature, 1-9.

Leach, A. R., & Leach, A. R. (2001). Molecular modelling: principles and applications. Pearson education.

Letko, M., Marzi, A., and Munster, V. 2020. "Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses", Nature microbiology, 1-8.

Li, F. 2008. "Structural analysis of major species barriers between humans and palm civets for severe acute respiratory syndrome coronavirus infections", Journal of virology, 82(14), 6984-6991.

Li, F. 2015. "Receptor recognition mechanisms of coronaviruses: a decade of structural studies", Journal of virology, 89(4), 1954-1964.

Li, F. 2016. "Structure, function, and evolution of coronavirus spike proteins", Annual review of virology, 3, 237-261.

Li, F., Li, W., Farzan, M., and Harrison, S. C. 2005. "Structure of SARS coronavirus spike receptor-binding domain complexed with receptor", Science, 309(5742), 1864-1868.

Li, W., Moore, M. J., Vasilieva, N., Sui, J., Wong, S. K., Berne, M. A., Somasundaran, M., Sullivan, J. L., Luzuriaga, K., and Greenough, T. C. 2003. "Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus", Nature, 426(6965), 450-454.

Li, W., Wong, S.-K., Li, F., Kuhn, J. H., Huang, I.-C., Choe, H., and Farzan, M. 2006. "Animal origins of the severe acute respiratory syndrome coronavirus: insight from ACE2-S-protein interactions", Journal of virology, 80(9), 4211-4219.

Lu, H., Isralewitz, B., Krammer, A., Vogel, V., and Schulten, K. 1998. "Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation", Biophysical journal, 75(2), 662-671.

Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., Wang, W., Song, H., Huang, B., and Zhu, N. 2020. "Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding", The Lancet, 395(10224), 565-574.

MacKerell Jr, A. D., Bashford, D., Bellott, M., Dunbrack Jr, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., and Ha, S. 1998. "All-atom empirical potential for molecular modeling and dynamics studies of proteins", The journal of physical chemistry B, 102(18), 3586-3616.

MacKerell Jr, A. D., Brooks, B., Brooks III, C. L., Nilsson, L., Roux, B., Won, Y., and Karplus, M. 2002. "CHARMM: the energy function and its parameterization", Encyclopedia of computational chemistry, 1.

MacKerell Jr, A. D. 2004. "Empirical force fields for biological macromolecules: overview and issues", Journal of computational chemistry, 25(13), 1584-1604.

Mallapaty, S. 2020. "Why does the coronavirus spread so easily between people?", nature, 579(7798), 183.

Memish, Z. A., Mishra, N., Olival, K. J., Fagbo, S. F., Kapoor, V., Epstein, J. H., AlHakeem, R., Durosinloun, A., Al Asmari, M., and Islam, A. 2013. "Middle East respiratory syndrome coronavirus in bats, Saudi Arabia", Emerging infectious diseases, 19(11), 1819.

Memish, Z. A., Perlman, S., Van Kerkhove, M. D., and Zumla, A. 2020. "Middle East respiratory syndrome", The Lancet.

Nomenclature, S. "Symbol Nomenclature for Glycans (SNFG)".

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., and Ako-Adjei, D. 2016. "Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation", Nucleic acids research, 44(D1), D733-D745.

Ouldridge, T. E. 2018. "The importance of thermodynamics for molecular systems, and the importance of molecular systems for thermodynamics", Natural computing, 17(1), 3-29.

Paraskevis, D., Kostaki, E. G., Magiorkinis, G., Panayiotakopoulos, G., Sourvinos, G., and Tsiodras, S. 2020. "Full-genome evolutionary analysis of the novel corona virus (2019-nCoV) rejects the hypothesis of emergence as a result of a recent recombination event", Infection, Genetics and Evolution, 79, 104212.

Park, J.-E., Li, K., Barlan, A., Fehr, A. R., Perlman, S., McCray, P. B., and Gallagher, T. 2016. "Proteolytic processing of Middle East respiratory syndrome coronavirus spikes expands virus tropism", Proceedings of the National Academy of Sciences, 113(43), 12262-12267.

Park, S., Khalili-Araghi, F., Tajkhorshid, E., and Schulten, K. 2003. "Free energy calculation from steered molecular dynamics simulations using Jarzynski’s equality", The Journal of chemical physics, 119(6), 3559-3566.

Park, S., and Schulten, K. 2004. "Calculating potentials of mean force from steered molecular dynamics simulations", The Journal of chemical physics, 120(13), 5946-5961.

Peiris, J., Guan, Y., and Yuen, K. 2004. "Severe acute respiratory syndrome", Nature medicine, 10(12), S88-S97.

Perlman, S., and Netland, J. 2009. "Coronaviruses post-SARS: update on replication and pathogenesis", Nature reviews microbiology, 7(6), 439-450.

Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L., and Schulten, K. 2005. "Scalable molecular dynamics with NAMD", Journal of computational chemistry, 26(16), 1781-1802.

Ponder, J. W., and Case, D. A. 2003. Force fields for protein simulations. In Advances in protein chemistry (Vol. 66, pp. 27-85): Elsevier.

Prabakaran, P., Xiao, X., and Dimitrov, D. S. 2004. "A model of the ACE2 structure and function as a SARS-CoV receptor", Biochemical and biophysical research communications, 314(1), 235-241.

Ryckaert, J.-P., Ciccotti, G., and Berendsen, H. J. 1977. "Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes", Journal of computational physics, 23(3), 327-341.

Saberi, A., Gulyaeva, A. A., Brubacher, J. L., Newmark, P. A., and Gorbalenya, A. E. 2018. "A planarian nidovirus expands the limits of RNA genome size", PLoS pathogens, 14(11).

Scott, W. R., Hünenberger, P. H., Tironi, I. G., Mark, A. E., Billeter, S. R., Fennen, J., Torda, A. E., Huber, T., Krüger, P., and van Gunsteren, W. F. 1999. "The GROMOS biomolecular simulation program package", The Journal of Physical Chemistry A, 103(19), 3596-3607.

Shang, J., Ye, G., Shi, K., Wan, Y., Luo, C., Aihara, H., Geng, Q., Auerbach, A., and Li, F. 2020. "Structural basis of receptor recognition by SARS-CoV-2", Nature, 1-8.

Smith, M., and Smith, J. C. 2020. "Repurposing therapeutics for COVID-19: supercomputer-based docking to the SARS-CoV-2 viral spike protein and viral spike protein-human ACE2 interface".

Tipnis, S. R., Hooper, N. M., Hyde, R., Karran, E., Christie, G., and Turner, A. J. 2000. "A human homolog of angiotensin-converting enzyme cloning and functional expression as a captopril-insensitive carboxypeptidase", Journal of Biological Chemistry, 275(43), 33238-33243.

Towler, P., Staker, B., Prasad, S. G., Menon, S., Tang, J., Parsons, T., Ryan, D., Fisher, M., Williams, D., and Dales, N. A. 2004. "ACE2 X-ray structures reveal a large hinge-bending motion important for inhibitor binding and catalysis", Journal of Biological Chemistry, 279(17), 17996-18007.

Varshney, A., Balkrishna, A., and Singh, J. 2020. "Withanone from Withania somnifera May Inhibit Novel Coronavirus (COVID-19) Entry by Disrupting Interactions between Viral S-Protein Receptor Binding Domain and Host ACE2 Receptor".

Walls, A. C., Park, Y.-J., Tortorici, M. A., Wall, A., McGuire, A. T., and Veesler, D. 2020. "Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein", Cell.

Walls, A. C., Tortorici, M. A., Snijder, J., Xiong, X., Bosch, B.-J., Rey, F. A., and Veesler, D. 2017. "Tectonic conformational changes of a coronavirus spike glycoprotein promote membrane fusion", Proceedings of the National Academy of Sciences, 114(42), 11157-11162.

Wang, M., Yan, M., Xu, H., Liang, W., Kan, B., Zheng, B., Chen, H., Zheng, H., Xu, Y., and Zhang, E. 2005. "SARS-CoV infection in a restaurant from palm civet", Emerging infectious diseases, 11(12), 1860.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., de Beer, T. A. P., Rempfer, C., and Bordoli, L. 2018. "SWISS-

MODEL: homology modelling of protein structures and complexes", Nucleic acids research, 46(W1), W296-W303.

Wilde, R. E., and Singh, S. 1998. Statistical mechanics: Fundamentals and modern applications: Wiley-Interscience.

Wong, S. K., Li, W., Moore, M. J., Choe, H., and Farzan, M. 2004. "A 193-amino acid fragment of the SARS coronavirus S protein efficiently binds angiotensin-converting enzyme 2", Journal of Biological Chemistry, 279(5), 3197-3201.

Woo, H., Park, S.-J., Choi, Y. K., Park, T., Tanveer, M., Cao, Y., Kern, N. R., Lee, J., Yeom, M. S., and Croll, T. 2020. "Modeling and Simulation of a Fully-glycosylated Full-length SARS-CoV-2 Spike Protein in a Viral Membrane", bioRxiv.

Woo, P. C., Huang, Y., Lau, S. K., and Yuen, K.-Y. 2010. "Coronavirus genomics and bioinformatics analysis", viruses, 2(8), 1804-1820.

Woo, P. C., Lau, S. K., Lam, C. S., Lau, C. C., Tsang, A. K., Lau, J. H., Bai, R., Teng, J. L., Tsang, C. C., and Wang, M. 2012. "Discovery of seven novel Mammalian and avian coronaviruses in the genus deltacoronavirus supports bat coronaviruses as the gene source of alphacoronavirus and betacoronavirus and avian coronaviruses as the gene source of gammacoronavirus and deltacoronavirus", Journal of virology, 86(7), 3995-4008.

World Health Organization, W. Data Accessed: June 12, 2020. Retrieved from https://www.who.int/emergencies/diseases/novel-coronavirus-2019

Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C.-L., Abiona, O., Graham, B. S., and McLellan, J. S. 2020. "Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation", Science, 367(6483), 1260-1263.

Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., Hu, Y., Tao, Z.-W., Tian, J.-H., and Pei, Y.-Y. 2020. "A new coronavirus associated with human respiratory disease in China", Nature, 579(7798), 265-269.

Yan, R., Zhang, Y., Li, Y., Xia, L., Guo, Y., and Zhou, Q. 2020. "Structural basis for the recognition of the SARS-CoV-2 by full-length human ACE2", Science.

Zhang, G., Pomplun, S., Loftis, A. R., Loas, A., and Pentelute, B. L. 2020. "The first-in-class peptide binder to the SARS-CoV-2 spike protein", BioRxiv.

Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., and Huang, C.-L. 2020a. "Discovery of a novel coronavirus associated with the recent pneumonia outbreak in humans and its potential bat origin", BioRxiv.

Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., and Huang, C.-L. 2020b. "A pneumonia outbreak associated with a new coronavirus of probable bat origin", nature, 579(7798), 270-273.

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., and Lu, R. 2020. "A novel coronavirus from patients with pneumonia in China, 2019", New England Journal of Medicine.

CURRICULUM VITAE

Name Surname : **Ceren Kılınç**

Place and Date of Birth : **Adana, 25.04.1994**

E-Mail : **c.cerenkilinc@gmail.com**

EDUCATION:

- **B.Sc.** : **2017**, Yıldız Technical University, Faculty of Science and Letters, Molecular Biology and Genetics
- **B.Sc.** : **2016-2017**, Wroclaw University of Science and Technology, Biochemistry – Erasmus Student Exchange Program
- **M.Sc.** : **Present**, Istanbul Technical University, Graduate School of Science, Engineering and Technology, Department of Molecular Biology-Genetics and Biotechnology

PROFESSIONAL EXPERIENCE AND REWARDS:

- **June 2020 – Present.** TÜBİTAK Trainee Researcher (TÜBİTAK-STAR) Scholarship Holder
- **January 2019 – Present**. Graduate Research Student in Istanbul Technical University Biomolecular Engineering Laboratory

PUBLICATIONS ON THE THESIS:

- **Gur, M., Taka, E., Yilmaz, S. Z., Kilinc, C., Aktas, U., & Golcuk, M.** 2020. “Conformational Transition of SARS-CoV-2 Spike Glycoprotein Between Its Closed and Open States.”, The Journal of Chemical Physics, 153(7).

PROJECTS:

- Kolon ve Meme Kanserlerinde Erken Uyarı Sistemi Oluşturulması ve Kişisel Sağlık Sistemi (E-Nabız) Veri Tabanına Entegrasyonu, TÜBİTAK project no. 318S129
- Exploring Binding and Fusion Mechanism of SARS-CoV-2 Spike Glycoprotein Using Molecular Dynamics Simulations, COVID-19 High-Performance Computing Consortium Research Awarded (USA)