Florida International University

FIU Digital Commons

6-21-2022

The Role of Conformational Changes in Viral and Bacterial Protein Functions

Md Lokman Hossen
Department of Physics, Florida International University, mhoss042@fiu.edu

Recommended Citation

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

THE ROLE OF CONFORMATIONAL CHANGES IN VIRAL AND BACTERIAL

PROTEIN FUNCTIONS

A dissertation submitted in partial fulfillment of

the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

PHYSICS

by

Md Lokman Hossen

2022

To: Dean Michael R. Heithaus
       College of Arts, Sciences and Education

This dissertation, written by Md Lokman Hossen, and entitled The Role of Conformational Changes in Viral and Bacterial Protein Functions, having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this dissertation and recommend that it be approved.

_____
Jorge L. Rodriguez

_____
Yuan Liu

_____
Bernard S. Gerstman, Co-Major Professor

_____
Prem P. Chapagain, Co-Major Professor

Date of Defense: June 21, 2022

The dissertation of Md Lokman Hossen is approved.

_____
Dean Michael R. Heithaus
College of Arts, Sciences and Education

_____
Andrés G. Gil
Vice President for Research and Economic Development
and Dean of the University Graduate School

Florida International University, 2022

DEDICATION

This dissertation is dedicated to my sister, Tahmina Salam, who means a lot to me. I am truly thankful to the Almighty for having her in my life. This work is also dedicated to my parents Mst Halima Begum and Abdus Salam, my wife Most Khadija Nahar, and lovely daughter Benoyee Hossen.

ACKNOWLEDGMENTS

ABSTRACT OF THE DISSERTATION

THE ROLE OF CONFORMATIONAL CHANGES IN VIRAL AND BACTERIAL

PROTEIN FUNCTIONS

by

Md Lokman Hossen

Florida International University, 2022

Miami, Florida

Professor Prem P. Chapagain, Co-Major Professor

Professor Bernard S. Gerstman, Co-Major Professor

Proteins are the machinery of life that allow a living system to maintain a low-entropy, non-equilibrium state. The functioning of many proteins relies on a protein's ability to change its conformational shape. The structural changes can be triggered by atomic or molecular binding with other molecules or by mutations in the amino acid sequence of the protein. In this dissertation, I investigate the molecular details of the conformational changes in a SARS-CoV-2 viral protein called the spike protein, and an archaea cellular protein called PfMATE and explored how conformational changes affect the functioning of these proteins. I also performed molecular docking-based drug screening targeting the corona virus E protein to create a set of drugs that can potentially be repurposed.

Recently, the Omicron variant of SARS-CoV-2 has undergone additional mutations and shown unparalleled transmissibility compared to other strains of the virus. I computationally investigated the role of mutations in the Omicron spike protein that binds the virus to the host cell's ACE2 protein. My results suggest that, compared to the WT and

Delta variant, the mutations in the Omicron spike protein facilitate a more efficient conversion between "down" and "up" protein conformations as well as more efficient attachment to ACE2. These effects, combined with conformational changes that allow antibody evasion, may have contributed to Omicron's dominance over Delta.

In the PfMATE protein project, I performed molecular dynamics (MD) simulations to investigate the flexibility of five different PfMATE structures. Various analysis techniques based on MD computations provide information on how protonation of specific amino acids or the presence of Na+ can induce cascading structural changes responsible for the transition between inward and outward facing configurations. These structural transitions allow the PfMATE to expel molecules, which can produce drug-resistance by allowing the expelling of antibiotic molecules before they perform the desired damage to the archaea cell.

The E protein of SARS-CoV-2 plays an essential role in assembling the virus, mediating the budding process, and releasing the progeny viruses into the host cells. I examined US Food and Drug Administration (FDA) approved and investigational drugs and used them to target the E protein to obtain drug-protein complexes using molecular docking. The top six complexes were selected based on the docking score and embedded in the ERGIC membrane and relaxed with unconstrained MD simulations to investigate their stability. Their pharmacological properties were also predicted. The top-scoring, most stably bound, and clinically safe compounds are proposed as potential candidates for drug repurposing.

ABBREVIATIONS AND ACRONYMS

CHARMM   Chemistry at Harvard Molecular Mechanics

LJ   Lennard-Jones

MD   Molecular Dynamics

NAMD   Nanoscale Molecular Dynamics

PDB   Protein Data Bank

RMSD   Root Mean Square Deviation

RMSF   Root Mean Square Fluctuation

MATE   Multidrug Transporter and Toxic Extrusion

MHC   Major Histocompatibility Complex

LD50   Lethal Dose 50

HIA   Human Intestinal Absorption

BBB   Blood Bain Barrier

%   Percentage

ns   Nanoseconds

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.   INTRODUCTION

1.1   Proteins and Their Conformational Changes

Proteins are the machinery of life that allow a living system to maintain a low-entropy, non-equilibrium state. Proteins vibrate, move, fold, and perform work. They are naturally occurring molecular machines and perform most of the tasks for a living cell[1,2]. A protein is a one-dimensional chain of amino acids and attain a particular three-dimensional folded shape through ionic and hydrogen bonds, hydrophobic and hydrophilic interactions, and disulfide bridges among amino acids. This three-dimensional shape of a protein is called its tertiary structure[3-7]. Based on the Plasma Proteome Database[8], there are about 20,000 proteins known to exist in the human body. Each of them has a specific shape, size, and function in the cell that determines the overall physiological condition of a human body. Proteins perform a variety of biochemical and biophysical functions: controlling growth and maintenance[9-12], giving structural rigidity to cells and tissues[13-15], performing biochemical reactions[16,17], producing antibodies to fight infections[18,19], acting as hormones and chemical messengers[20], maintaining pH[21-23], providing energy[24] etc.

Protein functions require biomolecular interactions that are highly dependent on protein structural conformations[25,26]. Some proteins perform their function with static structures as in the lock-and-key model[27] which does not require any structural changes, while allosteric transitions[28] are necessary for other proteins to function. The induced fit mechanism[29-31] is a versatile mechanism that requires conformational changes in the interaction process.

Though proteins perform all the essential functions to maintain the health of a cell, they are also responsible for infectious diseases[32-35] caused by pathogens. Viral proteins play a critical role in paving the way for viral material to enter the host cell and allowing the virus

1

to hijack the host cell's genetic machinery to make copies of the viral particles that are released throughout the host to infect other cells. For bacterial infections, proteins help attach to the host cell receptors, act as a toxin, offer drug resistance, etc.

Recently, the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) caused the coronavirus disease outbreak of 2019 (Covid-19) and brought about a global pandemic[36-38]. The molecular-level details of the proteins associated with the pathogenesis of the virus are of prime importance to identify the biological processes of the virus and develop vaccines and drugs that can target different proteins to inhibit the infection. Although successful vaccines are already on the market, the continuous mutations occurring in different viral proteins cause differences in the molecular dynamics, thereby changing the transmission intensity, disease severity, and possibly cause immune evasion[39-42]. Also, all the vaccines were made by concentrating on the corona virus spike protein that attaches to the human host cell protein Angiotensin Converting Enzyme 2 (hACE2) to mediate viral entry. Other viral proteins, like the envelope (E) protein, can also be a potential target for vaccines and drugs as it plays critical roles in acting like an ion channel and introducing host cell inflammsome[43,44]. It also takes part in the assembly, budding, and release of progeny viruses[45].

Like viruses, bacteria also use proteins to cause disease with toxins. In addition, bacterial membrane proteins that normally expel foreign molecules can also expel antibiotic molecules and thereby cause drug resistance in the bacterial cells. It's been reported that the antibiotic norfloxacin is affected by a Multidrug and Toxin Extrusion (MATE) protein found in single cell archaeon *Pyrococcus furiosus*[46]. Because of the ability to expel antibiotics before they can kill the cell, multidrug transporter proteins have received much

attention from researchers[46,47]. MATE transporter proteins are thought to function by switching their structural configuration between an inward-facing cone to an outward-facing cone. This structural change can allow a toxic molecule inside a cell to float into the inward facing configuration and float out of the cell after the MATE protein has switched to an outward-facing configuration. Thus, MATE proteins allow cells to extrude antibiotic molecules and develop resistance to a wide range of drugs, e.g., fluoroquinolones, anticancer agents, antibiotics, and some vitamins[48,49]. However, we still lack a molecular-level understanding of the underlying dynamics of the multidrug transport protein's ability to change shape[50].

1.2   SARS-CoV-2 Omicron Variant

Viral proteins allow the virus to bind to the host cell. Mutation induced conformational changes in viral proteins can allow the virus to bind more easily and also allow the virus to evade antibodies and vaccines. As the SARS-CoV-2 pandemic continues to spread worldwide, the virus continues to amass protein mutations that occasionally lead to increased virulence and immune escape. The SARS-CoV-2 Delta variant outpaced the Alpha and Beta variants, and recently the Omicron variant, also known as B.1.1.529, quickly took over. Consternation about Omicron put global health sectors on high alert due to its high transmissibility and resistance to existing therapeutic antibodies or those produced by vaccines and prior infections[51-53]. Its spread was so rapid that cases of B.1.1.529 were reported in over 90 countries in less than a month[39,54]. In just three weeks since its detection in the US, it became the most dominant variant (>70%), followed by the Delta (~30%)[55]. Early results indicated that the immunity developed after Omicron infection is able

3

to successfully neutralize Delta[56]. Therefore, there is a glimmer of hope that Omicron will be able to displace the more severe Delta and ultimately lessen the burden of COVID worldwide. However, new variants of Omicron, such as BA.1 and the "stealth" BA.2 have emerged, underscoring the need for caution and further research[54].

The virus attaches to a host cell through binding of its spike protein to the ACE2 protein on the surface of the host cell. The portion at the tip of the spike protein that attaches to ACE2 is called the Receptor Binding Domain (RBD). The RBD is also the target for antibodies. As shown in Fig. 1.1, variant strains of the virus occur when there are mutations in the amino acids constituting the RBD.
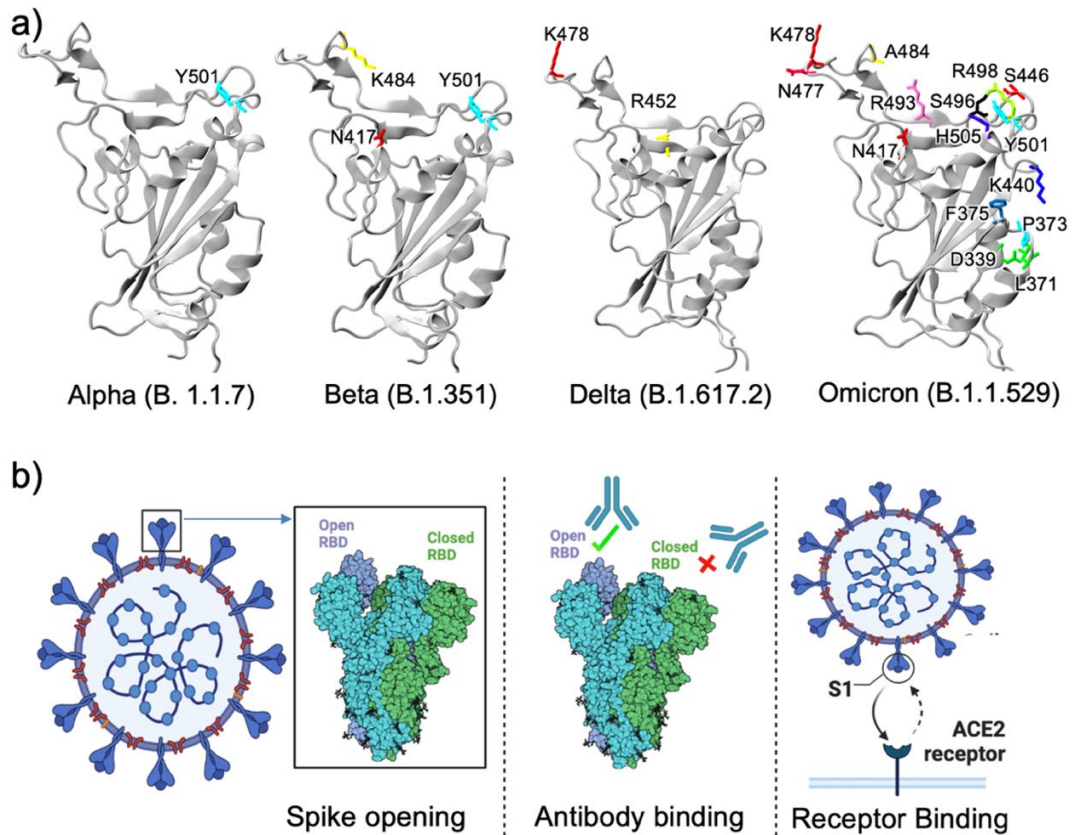
Figure 1.1 a) RBD structures of different variants. While Alpha, Beta, and Delta variants have less than three mutations in the RBD, Omicron has a remarkably large number of mutations. b) Three different ways RBD mutations may contribute to the high transmissibility of a variant (Created with BioRender.com).

Transmissibility of a respiratory viral infection such as with SARS-CoV-2 is a complex process involving a myriad of viral, host, and environmental factors[57] and there is a trade-off between transmissibility and virulence of SARS-CoV-2[58,59]. Considering only the RBD mutations, Omicron seems to have optimized its ability to infect in three different ways: 1) RBD "down" to "up" opening, 2) Antibody escape, and 3) ACE2 receptor binding. These three aspects of the RBD mutations are summarized in Figure 1.1(b).

1.3 Multidrug Transporters

Multidrug transporters are transmembrane proteins found in living cells. Transporter proteins are essential for regulating the molecules that travel in and out of a cell[60,61]. Transporters are able to selectively extrude toxic molecules from a cell. However, this ability allows disease-producing bacteria to expel antibiotics from cells before the antibiotics are able to perform their function, and thereby contributing to drug resistance in bacteria[62,63]. Multidrug resistant bacteria are a monumental problem because multidrug-resistant infectious diseases kill over two million lives and cost over $170 billion each year[64].

So far, seven families/superfamilies of multidrug transporter proteins have been discovered[50,60,65]. Among the seven, the family of multidrug and toxic compound extrusion (MATE) transporters are found in all three domains of life, e.g., eukarya, bacteria, and archaea[66,67]. The MATE transporters are divided into three phylogenetic branches based on similarities in their amino acid sequence: NorM, the DNA damage-inducible protein F (DinF), and the eukaryotic subfamilies[68]. The PfMATE protein found in archaea, is known

to undergo large scale conformational changes between an inward facing cone and an outward facing cone. Thus, PfMATE is an active transporter using electrochemical Na^+ or H^+ gradients as a source of Gibbs free energy to undergo conformational changes that allow the protein to extrude a broad spectrum of xenobiotics, metabolic organic cations (OCs), and antibiotics across the membrane[69-71]. By extruding antibiotic molecules, PfMATE helps bacteria to develop resistance to a wide range of drugs, e.g., fluoroquinolones, anticancer agents, antibiotics, and some vitamins[48,49]. Understanding the mechanisms of conformational changes of MATE transporter proteins may provide information to develop drugs that specifically target MATEs to prevent them from expelling antibiotics and, therefore reducing antibiotic resistance. As an important component of this problem, multidrug transporter proteins have received much attention from researchers[46,47], but we still lack a molecular-level understanding of the underlying mechanics of multidrug transport phenomena[50]. I investigated the molecular dynamics of conformational changes in the specific transporter protein, PfMATE.

2.   VIRAL AND BACTERIAL SYSTEMS STUDIED

2.1   SARS-CoV-2 Omicron Variant Spike Protein

Since the emergence of SARS-CoV-2 in late December of 2019 it is continuously mutating and changing its virulence. Sometimes this change makes situation worst as the Delta and Omicron variants showed recently. The change in the gene sequence causes the mutations in the amino acid sequence of the viral proteins and it occurs frequently in RNA viruses like Coronavirus. So far, five different major strains were identified and declared as Variant of Concerns by Centers for Disease Control and Prevention (CDC). Among these SARS-CoV-2 variants, Omicron was the most transmissible one with having 62 mutations in the total. The receptor binding domain (RBD) (residue 330-530) of the spike protein is an important part that mediates the attachment of the virus to the host cells. The large number of spike protein mutations sets the Omicron variant significantly apart from the other variants. Compared to the WT, the spike protein harbors more than 30 mutations, including 15 in the receptor binding domain (RBD) alone. The mutations in the RBD are - G339D, S371L, S373P, S375F, K417N, N440K, G446S, S477N, T478K, E484A, Q493R, G496S, Q498R, N501Y, and Y505H,[72] compared to only 501Y in B.1.1.7, 417N, 484K, and 501Y in B.1.351, and 478K and 452R in B.1.617.2 (Figure 2.1).
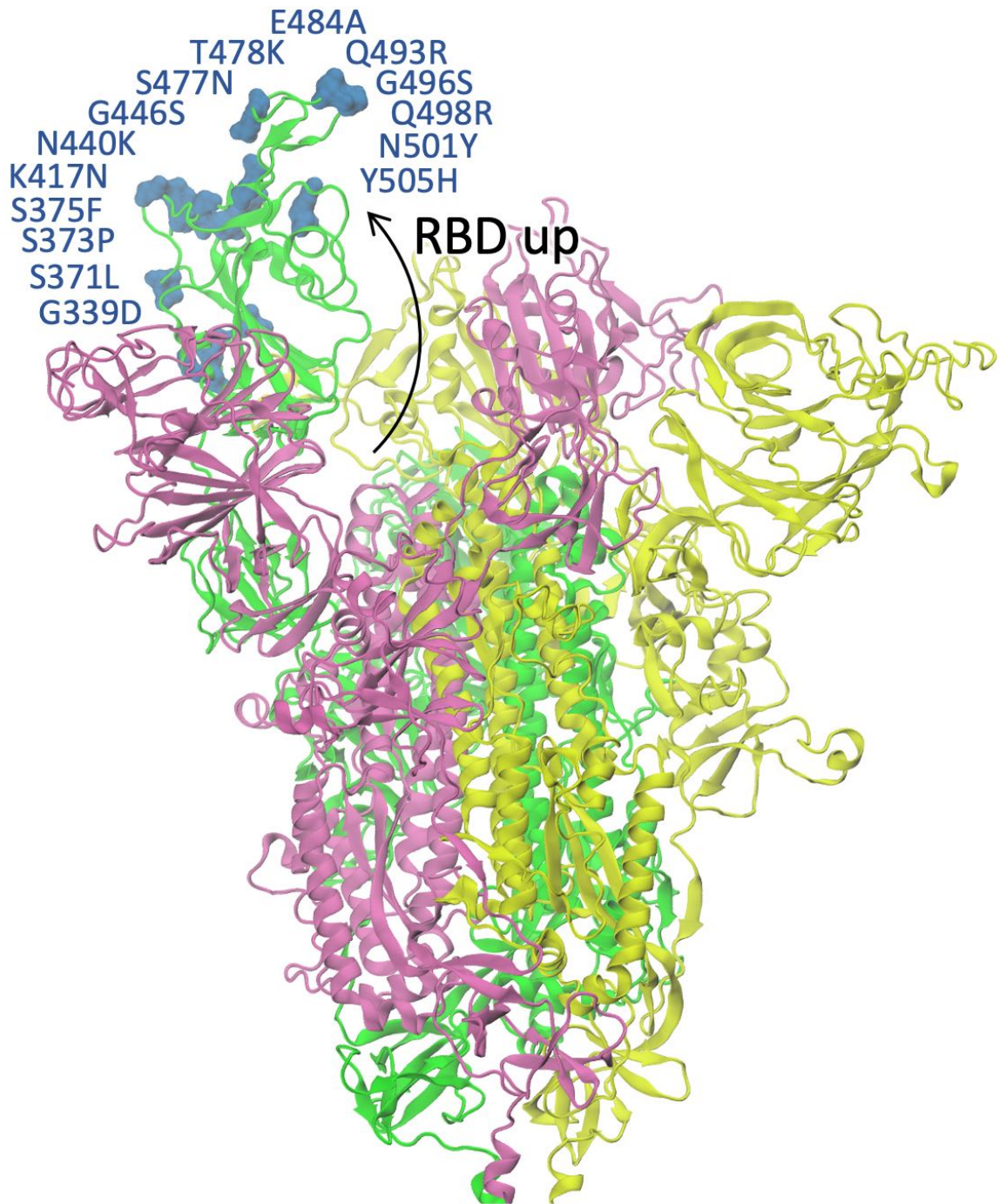
Figure 2.1 The head-only model of spike protein with one RBD in the "up" conformation with the associated mutations in the RBD of SARS-CoV-2 Omicron variant. The mutations are showed with surface representations in blue color.

Specific mutations can give a variant an edge on the fitness landscape. For example, the

P681R mutation in Delta is believed to have increased its transmissibility by enhancing the

spike protein cleavage[73]. In the RBD, previous studies show that the mutations Q498R and N501Y increase the affinity to bind with the human receptor ACE2[74], whereas mutations in the RBD loop region, e.g. E484K and others, are found to be associated with immune evasion[75].

2.2   PfMATE Protein

MATE proteins are thought to function by switching their structural configuration between an inward facing cone to an outward facing cone as shown in Figure 2.2(a,b) and 2.2(c,d), respectively. This can allow a toxic molecule inside a cell to float into the inward facing configuration, and then float out of the cell after the MATE protein has switched to an outward facing configuration. This configurational switching is fascinating from a molecular biophysics viewpoint because it requires a balancing of structural stability of either configuration (enthalpy) with the flexibility to switch configurations (entropy). Thus, MATE transporter proteins are a specific microcosm of the stability-flexibility, entropy-enthalpy tension that characterizes living systems. To understand the dynamics of the switching between outward-facing and inward-facing configurations requires knowledge of both structures, which until recently was not available for any MATE protein. Fortunately, the crystallographic structure of the inward-facing configuration of PfMATE has been determined[76]. The outward-facing configuration of PfMATE had been previously determined. PfMATE is a transmembrane protein grouped in the subfamily DinF and is extracted from the archaeon *Pyrococcus furiosus*. Based upon the structure reported by Lu *et al.*[77], the amino acids constituting each TM helix are the following: TM1: 17-46, TM2:

51-85, TM3: 95-106, TM4: 133-163, TM5: 166-188, TM6: 196-219, TM7: 239-269, TM8: 276-308, TM9: 316-348, TM10: 355-386, TM11: 391-411, TM12: 418-442.
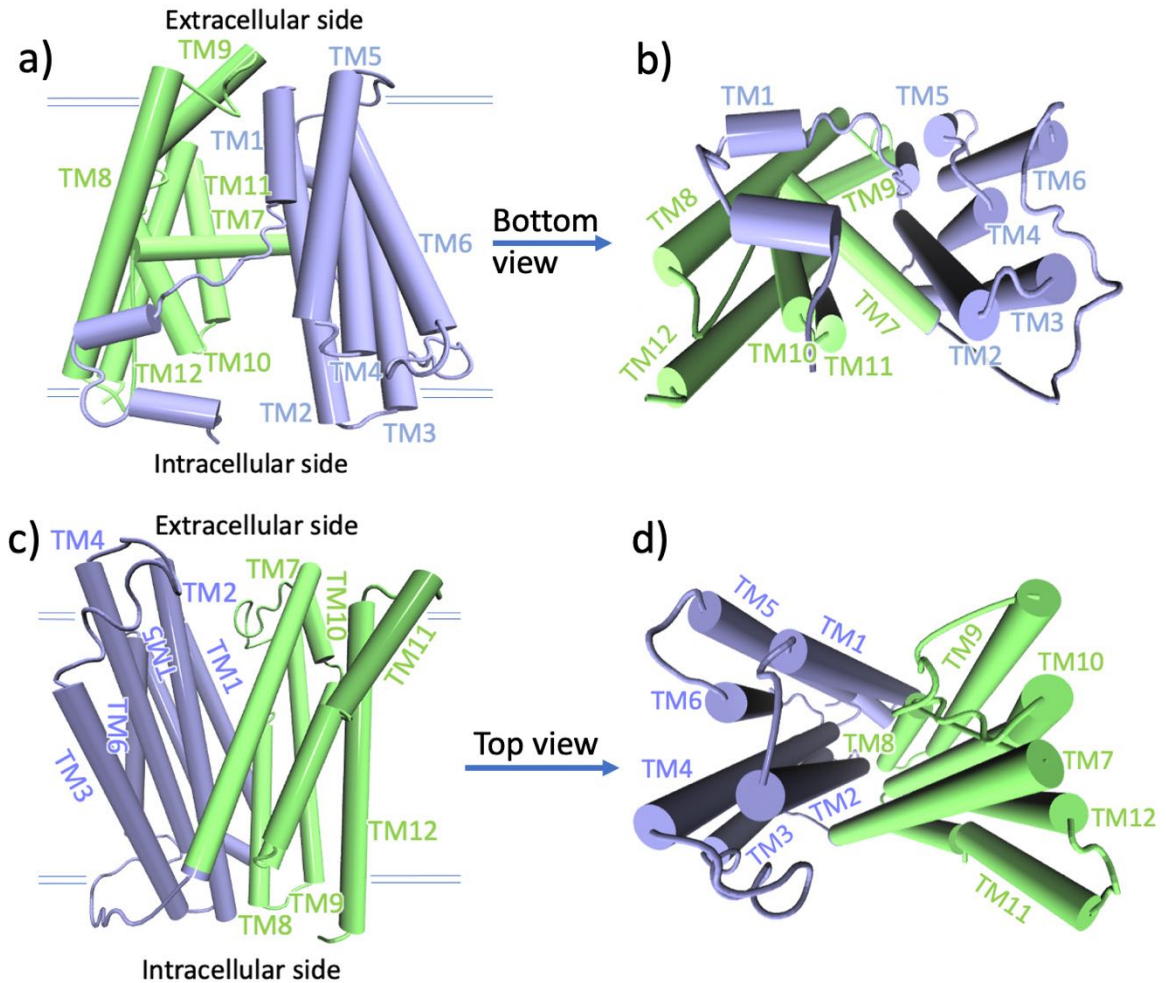
Figure 2.2 Structure of PfMATE. Helices are represented as cylinders. For IF: (a) side view, (b) bottom view of wide opening. For OF: (c) side view, and (d) top view of wide opening. The N-lobe (TM1-TM6) is colored in blue and the C-lobe (TM7-TM12) is colored in green.

Archaea are single-cell prokaryotes without a cell nucleus. Archaea organisms were originally classified as bacteria but are now a separate domain. Archaea differ from bacteria in a few ways, including the lipid molecules that constitute the membrane as described below. Archaea have the ability to live in extreme environments of temperature and chemical composition.

Investigation of the structural transition between inward-facing and outward-facing PfMATE configurations has been started experimentally and uncovered complicated topologies, but an understanding of the dynamics at the molecular level is lacking. There seem to be important interactions between the surrounding lipids, and protonation of specific amino acid residues may also be critical in facilitating the isomerization between the two configurations. I will use computational molecular dynamics (MD) simulations and analysis to investigate the atomic-level details of responsible for the configurational transitions of PfMATE, and the importance of protonation of different residues.

2.3   SARS-CoV-2 Envelope Protein

SARS-CoV-2 E protein is a homopentameric, integral, helical, and short protein with 75 amino acids (aa) which is short by one aa than SARS-CoV E protein[43,78-80]. Of the 75 residues, 1-13 and 35-75 are called are called topological domain and 14-34 are called transmembrane domain[81].

The E protein of SARS-CoV-2 has sequence similarity 94.7%, and identity 93.4% with the SARS-CoV E protein sequence(Accession: ABD75324.1)[82]. The amino acid sequence comparison of E protein of SARS-CoV-2 and SARS-CoV is given in **Figure 2.3**.

```
SARS-CoV-2    1 mysfvseetgtlivnsvllflafvvfllvtlailtalrlcayccnivnvs    50
                ||||||||||||||||||||||·||||||||||||||||||||||||||||
SARS-CoV      1 mysfvseetgtlivnsvllffafvvfllvtlailtalrlcayccnivnvs    50

SARS-CoV-2   51 lvkpsfyvysrvknlnssr-vpdllv         75
                ||||:·|||||||||||||· ||||||
SARS-CoV     51 lvkptvyvysrvknlnssegvpdllv         76
```

Figure 2.3 E protein amino acid sequence comparison of SARS-CoV-2 and SARS-CoV.

The E protein plays a critical role in CoVs biology. Among the transmembrane proteins of SARS-CoV-2, E is the only viroporin and acts as a cation selective channel through ERGIC membrane. The E protein also releases the progeny viruses by taking part into the budding process. E protein expels out the Ca^{2+} ions from the ERGIC lumen and cause host cell NLRP3 inflammation[44]. The pathogenicity of CoVs depends on E protein and can be attenuated by mutating its residues[78,83,84]. In addition, knocking out E protein from CoVs form immature and less infectious progeny, and thereby reduce the number of viruses. in the host cell[84].

3.   MATERIALS AND METHODS

3.1   Systems Building for SARS-CoV-2 Omicron Project

For the RBD interactions with the surrounding domains in the closed-form trimer, the initial frame of the simulation trajectory of the full spike protein trimer from the COVID-19 Data Sets of Amaro lab[85] (https://amarolab.ucsd.edu/covid19.php). This fully glycosylated spike trimer (RBD-down) was prepared from the cryo-EM structure (pdb ID 6VXX[86]). Because a full trimer structure of the Omicron spike protein is unavailable, I introduced the Omicron-specific mutations to the WT using CHARMM-GUI webserver[87,88]. These Omicron mutations are: G339D, S371L, S373P, S375F, K417N, N440K, G446S, S477N, T478K, E484A, Q493R, G496S, Q498R, N501Y, Y505H in the RBD region (residues 330-530). I used these systems for investigating the RBD opening mechanism. For the Omicron RBD-ACE2 system, a recently deposited cryo-EM structure (PDB ID 7T9L)[89] was obtained from the Protein Data Bank. Similarly, the cryo-EM structure (PDB ID 7W9I)[90] was used for the Delta RBD-ACE2 simulations. The RBD-only system for Omicron was prepared with the RBD from the RBD-ACE2 complex (PDB ID 7T9L)[89]. The simulation system was set up with the same procedure as in our earlier studies[91,92]. All the systems were prepared using the solution builder interface of CHARMM-GUI website[87,88]. All the systems studied in this project are listed in Table 3.1.

Table 3.1 Molecular systems simulated in this work. The WT and Delta simulations of the RBD-only systems in our earlier work (Baral et. al, 2021)[39] were extended from 600ns to 1000ns. All systems include appropriate glycosylations: glycosylated spike residues: N165, N234, N343, and glycosylated ACE2 residues: N53, N90, N103, N322, N546.

System	Variant/ (PDB ID)	# Residues	# Water	# Ions	Total # Atoms in the System	Box Size (Å)^3	Simulation Time (ns)
RBD-only	WT (6VXX)	201	24,764	144	77,736	95x95x95	1000
	Delta (Mutated on 6VXX)	201	26,654	156	84,000	95x95x95	1000
	Omicron (7T9L)	201	26,723	157	83,518	95x95x95	1000
RBD-hACE2	WT (7A92)	796	79,743	273	253,548	140x140x140	100
	Delta (7W9I)	791	75,956	454	242,060	138x138x138	100
	Omicron (7T9L)	796	70,801	420	226,700	134x134x134	100
RBD-down*	WT (6VXX from Amaro Lab)	969	87,924	509	280,520	145x145x145	100
	Omicron (Mutated on 6VXX from Amaro Lab)	969	87,814	518	280,367	145x145x145	100

3.2   Systems Building for PfMATE Protein Project

Structures for the IF and OF conformations (PDB ID 6FHZ and 6GWH respectively) of the PfMATE transporter recently determined by Zakrzewska *et al.*[76], were obtained from the Protein Data Bank. The 17 missing residues in the N-lobe of the IF structure were predicted using the protein homology recognition web portal Phyre2[93]. The missing segment was inserted into the protein with VMD software[94]. For both the IF and OF conformations, I prepared two different systems based upon the experimental observations[76,95]; one protonated, and the other unprotonated. For the IF system, the following five residues were protonated: ASP41, ASP184, GLU163, GLU273, and GLU331. For the OF system, I protonated residues ASP184 and GLU163 in the N-lobe,

and GLU 331 in the C--lobe. In addition to these four systems, I also performed MD simulations on a fifth system, non-protonated IF with an Na$^+$ ion placed at its experimentally observed site[76] near Asp41, Asn180, ASP184, and Thr202, close to the extracellular side in the N-lobe region. The proteins were then embedded into the archaeal lipid bilayer as shown in Figure 3.1(a) and 3.1(b).
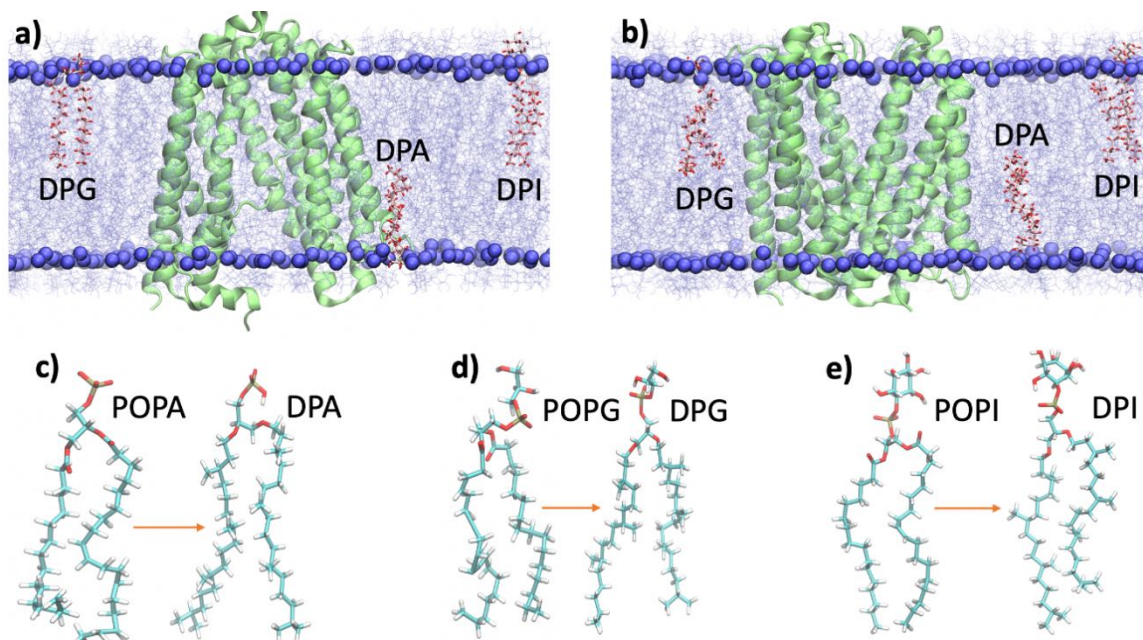
Figure 3.1 The initial setup of the membrane-protein system for: a) IF and b) OF PfMATE. The counterpart of the POPA, POPG and POPI lipids in the archaeon *Pyrococcus furiosus* are ether lipids: DPA, DPG and DPI respectively, obtained by changing the ester bond with ether bonds in the head group and tail junction and replacing one H with a -CH$_3$ in the tail, as depicted in (c), (d), and (e).

The PfMATE protein structures were embedded in a lipid bilayer using the Membrane Builder plugin of the CHARMM-GUI[87] webserver. For all five PfMATE systems, the membrane lipid composition was the same on both the inner and outer leaflets with a ratio of 35%:20%:45% of POPA, POPG and POPI, respectively. This lipid composition in the membrane allows archaea to cope with extreme thermal and acidic conditions[96,97]. Next, these systems were modified to mimic archaeal membranes. PfMATE is found in *Pyrococcus furiosus*, a hyperthermophilic archaeon. Unlike bacteria and eukaryotes,

archaeon cell membranes contain glycerol-ether lipids[98,99], and long isoprenoid tails with side chains[100] that make them capable of living in extreme environments, i.e., a wide range of temperature from 0°C to 100°C[101,102]. I replaced the ester linkage in POPA, POPG and POPE with an ether linkage and added -CH$_3$ branches in the tails to mimic the archaeal lipids diphytanyl phosphatidic acid (DPA), diphytanyl phosphatidyl N-acetyl hexose, diphytanyl phosphatidyl glycerol (DPG), and diphytanyl phosphatidyl inositol (DPI), respectively, as shown in Figures 3.1(c-e).

3.3   Systems Building for E-protein Drug-Screening Project

In this project I used recently solved crystal structure of SARS-CoV-2 E protein 7K3G[43] from RCSB Protein Data Bank website (rcsb.org)[103]. It has a total of 10 conformations, and I took the first structure to make the system. The structure contains only the transmembrane helices with residues 8-38, although the complete E protein of SARS-CoV-2 has total 75 amino acids in its genome sequence (Accession: QNH88662.1)[104]. The C-terminal of the structure was modelled using Target-Template Alignment tool of SWISS-MODEL webserver[105,106] targeting the crystal structure. Later, the transmembrane part of the complete modelled structure was overlapped with the transmembrane part of the crystal structure and the C-terminal part was cut and joined to the crystal structure using Visual Molecular Dynamics (VMD) software[94], thus I got a complete E protein for SARS-CoV-2 with the transmembrane part from crystal structure. For making a system of E protein embedded in the endoplasmic reticulum–Golgi intermediate compartment (ERGIC) membrane, I used CHARMM-GUI web-based membrane builder interface[87,107,108]. The ratio of the lipid composition for mimicking the ERGIC membrane was POPC : POPE : PI

: POPS : Chol = 45 : 20 : 13 : 7 : 15,[43] and the system contains 150 lipids on upper and lower leaflet each. The system was solvated with TIP3[109] water model and ionized with 150mM of $Ca^{2+}Cl^-$. The system was minimized and equilibrated with a six steps protocol generated by CHARMM-GUI for a total of 10,000 steps of 1 fs timestep and 1,115,000 steps of 2 fs timestep respectively, at 300 K temperature. A series of constrained production runs, with a decreased force constant applied to the heavy atoms, were performed to make the channel stable for a total of 100ns and was followed by an unconstrained production simulation for 1000 ns. All the molecular dynamics (MD) simulations were performed with NAMD2.14[110]. The transmembrane part of the E protein started intertwining from the straight form during the MD simulation as shown in Figure 3.2. Later, the structure was also retrieved using
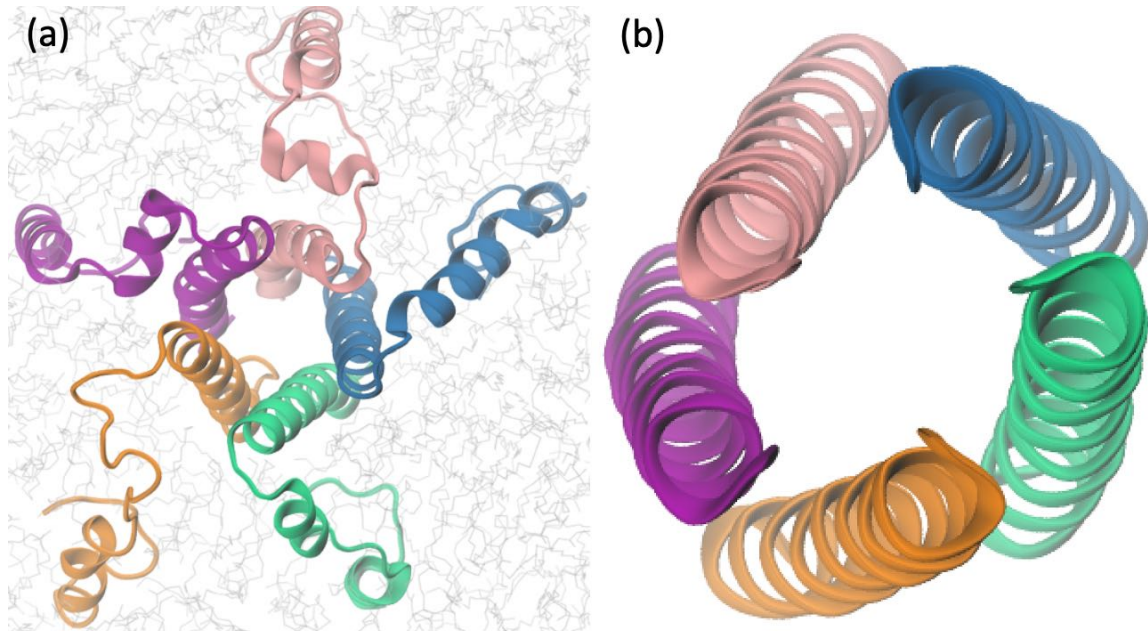
Figure 3.2 (a) SARS-CoV-2 E protein after 1000ns MD simulation in its helical form, (b) the similar kind of helical shape was also obtained by predicting the structure using AlphaFold.

DeepMind algorithm AlphaFold[111,112] (https://deepmind.com/) and found a similar pentameric E protein channels to the one I found in the MD simulation. The similarity of the two different E protein structures obtained in two different methods gave me confidence using the ensemble of the E protein in the drug screening.

3.4   Molecular Docking

The preparation of the ligands for docking was performed using same protocol as done by Pokhrel et al[113]. In short: the FDA-approved drugs library was downloaded from the web-interface e-Drugs3df[114]. The ligands were in SDF format and Open Babel[115] was used to convert them to pdbqt files. In this study, I am performing a screening against a total of 3800 FDA-approved and investigational drugs targeting E protein. An ensemble of 100 different conformations of the E protein was captured from the MD simulation trajectory. As all the protein structures were aligned, a common docking configuration box around them was made. The docking and screening was performed using AutoDock Vina 4.2[116]. The best docking hits were classified and graded based on the binding energy scores using custom scripts.

3.5   All Atom Molecular Dynamics (MD) Simulation for

3.5.1   Protein Only System

I performed molecular dynamics (MD) simulation following the procedures used in our previous work[92]. Briefly, constant pressure MD simulations were performed with NAMD 2.14[117] using CHARMM36 force-field[118,119]. Following 10,000-step minimization and 2 ns equilibrations, the production runs were performed at 303.15K temperature and 1 atm

pressure using 2fs timestep, and using Particle Mesh Ewald method to treat electrostatics[120,121] and SHAKE algorithm to constrain covalent bonds involving hydrogen atoms[122]. Visual Molecular Dynamics (VMD)[94] 1.93 was used for the structure and trajectory visualization, as well as for the hydrogen-bond analysis. Hydrogen bond analysis was performed for the RBD-ACE2 complexes for the 100 ns of the trajectories using a cut-off of 3.5 Å and 30°.

3.5.2   Membrane-Protein System

All-atom, explicit solvent molecular dynamics (MD) simulations were performed for the PfMATE-membrane systems with the CHARMM36 force field[123-126] using the GPU version of NAMD2.12 software[117]. The long-range ionic interactions were treated with the particle mesh Ewald method[120,121], where the nonbonded cutoff was set to 12 Å. The pressure and temperature were controlled using the Nose–Hoover Langevin-piston method with a piston period of 50 fs and a decay of 25 fs, and Langevin temperature coupling with a friction coefficient of 1 ps^{-1}, respectively[127,128]. Bond lengths for hydrogen atoms were fixed using the SHAKE algorithm[122]. The energy was minimized for 10,000 steps with 1 fs timesteps using the conjugate gradient and line search algorithm and was followed by a six-step protocol of minimization (10,000 steps of 1 fs each) and equilibration (1,115,000 steps of 2 fs each), at 303.15 K temperature. I performed MD simulations at the ambient temperature for archaeal of 373K. To prevent the MD simulations from crashing due to a large temperature change, after equilibration I first ran a 50 ns NPT (constant pressure and temperature) MD production run with 2 fs timesteps at 353 K. This was followed by another 100 ns of MD production run at 373 K. Output files were updated at 20 ps intervals.

3.6   Targeted Molecular Dynamics (TMD) Simulation

The structural conversions between OF and IF require a tremendous amount of time using standard molecular dynamics. To shorten the time so that I was able to investigate the dynamics of the structural transitions between the OF and IF conformations, I used Targeted Molecular Dynamics (TMD) computations, in which external forces are exerted on the C_α atoms to guide them to a target structure. The forces change during the TMD simulation and are determined from the gradient of a potential energy function:

$$U = (k/2N)[\text{rmsd}(t) - \text{RMSD}(t)]^2 \qquad (1)$$

where, rmsd(t) is the instantaneous root-mean-squared distance (rmsd) of the current coordinates with respect to the target coordinates, and RMS$D(t)$ is the assigned target rmsd, which decreases over time. For all five systems, the final RMSD was less than 1 Å.

TMD simulations were performed at 373 K for 40 ns for all five systems to transition each system from its initial configuration to the other configuration (OF→IF or IF→OF). The initial configuration used for the TMD runs was the configuration of the system at the end of the 100 ns 373K MD simulation. An elastic force constant per atom of $k_{\text{atom}} = k/N = 0.43$ kcal/(mol Å^2) was used in the TMD runs.

3.7   Potential of Mean Force (PMF) Simulation

The structural transitions between IF and OF configuration were further characterized by the determination of free energy landscapes from potential of mean force (PMF) calculations from the ensemble average of the TMD trajectories[129-133]. The TMD trajectories provided the reaction coordinates for the PMF calculation to determine how the free energy of a system varies with a reaction coordinate. The umbrella sampling

method was implemented along the RMSD reaction coordinate of the TMD trajectory at 0.5 Å increments in order to have at least 10 PMF points for the system that had the smallest change in RMSD. Each of the windows was then run for a short minimization and equilibration. An unconstrained MD simulation run was performed for each window for 2 ns at 373 K.

3.8   Dynamical Network Analysis

Dynamical network analysis (DNA) has been performed to examine the dynamical connections between amino acids in various protein structures such as Rfah, tRNA and protein kinase[134-136]. DNA examines MD trajectories to uncover groups of amino acids (communities) whose motions are highly correlated. The C_α of each amino acid is called a node, and the correlation values between the nodes are called weighted edges. The node pairs are connected with edges only when the centers of the nodes are within 4.5 Å for a minimum of 75% of the total simulation time. The correlation motions are used to create the time averaged connectivity of the nodes using the Girvan−Newman algorithm[137]. I performed DNA using the Carma[138] software package. The CatDCD software package was used to slice the MD trajectory into pieces to make it suitable for Carma.

3.9   Transfer Entropy

Transfer entropy (TE) is a method to identify communications between components of a system such as information flow between amino acids[134,139] during allosteric activity of proteins. The parts of a protein that drive the structural change (leaders) and the parts that respond (followers) during a transition can be measured using the directionality matrix

$T_{i\to j}(\tau)$. Here, i and j are the index of the driver and responder residues, respectively, and $\tau=n\tau_o$ is the time delay required for evaluating TE in units of τ_o[140,141] where τ_o is the time between frames that are output in the MD simulations (my τ_o=20 ps). I calculated TE as described by Michel in ref.[142] and used n=5. In this method, the TE was calculated using the variance-covariance matrix obtained from the change in position of each Cα atom during the MD trajectory using the ProDy Python package[143]. I performed the TE calculations based on the protocol reported by ref.[144]. I am interested in determining how information is transferred through the protein to trigger the structural change. The OF(np) configuration is stable, but protonation of the OF configuration seems to cause the OF◊IF transformation. Therefore, for comparison I calculated TE of amino acids in the stable OF(np) configurations to compare with the OF(p) configuration that will transform to IF. Likewise, IF(np) is stable, but the binding of Na$^+$ causes the IF◊OF transformation. For comparison, I calculated TE for amino acids in IF(np) and IF(Na$^+$). The TE is calculated from the trajectories of the 100 ns all-atom MD simulations for each of the four systems.

4.   SIGNIFICANCE OF THE RBD MUTATIONS IN THE SARS-COV-2 OMICRON: FROM SPIKE OPENING TO ANTIBODY ESCAPE AND CELL ATTACHMENT

This chapter was published in 2022: Hossen, M.L., Baral, P., Sharma, T., Gerstman, B. and Chapagain, P., 2022. Significance of the RBD mutations in the SARS-CoV-2 Omicron: From spike opening to antibody escape and cell attachment. Physical Chemistry Chemical Physics, 24(16), pp.9123-9129.

In this project I computationally investigated the role of the Omicron RBD mutations on its structure and interactions with surrounding domains in the spike trimer as well as with ACE2. My results suggest that, compared to the WT and Delta, the mutations in the Omicron RBD facilitate a more efficient RBD "down" to "up" conformation as well as ACE2 attachment. These effects, combined with antibody evasion, may have contributed to its dominance over Delta.

4.1   Effects of Mutations on the RBD Interactions in the Closed-form Trimer

The binding of the RBD with the ACE2 receptor requires the RBD to be in the "up" conformation from "down" in the prefusion state of the spike trimer. RBD's successful attachment to ACE2 is also facilitated by the spike protein's flexibility provided by the three hinges in the stalk domain (residues 1140-1234)[145,146], which is highly conserved in different variants. Several structures of the spike trimer solved with or without ACE2, including the structure with the three RBDs in the down and the structure with one RBD in the up-conformation[86,147-149], provide insights into the mechanism of the cell attachment of the SARS-CoV-2. The RBDs in the "down" conformation are held together by symmetrically arranged, centrally clustered,

interdomain hydrogen bonds[150], which break during the RBD "up" conformation. In order to calculate the inter-domain hydrogen bonds formed with a RBD in a closed-form spike trimer, I prepared a simulation system for a RBD surrounded by the interacting domains of the spike trimer using the full spike trimer structure[151]. To minimize the computational time, I prepared a truncated trimer including only the domains that directly interact with one of the RBDs (here, I chose the RBD of chain A), as shown in Figure 4.1 and Figure 4.2.

(a) RBD down

RBD down conformation

(b) RBD up

Mutations in Omicron RBD

G339D	T478K
S371L	E484A
S373P	Q493R
S375F	G496S
K417N	Q498R
N440K	N501Y
G446S	Y505H
S477N	

RBD up conformation

(c) Chain A, Chain C, Chain B

(d) Chain A, Chain C, Chain B

Chain A: 330-530
Chain B: 30-530, 968-1000
Chain C: 330-530, 968-1000
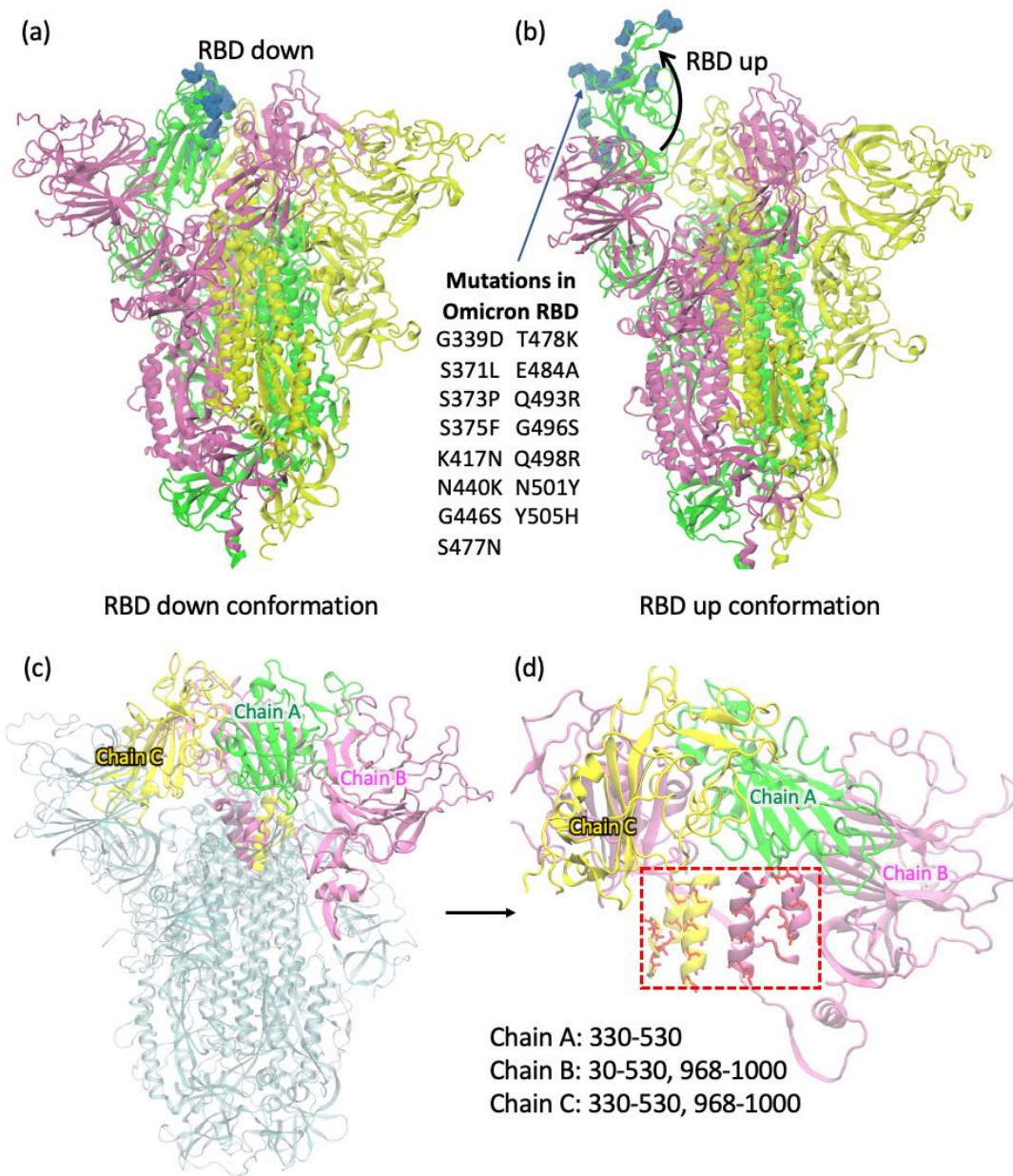
Figure 4.1 The spike protein trimer of SARS-CoV-2 in its prefusion state in the a) closed (RBD down) and b) open (RBD up) conformations. c) The closed form conformation of the spike protein highlighting the protein domains considered for the MD simulation. The trimer structure used here was the first frame of an MD trajectory obtained from the Amaro Lab[151] (trajectory of spike opening: https://amarolab.ucsd.edu/covid19.php). The RBD of chain A (green) and the surrounding domains (chain B – magenta and chain C – yellow) are highlighted. b) The colored part shows the truncated system consisting of the Chain A RBD and the surrounding domains considered for MD simulations. All C_α atoms >12 Å from the RBD of chain A are harmonically restrained for MD simulations. The helical segments highlighted in red dashed box are the non-contiguous segments of chain B and C that interact with the chain A RBD. For the Omicron system, all mutations within 12 Å of RBD of chain A and the surrounding were considered.

a) WT

b) Omicron

c) WT — Hydrogen-bond pairs and % occupancies

Chain A \	Chain B											Chain C	
	K113	Q115	D198	G232	D364	Y369	N370	F374	R983	D985	E988	N481	K986
A372												66	
S383									68	87	37		
T385								22					
Y421							59						
D427													38
R457					80	20							
K462				30									
R466						56							
I468		87											
S469	51												
E471	34												
Y505									87				

d) Omicron — Hydrogen-bond pairs and % occupancies

Chain A \	Chain B											Chain C		
	K113	Q115	D198	Y200	G232	D364	Y369	N370	R983	D985	E988	F456	G476	K986
N370												28	24	
A372													51	
S383									73	74	77			
T385										63				
Q414														45
G416						29								
N417									46					
D420										39				
Y421									40					
R457						93								
K462					56									
R466							77							
I468		69												
S469	24													
E516						55								

e)

f)

Figure 4.2 Major hydrogen bonds formed between the RBD of chain A (green) and the surrounding domains in the closed-form spike trimer for a) WT and b) Omicron. Additional interactions are shown in Figure 4.1 (from different views). Hydrogen-bond pairs and % occupancies for the c) WT and d) Omicron, with the color scale from red (maximum) to white (minimum). e) The locations of three glycans N165, N234, and N343 in the RBD of chain B that interact directly with the RBD of chain A. The RBDB glycans-RBDA hydrogen-bonds (red dotted lines) break to make way for the RBD opening. f) Histogram of the hydrogen-bonds made by RBDB glycans at N165, N234, and N343 with RBDA, though majority of the contribution comes from N165 and N343.

These domains include the residues 330-530 of chain A (RBD), 16-530 and 968-1000 of chain B, and 330-530 and 968-1000 of chain C. The domains surrounding the RBD of chain A are harmonically restrained by applying harmonic forces to all C_α atoms that are >12 Å away from the RBD of chain A. This allows flexibility of the RBD in the trimer but maintains the integrity of the domains mimicking the full trimer. I performed 100 ns simulations for the WT and Omicron and calculated the hydrogen bonds for the last 50 ns of the trajectories. To confirm that the RBD hydrogen bonding is adequately represented by the truncated system, I calculated and compared the % hydrogen bonds for the closed-form RBD to the simulation of the full system performed by the Amaro Lab[151]. The RBD hydrogen bonding pattern has good agreement between the truncated trimer and the full trimer.

I show in Figure 4.2 the major hydrogen bonds that an RBD forms with its surrounding domains for both the WT and Omicron. The % occupancies for the major hydrogen bonds for both the WT and Omicron are given in the matrices in Figure 4.2(c-d). Major hydrogen bond interactions with %occupancy >80% include R457(A)-D364(B), Y505(A)-F374(B), S383(A)-D985(B), and I468(A)-Q115(B). Residue S383 of the RBD makes three stable hydrogen bonds with the helix domain comprised of residues 968-1000 of chain B. However, only a few, weak interactions are observed between the RBD and the helix domain of chain C. The helix domains lie just below the RBD, as shown in Figure 4.2 and Figure 4.1. I also calculated the %hydrogen bonds (Figure 4.2(d)) for the Omicron RBD displayed in Figure 4.2(b). Most of the major hydrogen bonds found in the WT are also present in Omicron, including R457(A)-D364(B), S383(A)-R983(B), S383(A)-D985(B), S383(A)-E988(B), I468(A)-Q115(B), K462(A)-D198(B), and E516(A)-Y200(B). However, one of the major WT hydrogen bonds with nearly 90% occupancy, Y505(A)-

F374(B), is lost in Omicron due to the Y505H mutation. In addition, the polar to hydrophobic mutation S371L also abrogates minor hydrogen-bonding with multiple residues. When the Y505(A)-F374(B) hydrogen bond is broken, the RBD slightly repositions to make relatively weaker but new hydrogen bonds, mostly with Chain C residues. The new interactions in Omicron include N370(A)-F456(C), N370(A)-G476(C), A372(A)-G476(C), Q414(A)-K986(C), and D420(A)-E988(B). Although this loss of these hydrogen bond interactions in Omicron appears to be compensated by the increase in the %hydrogen bond or new hydrogen bonds. Therefore, it's difficult to assess the stability of the RBD based on the total %hydrogen bonds alone. However, the RBD opening from the closed form trimer may still be affected for the following reasons. With the Y505(A)-F374(B) hydrogen bond in the WT, the RBD of chain A is held in a position slightly away from these residues, which are in the RBD and the helix domain of chain C. These new interactions can still form in the WT if the Y505(A)-F374(B) hydrogen bond is broken and vice-versa. Therefore, my analysis shows that compared to the Omicron RBD, the WT RBD is more protected from opening from the closed-form trimer due to the possibility of either being held by Y505(A)-F374(B) or the additional interactions with chain C, whereas the Omicron RBD lacks the Y505(A)-F374(B) interaction. In addition to the interdomain residue interactions, my results suggest that the glycan interactions (Figure 4.2(e)) are also weaker in Omicron. Glycan-gating is shown to play a crucial role in the opening of the RBD to "up" conformation[151]. As shown in Figure 4.2(f), the chain B RBD glycans (mostly at N165 and N343) form 2.3±1.5 hydrogen bonds on average in Omicron vs. 5.5±1.9 in WT. I note that the interactions of RBD with the neighboring domains are transient and may change as the RBD shifts. While I investigated the interactions in the closed-form

state, further work is needed to determine exactly how these mutations affect the RBD opening along the opening pathway.

4.2   Structural Changes, Antibody Binding, and Antigenic Shift in the Omicron RBD

Once the RBD springs out from the "down" conformation, it is vulnerable to antibody detection and binding due to the loss of shielding by glycans[151]. Due to mutation-induced changes in both the residue type as well as the RBD structure, antibodies elicited with prior infections or vaccines may not be able to optimally bind to the RBD. In an earlier work, I showed that the changes in the Delta RBD structure, including in the receptor-binding motif (RBM) loop segment, cause some antibodies to be ineffective[39] at binding the RBD. With the significant number of mutations in the Omicron RBD, such effects can be extensive. To explore the structural changes in the Omicron RBD, I performed 1 μs simulation of the RBD-only system for Omicron (PDB ID 7T9L) and compared with that of WT[92] and I find significant differences in the RBD structure for the isolated RBD (i.e., not complexed with ACE2). Specifically, the motif consisting of residues 364 to 375, which contains the mutations S371L, S373P, and S375F, shows an extensive structural change (Figure 4.4(a)). All these three mutations are from polar to hydrophobic and this causes the motif to realign and make non-specific interactions with F342, A435, and W436 in the hydrophobic pocket (Figure 4.3).
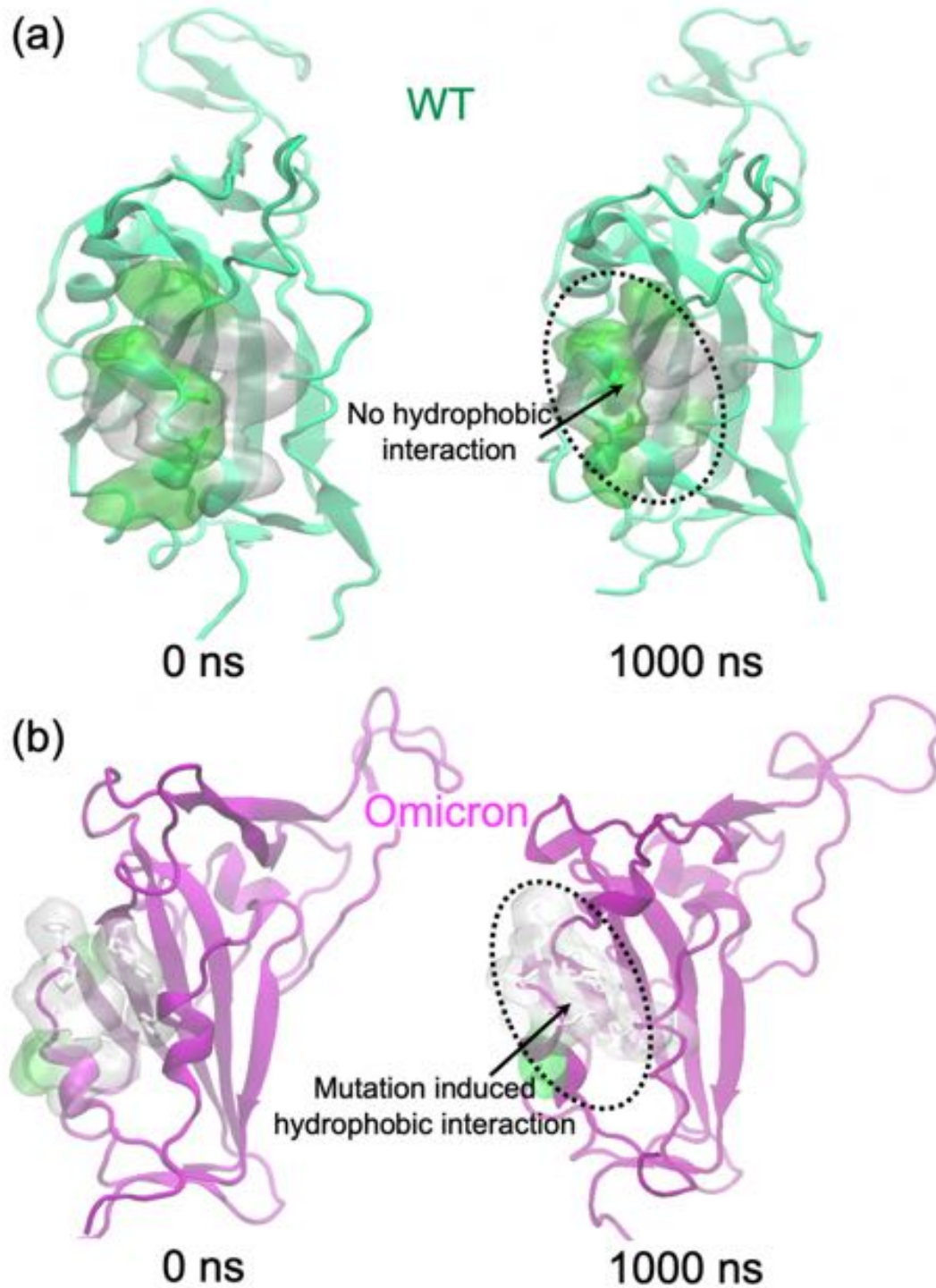
Figure 4.3. Comparison of the motif structure (residues 364-375) at 0 ns and 1000ns for the a) WT and b) Omicron RBDs. The polar to hydrophobic mutations S371L, S373P, S375F in Omicron allow interactions with the nearby hydrophobic residues. The polar residues are highlighted in green surface and the hydrophobic residues are highlighted in white/gray surface.

As shown in Figure 4.4(a), the distance between the C$_\alpha$ atoms of residues 371 and

375 in the motif is
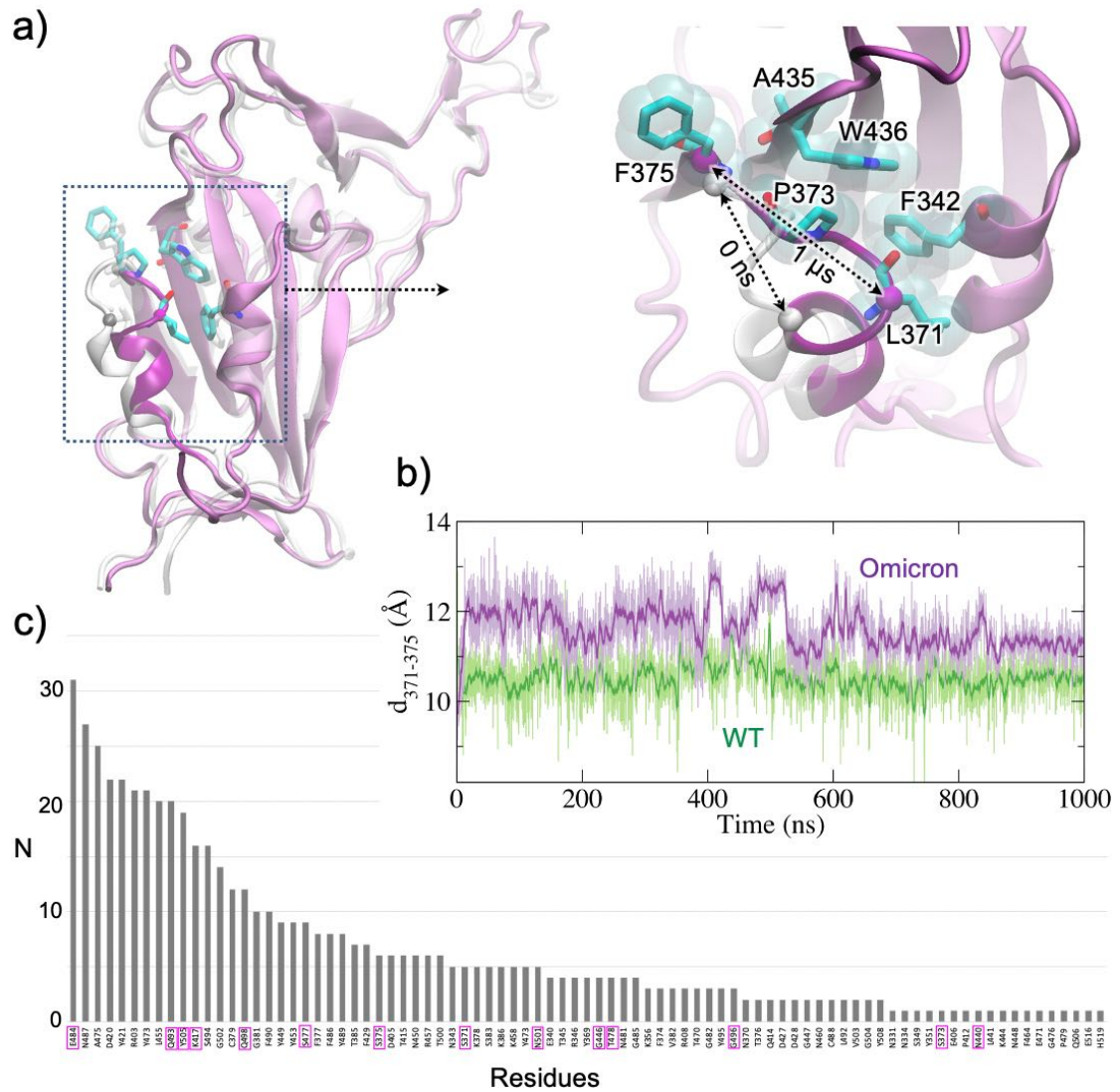
Figure 4.4 a) Structural changes in the Omicron RBD motif containing the mutations S371L, S373P, and S375F (motif highlighted in bright purple), which form a hydrophobic cluster (right). b) The Cα-Cα distance between residues 371 and 375 showing the difference in WT vs Omicron. c) Number of times the RBD residues found to hydrogen bond with the antibodies in 105 RBD-Ab complexes from the Protein Data Bank. The mutated residues in Omicron are highlighted in purple boxes along the x-axis.

relatively stable in WT, whereas it separates significantly in Omicron. Both of these

residues are binding sites for antibodies (e.g. RBD-Ab complexes 7KN5 and

7M7B). This separation in the antibody-binding region can reduce or abolish the binding of the antibodies specific to these sites. Subtle structure changes in other sites in the RBD may also affect antibody binding, allowing the Omicron RBD to escape antibody detection. Figure 4.4(c) shows the antibody-binding sites identified from the RBD-Antibody complexes available in the Protein Data Bank. Almost all mutations in the Omicron RBD are located in important antibody-binding sites (residues indicated by purple boxes) and therefore can directly affect the binding of antibodies specific to the WT and other variants.

To explore the antigenic shifts due to the mutations, I first identified the RBD epitopes using various MHC-I and MHC-II prediction methods as well as sequence and structure-based B-Cell epitope prediction methods as described in the Supporting Information and used a consensus approach[152] to select the epitopes for further analysis. The consensus epitopes that contain the mutations in RBD are given in the Supplementary Table 4.1.

Table 4.1 Predicted RBD epitopes that involve mutations in different variants of concern - WT, Alpha (B.1.1.7), Beta (B.1.351), Delta (B.1.617.2), Mu (B.1.621), and Omicron (B.1.1.529) and their corresponding antigenicity values. The mutations in the epitope are underlined. Significantly increased antigenicity in Omicron epitopes E2, E3, and E9 are highlighted in green and a moderately decreased antigenicity in E5 is highlighted in light red.

Epitope	Residues	Variant	Epitope Sequence	Antigenicity
E1	338-356	WT	FGEVFNATRFASVYAWNRK	0.24
		B.1.1.529	FDEVFNATRFASVYAWNRK	0.28
E2	370-378	WT	NSASFSTFK	0.12
		B.1.1.529	NLAPFFTFK	1.34
E3	372-380	WT	ASFSTFKCY	0.28
		B.1.1.529	APFFTFKCY	1.20
E4	417-425	WT	KIADYNYKL	1.66
		B.1.351	NIADYNYKL	1.55
		B.1.621	NIADYNYKL	1.55
		B.1.1.529	NIADYNYKL	1.55

E5	437-448	WT	NSNNLDSKVGGN	0.70
		B.1.1.529	NSN<u>K</u>LDSKV<u>S</u>GN	0.24
E6	439-451	WT	NNLDSKVGGNYNY	0.94
		B.1.1.529	N<u>K</u>LDSKV<u>S</u>GNYNY	0.69
E7	447-471	WT	GNYNYLYRLFRKSNLKPFERDISTE	0.16
		B.1.617.2	GNYNY<u>R</u>YRLFRKSNLKPFERDISTE	0.42
E8	455-478	WT	LFRKSNLKPFERDISTEIYQAGST	0.13
		B.1.617.2	LFRKSNLKPFERDISTEIYQAGS<u>K</u>	0.11
		B.1.1.529	LFRKSNLKPFERDISTEIYQAG<u>NK</u>	0.10
E9	483-493	WT	VEGFNCYFPLQ	0.56
		B.1.351	V<u>K</u>GFNCYFPLQ	0.60
		B.1.621	V<u>K</u>GFNCYFPLQ	0.60
		B.1.1.529	V<u>A</u>GFNCYFPL<u>R</u>	1.23
E10	491-505	WT	PLQSYGFQPTNGVGY	0.34
		B.1.1.7	PLQSYGFQPT<u>Y</u>GVGY	0.41
		B.1.351	PLQSYGFQPT<u>Y</u>GVGY	0.41
		B.1.1.529	PL<u>R</u>SY<u>S</u>F<u>R</u>PT<u>Y</u>GVG<u>H</u>	0.41
E11	502-510	WT	GVGYQPYRV	1.36
		B.1.1.529	GVG<u>H</u>QPYRV	1.01

I calculated the antigenicity of the mutated sequences using VaxiJen[153] and compared with the corresponding WT sequence to assess the antigenic characteristic introduced by the mutation. Most of the epitopes listed in Table 4.1 have similar antigenicity after mutations in Alpha, Beta, Delta, or Omicron. However, three epitopes in the Omicron (E2, E3, and E9 in Table 4.1) are found to have significantly increased antigenicity compared to the WT. These Omicron epitopes include 370-NLAPFFTFK-378 involving the mutations S371L, S373P, and S375F, 372-APFFTFKCY-380 involving the mutations S373P and S375F, and 483-VAGFNCYFPLR-493 involving the mutations E484A and Q493R. The antigenicity of E2 is 1.34 for Omicron vs 0.12 for WT. A similar increase is observed for epitope E3,

which has overlap with E2. Similarly, the antigenicity of E9 is 1.23 for Omicron vs 0.56

for WT. The locations of these epitopes are shown in Figure 4.5.
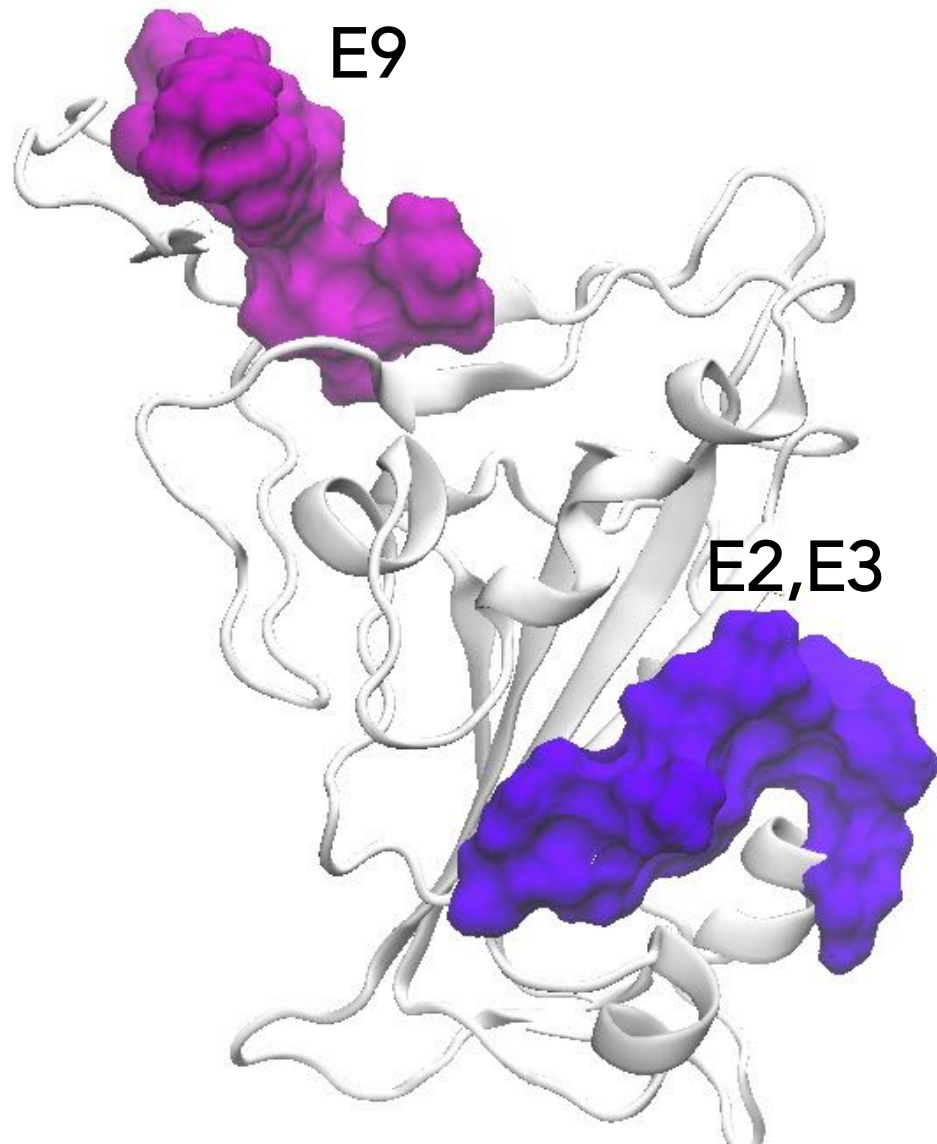
Figure 4.5 Location of the epitopes E2, E3, and E9 with increased antigenicity in the Omicron RBD. All epitopes have good surface accessibility for the antibody binding.

While the mutation-induced antigenic shifts render reduced sensitivity for the WT-specific

antibodies, the increased antigenicity in the Omicron epitopes suggests a more potent

immune response from these epitopes.

4.3   Effects of the Mutations on ACE2-Binding of the Omicron RBD

The mutations in the RBM region can directly affect the binding affinity of the RBD to bind ACE2. The RBD binding with ACE2 in different SARS-CoV-2 variants has been a topic of intense research[39,92,154-159]. Kim *et al* showed differences in the force required to dissociate the RBD from ACE2 for different variants of concern (Alpha, Beta, Gamma, and Delta)[160]. With 10 mutations in the RBM region of Omicron, the effects on ACE2 attachment can be significant compared to other variants. Early results by Wu et al. suggested that Omicron RBD-ACE2 interactions is weaker than in Delta[161]. However, recent cryo-EM structure-based analysis of the RBD-ACE2 complexes of both the Omicron and Delta variants show that the Omicron RBD-ACE2 interface has better optimized interactions compared to that of Delta[158].

Effort has been made to assess the binding affinity of RBD and ACE2 by launching online tool[162]. In this study, to investigate the effects of the Omicron mutations on the RBD attachment to the ACE2 receptor, I performed an MD simulation of the Omicron RBD-ACE2 complex (PDB ID: 7T9L)[163] and compared the RBD-ACE2 interfacial interactions with that of WT (PDB: 7A92)[164] and Delta (PDB ID: 7W9I)[90]. The short MD simulations provide the dynamic nature of the interactions and allow us to calculate the probability (% occupancy) of each hydrogen bond. As shown in Figure 4.6(a-d), the occupancy of the inter-protein hydrogen bond in Omicron is noticeably higher than in Delta suggesting a much stronger ACE2-binding in Omicron. The major RBD-ACE2 hydrogen-bond pairs in Omicron with >70% occupancy include Y453-H34, G502-K353, N487-Y83, T500-D355, and R493-E35, and H505-K353 and N477-S19 with >50% occupancy. In contrast, Delta has only two interactions (N487-Y83, G502-K353) with >70% occupancy and four

(K417-D30, Q493-E35, Y505-E37, and T500-D355) with >50% occupancy, suggesting a relatively weaker interfacial interactions in Delta compared to Omicron. This is consistent with a recent work by Genovese et al.[165] that analyzed the interfacial

a) Delta RBD vs ACE2

Delta RBD \ ACE2	Q24	D30	E35	E37	D38	Y41	Q42	Y83	K353	D355
K417		54								
Y449				35						
A475	22									
N487								77		
Q493			50							
Q498							25			
T500						34				54
G502									76	
Y505				60						

b) Omicron RBD vs ACE2

Omicron RBD \ ACE2	S19	H34	E35	Q42	Y83	K353	D355
Y453		72					
A475	48						
N477	54						
N487					86		
R493			79				
S496						26	
R498				25			
T500							76
G502						82	
H505						58	
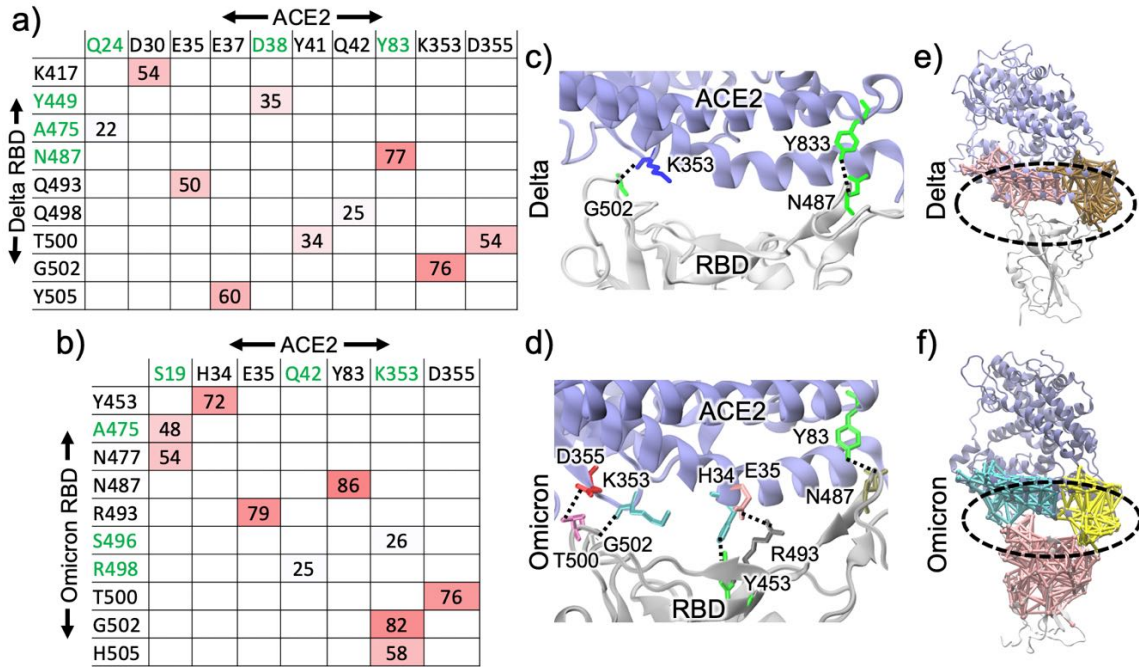
Figure 4.6 Percent occupancies of the hydrogen bonds between the RBD and ACE2 for the a) Delta and b) Omicron variants. The Unique interfacial hydrogen-bonds found in Omicron are highlighted in Green. Hydrogen bonds with >50% occupancy are shown for c) Delta and d) Omicron. The communities that span both the RBD and ACE2 are shown for the e) Delta and f) Omicron RBD-ACE2 complexes.

interactions in the RBD-ACE2 complex using ab-initio and quantum mechanical calculations. Lupala et al.[166] also found a significantly increased binding affinity of the Omicron RBD to ACE2, compared to the Delta RBD, suggesting an increase in infectivity. I compared the RBD-ACE2 hydrogen bonding with that in WT (Figure 4.7) and observed additional unique hydrogen bonds (e.g. A475-S19) in Omicron compared to WT. Since the presence of glycans at the interface can enhance the RBD-ACE2 interactions[167], I calculated the hydrogen bonding with glycan at N90

at the interface and found that Omicron has a relatively stronger interaction with 80% hydrogen bond occupancy in Omicron vs. 56% in Delta.

WT RBD	S19	D30	H34	E35	E37	Y41	Q42	K353	D355
K417		80							
Y453			31						
A475	21								
Q493				52					
Q498							50		
T500						55			26
G502								85	
Y505					71				

Delta RBD	Q24	D30	E35	E37	D38	Y41	Q42	Y83	K353	D355
K417		54								
Y449					35					
A475	22									
N487								77		
Q493			50							
Q498							25			
T500						34				54
G502									76	
Y505				60						

Omicron RBD	S19	H34	E35	Q42	Y83	K353	D355
Y453		72					
A475	48						
N477	54						
N487					86		
R493			79				
S496						26	
R498				25			
T500							76
G502						82	
H505						58	
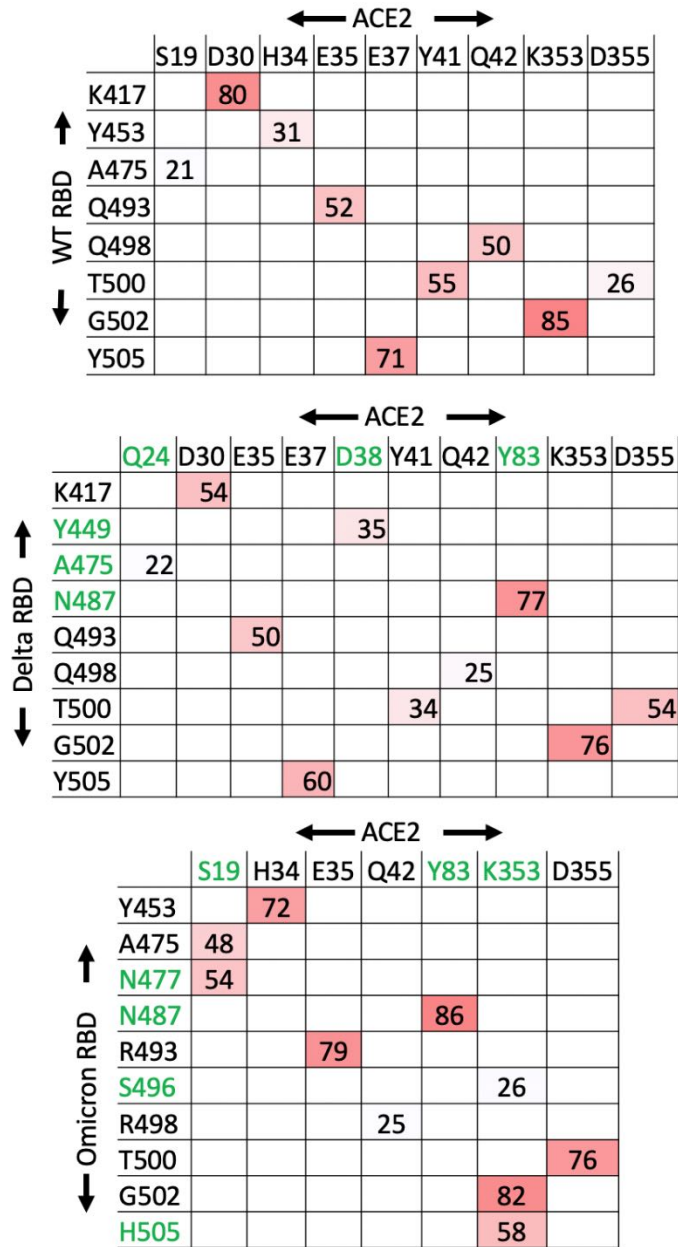
Figure 4.7 Hydrogen bond pairs for the interprotein interactions between the RBD and ACE2 for the WT, Delta, and Omicron. The matrices for the Delta and Omicron are the same as in Figure 4.6(a,b) but given here for comparison with WT. The matrix for residues participating in the unique hydrogen-bond pairs in Omicron are highlighted in green.

In addition to the increased hydrogen-bonding at the interface, the stability of the complex is also displayed by the dynamic network analysis[168]. I used the last 50 ns of the trajectory for RBD-ACE2 complexes of the Delta and Omicron variants to calculate the dynamic network communities as shown in Figure 4.8 I performed the community analysis by partitioning the network into subnetworks (community). A community is a collection of nodes (amino acids) that are connected with edges (connections). The communities that span across the RBD-ACE2 interface are shown in Figure 4.6(e,f) and the corresponding residues are shown in Figure 4.8.
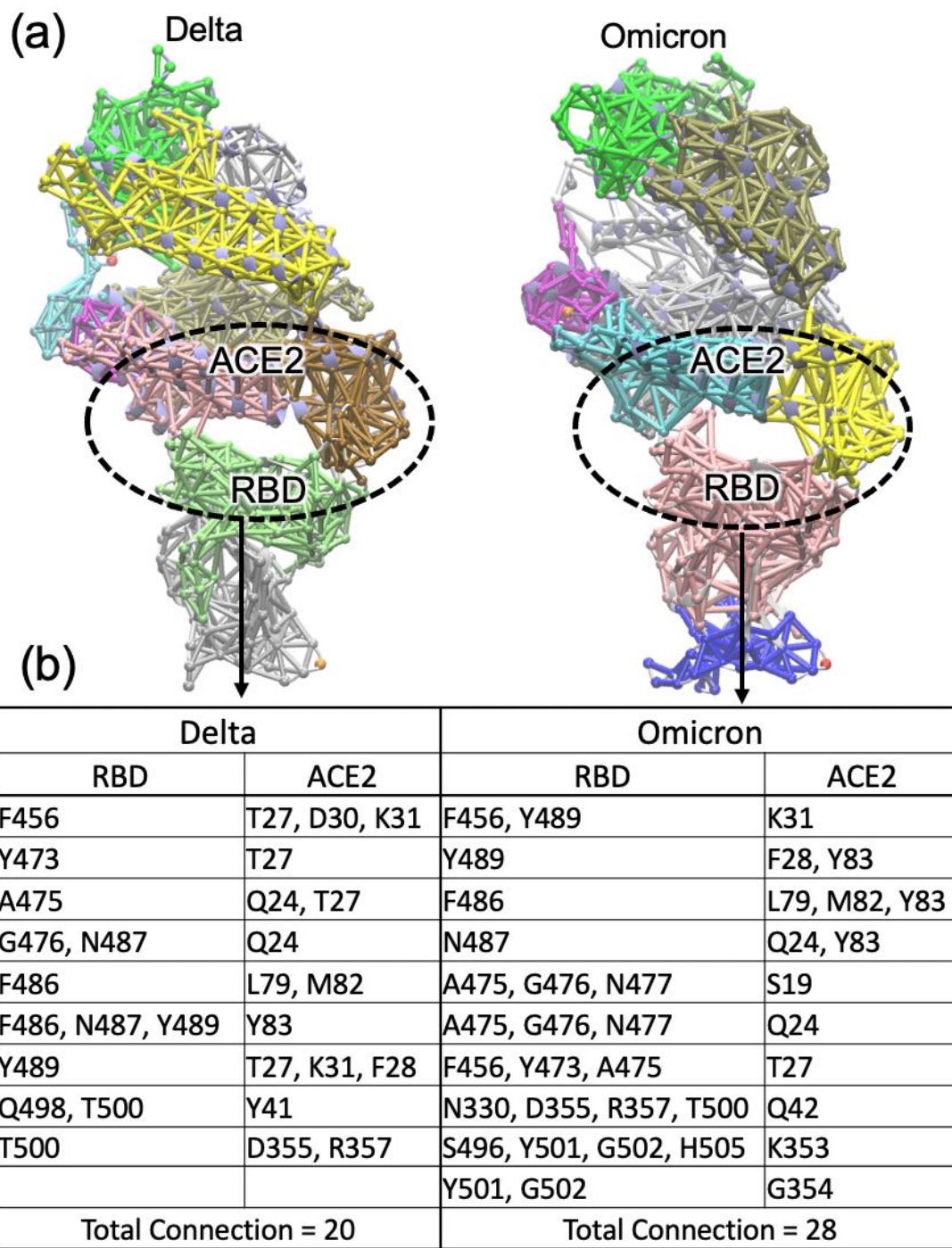
Figure 4.8 Community analysis performed for the first 50 ns of the trajectories for the Delta and Omicron variants. a) Each identified community is represented by a different color and the region that span both the RBD and ACE2 are circled. b) The amino acid residues involved in the dynamic network of the communities spanning the two proteins (circled region).

There are a total of 28 connections that occur between the RBD and ACE2 in Omicron, compared to 20 in Delta. This suggests a better binding of the RBD to ACE2 with increased interactions that stabilize the RBD-ACE2 complex in Omicron. While the network analysis and hydrogen-bond analysis provide a general assessment of the RBD-ACE2 binding, free-energy calculations with potential mean-force[134] or work done with steering force[160] are needed to compute more reliable binding affinities.

5.   THE ROLE OF PROTONATION AND NA$^+$ IN THE PFMATE TRANSPORTER

PROTEIN STRUCTURAL TRANSITIONS

This project was submitted to Physical Chemistry B journal and is currently under review:

Hossen, M.L., Bhattarai, N., Chapagain, P., and Gerstman, B., The role of protonation and

Na$^+$ in the PfMATE transporter protein structural transitions. Physical Chemistry B.

In this project, for the amino acids in PfMATE, I present computational results on the

dynamics, bonding, thermodynamic stability, connectivity and information transfer within

the different PfMATE structures. In the figures below, I used the following nomenclature

for the PfMATE structures: OF(np) for non-protonated outward-facing, OF(P) for

protonated outward-facing, IF(np) for non-protonated inward-facing, IF(p) for protonated

inward-facing, and IF(Na$^+$) for inward-facing without protonation but with Na$^+$ near

Asp41, Asn180, Asp184, Thr202.

5.1   Protein Structural Equilibration

The results presented in the following sections for the dynamics of the amino acid residues

in the five different PfMATE systems are determined from the 100 ns all-atom MD

simulation trajectories. To show that my MD structures have settled into equilibrium

structures, in Figure 5.1 for each system I display the root means square deviation (RMSD)

averaged over all of the Cα atoms relative to the initial positions. For all five systems, the

RMSD levels-off to a stable configuration by 50 ns, with little subsequent change in

RMSD. I use the subsequent 50 ns of the MD simulations for the analyses in the following

sections.
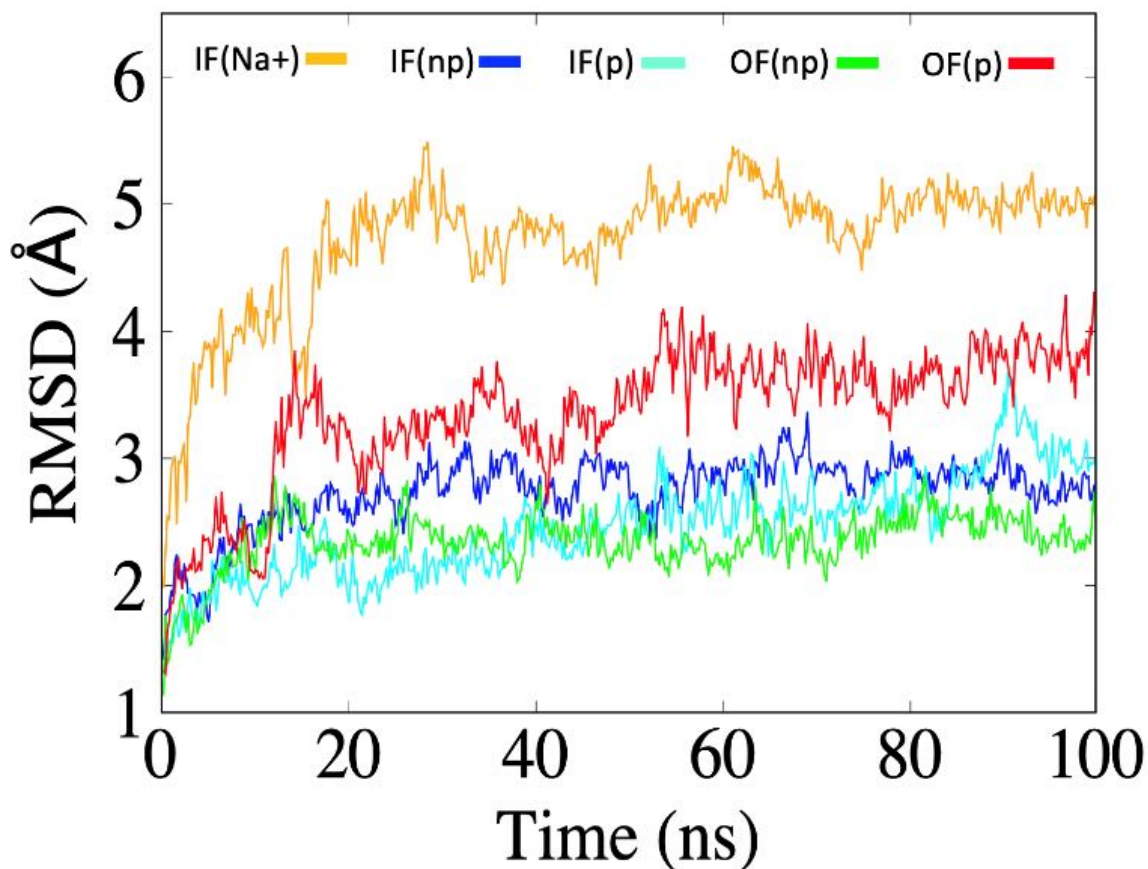
Figure 5.1 RMSD averaged over all C_α as a function of time for the MD trajectories of the five PfMATE systems. All five protein systems have settled into a structural configuration by 50 ns.

Figure 5.1 shows an RMSD trend among the five systems that is consistent with whether a system prefers the IF or OF configuration. The initial configuration for all five curves is based upon the experimental structures, which were IF(np) or OF(np). Figure 5.1 shows that the OF(np) structure underwent the smallest change in the positions of its Cα during its 100 ns MD simulation. This is expected since the np system prefers to be in the OF configuration. In contrast, the addition of Na^+ to the IF(np) structure results in a large RMSD, which is consistent with idea that the insertion of Na^+ causes a change in the interactions between the residues in the IF structure that will eventually (on much longer

time scales) lead to a transition to the OF structure. The changes in the residue interactions are examined in detail in later sections.

In addition, I also validated that my 40 ns TMD computations properly switched the PfMATE configuration between IF and OF for the five different systems. Table 5.1 displays the distance at the beginning and end of the TMD simulations between select amino acid pairs that are on either the outer lobe or inner lobe of the protein. The change in separation of these amino acids is consistent with the experimental data and show that the TMD simulations changed the structure between IF and OF.

Table 5.1. Distances between select amino acid pairs before and after Targeted MD simulations: inner lobe 95-318 and 96-318, outer lobe 44-316. The change in the distances are consistent with the experimental data and show that the TMD simulations changed the structure between IF and OF.

	OF(p)→IF(p)		OF(np)→IF(np)		IF(Na⁺)→OF(Na⁺)		IF(np)→OF(np)		IF(p)→OF(p)	
	Before TMD	After TMD	Before TMD	After TMD	Before TMD	After TMD	Before TMD	After TMD	Before TMD	After TMD
95-318 (TM3-TM9)	29.6	44.1	29.4	43.9	41.9	29.5	40.5	29.3	41.9	29.1
44-364 (TM1-TM10)	32.7	22.4	34.2	21.5	21.2	28.7	21.7	29.0	21.4	28.5
96-318 (TM3-TM9)	27.1	40.8	27.0	40.6	38.6	26.9	37.3	26.7	38.5	26.6

5.2   Amino Acid Rigidity: RMSF

I used the last 50 ns of the MD simulations to compare the dynamics of various amino acid residues among the five different PfMATE systems. I first examined the root mean square fluctuation in the position of the Cα of amino acids that were reported[76,95] to play important roles in the structural transformations that occur upon OF protonation or the IF binding of

Na$^+$. Compared to their behavior in the OF(np), in OF(p) the residue D184 experienced a large decrease in RMSF of 35%, residue E331 had a 28% decrease in RMSF, and E163 had a decrease of 9%. Upon the binding of Na$^+$ in the IF structure, D41, N180, and D184 experience large decreases in RMSF of 35%, 27%, and 10%, respectively. These results imply that the structural transformations involve important amino acids becoming more rigid in their position, perhaps acting as fulcrums. The importance of these amino acids in the structural transformation process is further explored below.

5.3   Pore Radius

I used the PDB IF and OF structures to calculate the radius of the PfMATE channel as a function of position along the channel axis traversing the membrane using the Hole program[169], a broadly used computational tool in the study of membrane protein channels[170-172]. The Hole program calculates the pore radius using the van der Waals radius of the atoms. Figure 5.2 shows the pore radius for IF and OF PfMATE structures, where the inner leaflet of the membrane has a negative channel coordinate. The pore radii were calculated taking the average of the radius over the last 50 ns of each 100 ns MD trajectory. The initial configuration was either the IF or OF structure from the PDB. As seen in Figure 5.1, none of the four structures undergo a structural transition during the relatively short 100 ns all-atom MD simulations. The IF(Na$^+$) pore has been found experimentally to eventually transition to an OF structure, whereas the IF(p) configuration is the preferred structure for the protonated protein. Interestingly, in Figure 5.2(a) the IF(Na$^+$) pore radius is significantly less uniform along the channel axis then the IF(p) pore. Likewise, as seen in Figure 5.2(b), the OF(np) pore, which is the preferred structure for the non-protonated

protein, is significantly smoother than the OF(p) pore that will eventually transform to the IF(p)
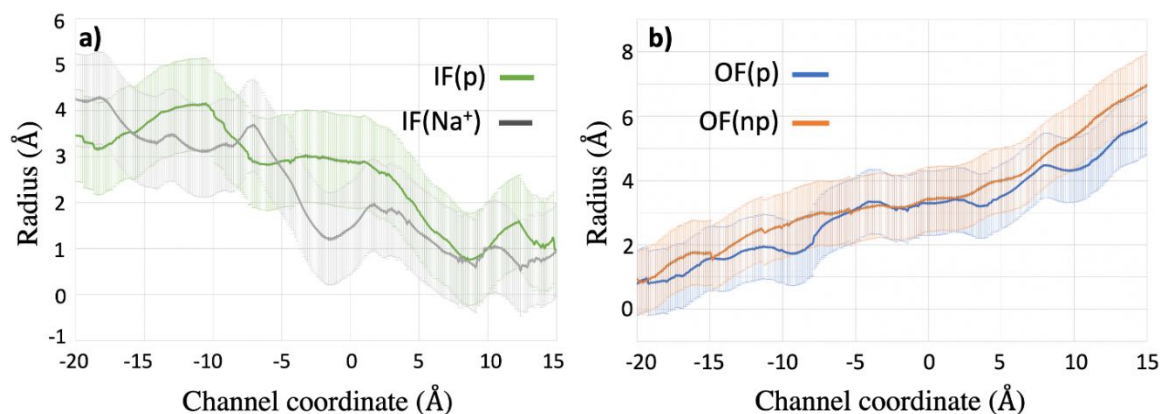
Figure 5.2 Pore radii along the coordinate axis traversing the membrane, where negative channel coordinate is for the membrane's inner leaflet. a) IF states, and b) OF states.

structure. Thus, even though all of the structures in Figure 5.2 appear to be stable on the 100 ns timescale of the all-atom MD simulations, the MD simulations can still provide information on the long-term stability of each PfMATE configuration.

5.4   Free Energy Along Transition Pathways

In Figure 5.3 I plot the free-energy surface (PMF) for several PfMATE structural transformations. The PMF curves are calculated from my Targeted Molecular Dynamics (TMD) computations as explained in the Methods section. I chose the RMSD as the reaction coordinate[130,173] between the initial and final configurations. The free-energy surface for all of the structural transitions is relatively smooth with no major kinetic barriers. Figure 5 shows that the transition IF(Na$^+$)→OF(Na$^+$) upon the addition of Na$^+$ to IF is especially favorable, as is the transition OF(p)→IF(p) upon protonation of OF. This is consistent with Jagessar's proposal[95] that the IF binding of a positively charged ligand such as Na$^+$ [76] causes the IF(Na$^+$)→OF(Na$^+$) transition, which is followed by a loss of Na$^+$

resulting in the OF(np) state. Then, protonation of the OF state causes the OF(p)→IF(p) transition.
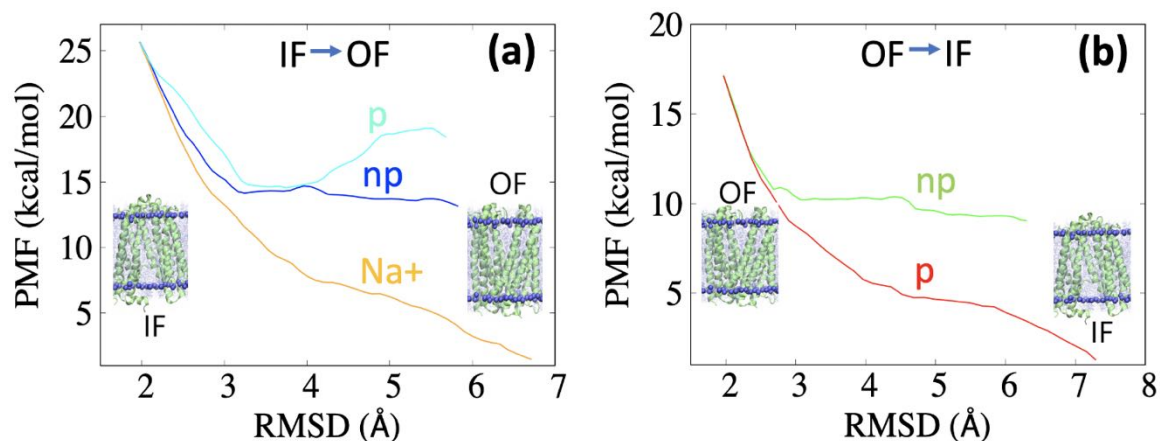
Figure 5.3 PMF (free energy landscape) for different structural transition of PfMATE a) IF → OF, b) OF → IF. On each graph, the curves were rescaled in such a way that their starting points are the same.

5.5   Energy and Hydrogen Bond Analysis

I used the TMD simulations displayed in Figure 5.3 to determine which amino acids experience significant changes in their bonding and interactions with other amino acids during the transition process. I first focused my attention on the amino acids that were discussed in refs.[76,95].

5.5.1   Hydrogen Bond and Energy Analysis for Specific Amino Acids

In Figure 5.4(a) I plot the number of hydrogen bonds made by the amino acids at the protonation site during the transition from OF(p)→IF(p). In Figure 5.4(b), I plot the number of hydrogen bonds made by the amino acids at the Na^+ binding site during the transition from IF(Na^+)→OF(Na^+). H-bonds are calculated using a cutoff distance of 3.5 Å and angle of 30°.
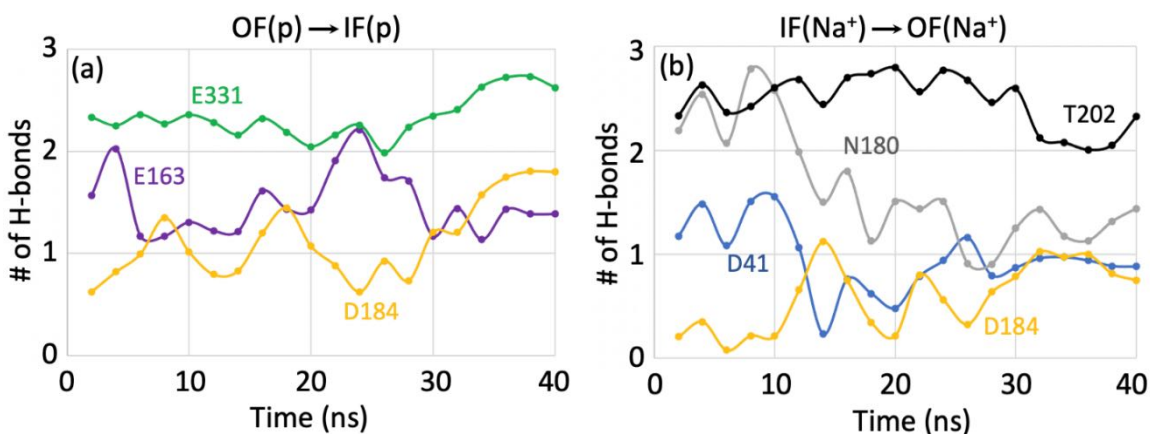
Figure 5.4 From the TMD simulations, the number of hydrogen bonds made by amino acids at the binding site during the transition for: a) protonation that induces OF(p)→IF(p), b) Na$^+$ binding that induces the transition from IF(Na$^+$)→OF(Na$^+$).

Figure 5.4(a) shows that amino acid E331 has a smooth and relatively small change in the number of H-bonds during the protonation induced OF(p)→IF(p) structural transition, whereas residues E163 and D184 undergo significant increases and decreases in the number of H-bonds throughout the transition. During the structural transition from IF(Na$^+$)→OF(Na$^+$) that occurs upon binding of Na$^+$, residues N180 undergo a significant decrease in the number of H-bonds during the structural transition, whereas residue D184 experiences an increase in the number of H-bonds during the transition. I also observe in Fig.6b that three of the amino acids (D41, N180, D184) experience significant fluctuations in the number of H-bonds during the transition. In both panels in Figure 5.4, D184 is involved in making and breaking H-bonds during the structural transitions. Further insight into the importance of D184 in the structural transitions is provided in the section on transfer entropy.

Table 5.2 provides changes in the electrostatic energy for important amino acids during the OF(p) → IF(p) transition and the IF(Na$^+$) → OF(Na$^+$) transition. For the OF(p) → IF(p)

47

transition, E163 experiences a small decrease in the strength (less negative) of its electrostatic energy. Since E163 is on the intracellular side of TM4, this is consistent with the inward side of the protein opening wider during the transition resulting in weaker residue interactions. In contrast, during the OF(p) → IF(p) transition, both D184 and E331 experience a significant increase in the strength of their electrostatic interactions that is consistent with the closing of the OF side of the protein during the transition.

For the IF(Na$^+$)→OF(Na$^+$) transition, I see that D41 has weaker electrostatic interactions after the structure has switched to OF, which is consistent with D41's location on the OF side of TM1. The electrostatic energies of the other residues in the same region (N180, D184, T202) are affected to a much smaller extent.

Table 5.2 Non-bonded electrostatic energy changes for important amino acids during structural transitions: (a) OF(p)→IF(p), (b) IF(Na$^+$)→OF(Na$^+$).

(a)	OF(p)→IF(p)			(b)	IF(Na$^+$)→OF(Na$^+$)		
Residue	Energy Before TMD (kcal/mol)	Energy After TMD (kcal/mol)	ΔE (kcal/mol)	Residue	Energy Before TMD (kcal/mol)	Energy After TMD (kcal/mol)	ΔE (kcal/mol)
E163	-10.9	-10.8	0.11	D41	-18.2	-9.67	8.57
D184	-34	-37.6	-3.6	N180	-35.5	-37.3	-1.8
E331	-9.62	-13.1	-3.4	D184	-14.6	-15.2	-0.6
-	-	-	-	T202	9.5	10.05	0.55

5.5.2 Interlobe H-Bonds

In addition to my investigations discussed above of H-bonds between amino acids that were signified as important in refs.[76,95], I also investigated H-bond that might reflect the large scale motions of the N- and C- lobes and TM helices upon the structural transitions. I focused on two H-bonds that help stabilize the OF structure as reported by ref. [76], the interlobe H-bond R88-E310, and H-bond R244-E393 between TM7-TM11. These bonds

are displayed in Figure 5.5(a). TM7 and TM11 are both in the C-lobe in the OF structure, but TM7 undergoes a large transverse shift to the N-lobe during the structural transition to the IF configuration. Figure 5.5(b) displays three interlobe H-bonds that are important in stabilizing the IF configurational structure.
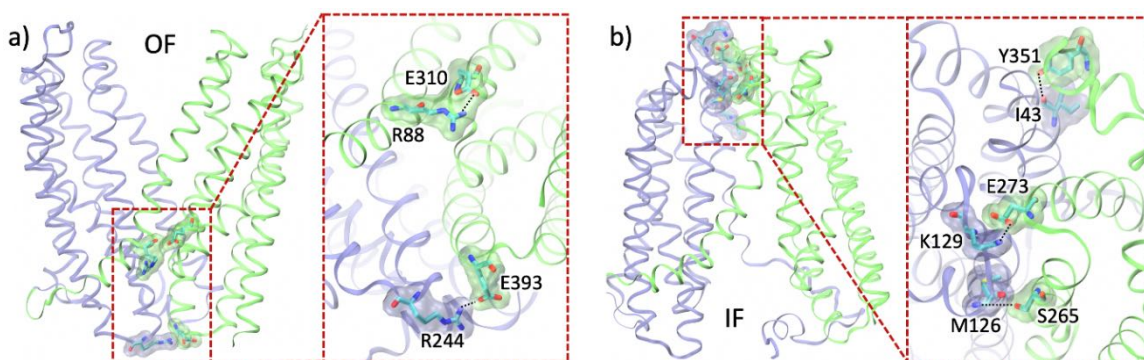
Figure 5.5 Important H-bonds that stabilize the structural conformation. a) OF interlobe H-bond R88-E310, and H-bond R244-E393 between TM7-TM11. b) Three interlobe H-bonds that are important in stabilizing the IF configurational structure.

In Table 5.3, I report the percentage of time that these H-bonds exist when the protein is in the OF configuration compared to when the protein is in the IF configuration. The H-bond percentages are determined from the last 20 ns of the 100 ns MD simulation for the OF(p) configuration and from the last 20 ns of the 100 ns MD simulation for the IF(Na$^+$) configuration. My H-bond results confirm the relative motion of the N- and C- lobes, as well as the displacement of TM7 away from TM11 discussed by Zakrzewska[76].

Table 5.3 Percentage of time that interlobe (IL) or TM-TM H-Bonds exist in the OF configuration compared to the IF configuration.

T	OF	IF
R88-E310 (IL)	74.4%	0.0%
R244-E393 (TM7-TM11)	74.6%	0.0%
K129-E273 (IL)	0.0%	50.3%
S265-M126 (IL)	0.0%	37.3%
Y351-E273 (IL)	0.0%	50.1%

5.6   Dynamic Network Analysis of Amino Acid Connectivities for Different Structural Conformations

Jagessar et al.[95] propose a mechanism for isomerization from the OF to IF configuration upon protonation. In their model, the IL6-7 loop functions as a belt that stabilizes the OF state. Upon E163 protonation, the interaction between IL4-5 and IL6-7 is disrupted and IL6-7 moves away from the N-lobe, which allows TM3 to tilt relative to the inner leaflet side of the membrane and facilitates the structural transition to the IF configuration. In order to investigate this proposal on the level of dynamic structural correlations among amino acids, and to uncover other structural cascading possibilities, I calculated the strength of structural network connections among amino acids when the PfMATE protein is in different configurations.

I used data from the last 20 ns of the 100 ns MD trajectories to perform network analysis. Figure 8a-e displays the network communities of the different PfMATE systems. The small spherical nodes are the Cα of the amino acids. Nodes within a community are connected by cylindrical rods whose thickness represents the strength of the connection. I found a different number of communities in the different systems: OF(np)-10, OF(p)-14, IF(p)-7, IF(np)-11, and IF(Na$^+$)-11. The list of communities and the participating amino acids are provided in Table 5.4.

Table 5.4 Data from the last 20 ns of the 100 ns MD trajectories was used to perform the dynamical network analysis. This is a list of amino acids participating in all of the network communities of the different PfMATE systems. The communities are pictured in Fig. 8 in the main article.

Structure	Community	Residue number
OF(np)	1	4, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 80, 81, 82, 97, 101, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168,

		169, 170, 171, 172, 173, 175, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 227
	2	5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 83, 85, 86, 87, 88, 305, 306, 307, 308, 309, 310, 311, 312, 315
	3	251, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 272, 403, 404, 405, 408, 409, 410, 411, 412, 413, 414, 415, 416, 417, 418, 419, 420, 421, 422, 423, 424, 425, 427, 429
	4	42, 43, 44, 45, 46, 47, 48, 179, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205
	5	271, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 336, 338, 340, 341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 368, 426
	6	34, 36, 37, 38, 39, 40, 41, 174, 176, 177, 178, 180, 181
	7	252, 369, 370, 371, 372, 373, 374, 375, 376, 378, 379, 382, 383, 386, 391, 392, 393, 394, 395, 396, 397, 398, 399, 400, 401, 402, 406, 407, 428, 430, 431, 432, 433, 434, 435, 436, 437, 438, 439, 440, 441, 442, 443, 444, 445, 446, 447, 449, 450, 453, 455
	8	84, 89, 90, 91, 92, 93, 94, 95, 96, 98, 99, 100, 102, 103, 226, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250
	9	15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 35, 79, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 313, 314, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 331, 332, 333, 334, 335, 337, 339, 377, 380, 381, 384, 385, 387, 388, 389, 390, 448, 451, 452, 454
	10	49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145
OF(p)	1	4
	2	5, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 27, 288, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 374, 376, 377, 378, 379, 380, 381, 382, 383, 384, 385, 388, 389
	3	6, 7, 10, 11, 12, 13, 14

	4	250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 372, 375, 397, 399, 400, 401, 402, 403, 404, 405, 406, 407, 408, 409, 410, 411, 412, 413, 414, 415, 416, 417, 418, 419, 420, 421, 422, 423, 424, 425, 426, 427, 428, 429, 430, 431, 432, 433, 434, 435, 436, 437, 439
	5	26, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 46, 47, 48, 174, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 196, 197, 198, 200, 201, 202, 203, 204, 205
	6	116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129
	7	330, 331, 332, 333, 334, 335, 336, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 348, 349, 351, 353
	8	8, 9, 77, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 102, 159, 160, 161, 162, 163, 164, 165, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 249
	9	149, 153, 157, 166, 167, 168, 169, 170, 171, 172, 173, 175, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223
	10	286, 287, 289, 350, 352, 354, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 368, 369, 370, 371, 373
	11	386, 438, 440, 441, 442, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452, 453, 454, 455
	12	45, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 63, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 195, 199
	13	62, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 78, 101, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 147, 148, 150, 151, 152, 154, 155, 156, 158
	14	248, 387, 390, 391, 392, 393, 394, 395, 396, 398
IF(p)	1	4, 5, 7, 9, 11, 13, 378, 379, 380, 382, 383, 385, 386, 387, 389, 390, 391, 392, 393, 394, 395, 396, 397, 398, 399, 400, 402, 440, 441, 442, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452, 453, 454, 455
	2	30, 32, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 47, 48, 49, 51, 52, 53, 56, 60, 149, 165, 167, 168, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223

	3	31, 33, 46, 50, 274, 278, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 333, 335, 336, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 368, 369, 370, 371, 372, 373, 374, 376, 426
	4	55, 58, 59, 61, 62, 63, 64, 66, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 150, 257
	5	65, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 166, 169, 213, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253
	6	54, 57, 126, 254, 255, 256, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 275, 276, 277, 279, 375, 401, 403, 404, 405, 406, 407, 408, 409, 410, 411, 412, 413, 414, 415, 416, 417, 418, 419, 420, 421, 422, 423, 424, 425, 427, 428, 429, 430, 431, 432, 433, 434, 435, 436, 437, 438, 439
	7	6, 8, 10, 12, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 293, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 331, 332, 334, 377, 381, 384, 388
IF(np)	1	4
	2	5, 7
	3	6, 8, 10, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 29
	4	28, 33, 34, 35, 36, 38, 149, 164, 165, 166, 167, 168, 170, 171, 172, 173, 174, 175, 176, 177, 178, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 227
	5	39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 196, 197, 198, 200, 201, 202, 203, 204, 351, 353
	6	9, 11, 37, 252, 289, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 331, 332, 333, 334, 335, 336, 337, 340, 370, 373, 374, 376, 377, 378, 379, 380, 381, 382, 383, 384, 385, 386, 387, 388, 389, 390, 391, 392,

		393, 394, 395, 396, 397, 398, 399, 400, 402, 440, 441, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452, 453, 454, 455
	7	64, 256, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 290, 338, 339, 341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 355, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 369, 404, 422, 423, 425, 426
	8	30, 31, 32, 61, 63, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 148, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 169, 225, 226, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 253, 254, 255, 257
	9	403, 408, 409, 411, 412, 413, 414, 415, 416, 417, 418, 419, 420, 421, 424, 427
	10	49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 62, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 195, 199, 352, 354, 356
	11	368, 371, 372, 375, 401, 405, 406, 407, 410, 428, 429, 430, 431, 432, 433, 434, 435, 436, 437, 438, 439, 442
IF(Na$^+$)	1	4, 6, 8, 10, 297, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 377, 378, 379, 380, 381, 382, 383, 384, 385, 386, 387, 388, 389, 390, 391, 392, 393, 394, 395, 396, 397, 398, 399, 400, 439, 440, 441, 442, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452, 453, 454, 455
	2	36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 142, 146, 168, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 222

	3	31, 33, 35, 278, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 331, 332, 333, 334, 335, 336, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 368, 369, 370, 371, 372, 373, 374, 375, 426
	4	65, 66 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 169, 221, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 245
	5	244, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 261, 403
	6	49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 143, 273
	7	105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 144, 145, 147
	8	260, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 274, 275, 276, 277, 279, 418, 419, 420, 421, 422, 423, 424, 425, 427
	9	13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 299
	10	5, 7, 9, 11, 12, 30, 32, 34, 290, 291, 292, 293, 294, 295, 296, 298
	11	376, 401, 402, 404, 405, 406, 407, 408, 409, 410, 411, 412, 413, 414, 415, 416, 417, 428, 429, 430, 431, 432, 433, 434, 435, 436, 437, 438

The number of communities does not necessarily correlate with the stability of a structure. Important details of the communities such as size and scope across the protein are crucial in understanding how structural stability is controlled by amino acid connectivity. In Figures Figure 5.6(a'), 5.6(c'), 5.6(d'), 5.6(e'), I show only the network communities that have amino acids that span both N- and C-lobes. Amino acids not in these lobe-spanning communities are displayed as ribbon structures. Figure 8(b') shows two communities for the OF(p) structure. Neither of these two OF(p) communities span the N- and C-lobes.

They are highlighted because they contain amino acids with strong connectivity that break upon conversion to the IF(p) conformation, as discussed below. For all panels in Figure 5.6 the N-lobe (violet shading) is on the left and the C-lobe (light green) is on the right, except for Figure 5.6(a'). The communities in Figure 5.6(a') are more easily viewed from the other side of the protein, and thus the N-lobe is on the right and the C-lobe is on the left.
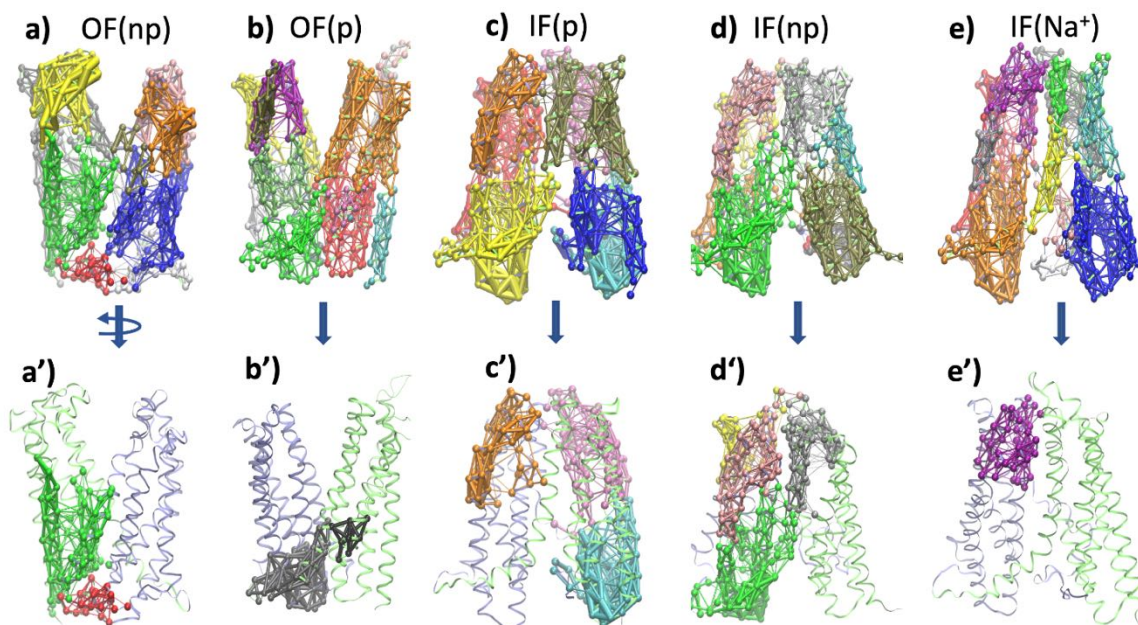
Figure 5.6 Communities from dynamical network analysis for the last 20 ns of the 100 ns MD trajectories. Communities in each system are shown in different colors. The small spherical nodes represent amino acids. Thicker lines between nodes represent stronger correlations in positional fluctuations. (a) OF(np), (b) OF(p), (c) IF(p), (d) IF(np), (e) IF(Na$^+$). Communities with amino acids spanning the N- and C-lobes are displayed in (a'), (c'), (d'), (e'). Figure 8b' displays two communities in OF(p). Neither community spans the N- and C-lobes but both contain strong connections that break when the structure converts to IF(p). For all figures except Figure (a'), the PfMATE N-lobe (lavender shading) is on the left and the C-lobe (light green) is on the right. For Figure 5.6(a'), the highlighted communities are better viewed from the other side of the protein and thus the lobe positions are reversed.

5.6.1 Protonation Induced Changes in Network Communities

Communities that span both the N- and C- lobes stabilize the overall structure of the protein. Table 5.5 shows network connections that must be broken in order for the structure

to convert. I found that in the stable OF(np) structure (Fig. 8a') there are two communities that span the N- and C- lobes (Table 5.5, OF(np) Communities 1 and 8) with many connections between the lobes. In contrast, upon protonation to the unstable OF(p) structure (Fig. 8b') there are no communities that span the N- and C- lobes (Table 5.5, OF(p)). However, Fig. 8c' and Table 5.5 show that upon conversion to the IF(p) structure, three communities are created that stabilize the IF(p) structure by spanning the N- and C-lobes with many connections. Also provided for comparison in Table 5.5 is the sum of the strengths of the connections for protonated versus non-protonated configurations. The strengths are given in arbitrary units.

Further insight into the structural conversion process upon protonation of the OF structure can be found in two communities in the unstable OF(p) configuration. The unstable OF(p) configuration contains Community 7 (N-lobe) that has strong internal amino acid connections that are broken during the structural conversion to IF(p), as does OF(p) Community 13 (C-lobe). Table 5.5 shows that within Community 7, residue G8 has strong connections to residue E163 and G164, and residue V9 has strong connectivity to I85, E163, and Y224. Residues G8 and G9 are on helix TM1 and E163 is on TM4. In the OF structure G8, G9 are near E163. However, in the IF structure, G8 and G9 are far from E163 and the connections between them are broken. As proposed by Jagessar[95], this shows that E163 plays a crucial role in the structural transition from OF(p) to IF(p).

Table 5.5 Network connections that are broken if the PfMATE converts to the other structure. The stable OF(np) structure contains many inter-lobe and intra-lobe connections (26 total), whereas protonation changes the connectivity so that only a small number (8 total) of intra-lobe connections must be broken for OF(p) to convert. The IF(p) structure contains many inter-lobe connections that must be broken upon structural conversion. The strength of the connections is given in arbitrary units.

Community	OF(np) Between		OF(p) Between		IF(p) Between	
	N-lobe	C-lobe	N-lobe	C-lobe	N-lobe	C-lobe
1	V9, L12, R13 -I85	-	-	-	-	-
	R13-R88	-				
	G86-	A306				
	A87-	E310				
	R83-	A306, G309				
2	-	-	-	-	M31-	V291, P292
					S46-	L348, F349, Y351
					P50-	N274, S354, V357
3	-	-	-	-	P61, M64	-S257
6	-	-	-	-	T6-	G305, G388
7	-	-	G8- E163, G164	-	-	-
			V9- I85, E163, Y224	-	-	-
8	M31, S32, T35	-V291	-	-	-	-
	M28-	P292, I330				
	I20, S24	-M296				
	K17-	Y322				
	S24-	A299				
	I20-	T300				
	M27, M31	-L294				
	L23-	A299, S302				
	A19-	V303				
	S79-	T301, Q387, G388				
13	-	-	-	P248- G392, E393, L396	-	-
Total Connections	26		8		12	
Sum of Connection Strengths	35.8		9.1		11.8	

5.6.2 Na$^+$ Induced Changes in Network Communities

I investigated how the presence of Na$^+$ changes the connection networks for the IF structure in comparison to IF(np). All of the networks are provided in Table 5.4. Table 5.6 shows network connections that must be broken in order for the IF structure to convert. The IF(np) structure contains a moderate number of inter-lobe connections that must be broken for conversion to OF, whereas the presence of Na$^+$ greatly reduces the number and strength of stabilizing inter-lobe connections, which facilitates the transition.

Table 5.6 Network connections that are broken if the PfMATE converts to the other structure. The presence of Na⁺ great reduces the number and strength of the connections that must be broken. The strength of the connections is given in arbitrary units.

Community	IF(np) Between		IF(Na⁺) Between	
	N-lobe	C-lobe	N-lobe	C-lobe
4	I43, W44, S46, G47	-Y351	-	-
5	-	-	A54, K129, G130	-E273
6	M64-	M256, M287	-	
7	P61, G65	-S257	-	-
	A68-	L254		
9	G49-	S352	-	-
	P50-	S354		
Total Connections	11		3	
Sum of Connection Strengths	14.1		5.2	

5.6.3 Size and Strength of Network Connectivities

I also used my network analysis to further quantify the size and strength of the networks containing amino acids that are involved in structural conversion from OF→IF upon protonation (E163, D184, E331) or involved in Na⁺ binding (D41, N180, D184, T202) that causes the IF OF structural conversion. Figures 5.7(a,b) show all connections involving the communities that contain E163, D184, E331 in the unstable OF(p) and stable OF(np), respectively. Figure 5.7(c) shows all connections involving the communities that contain D41, N180, D184, T202 in the unstable IF(Na⁺) as well as the unprotonated D41, E163, D184, E273, E331 residues. Figure 5.7(d) shows all connections in the stable IF(p) involving the communities that contain D41, N180, D184, T202 in the absence of Na⁺ but with residues D41, E163, D184, E273, E331 protonated in the IF configuration. The figures show that the stable configurations have larger networks with stronger connectivity (edges). I quantified this by summing the strengths of all the connections in each structure.

The results are: unstable OF(p): 9.1, stable OF(np): 35.8, unstable IF(Na^+): 5.2, stable IF(p): 11.8. A clear trend is apparent between the stability of a structure and the sum of the strengths of the connections in the network communities involving key amino acid.
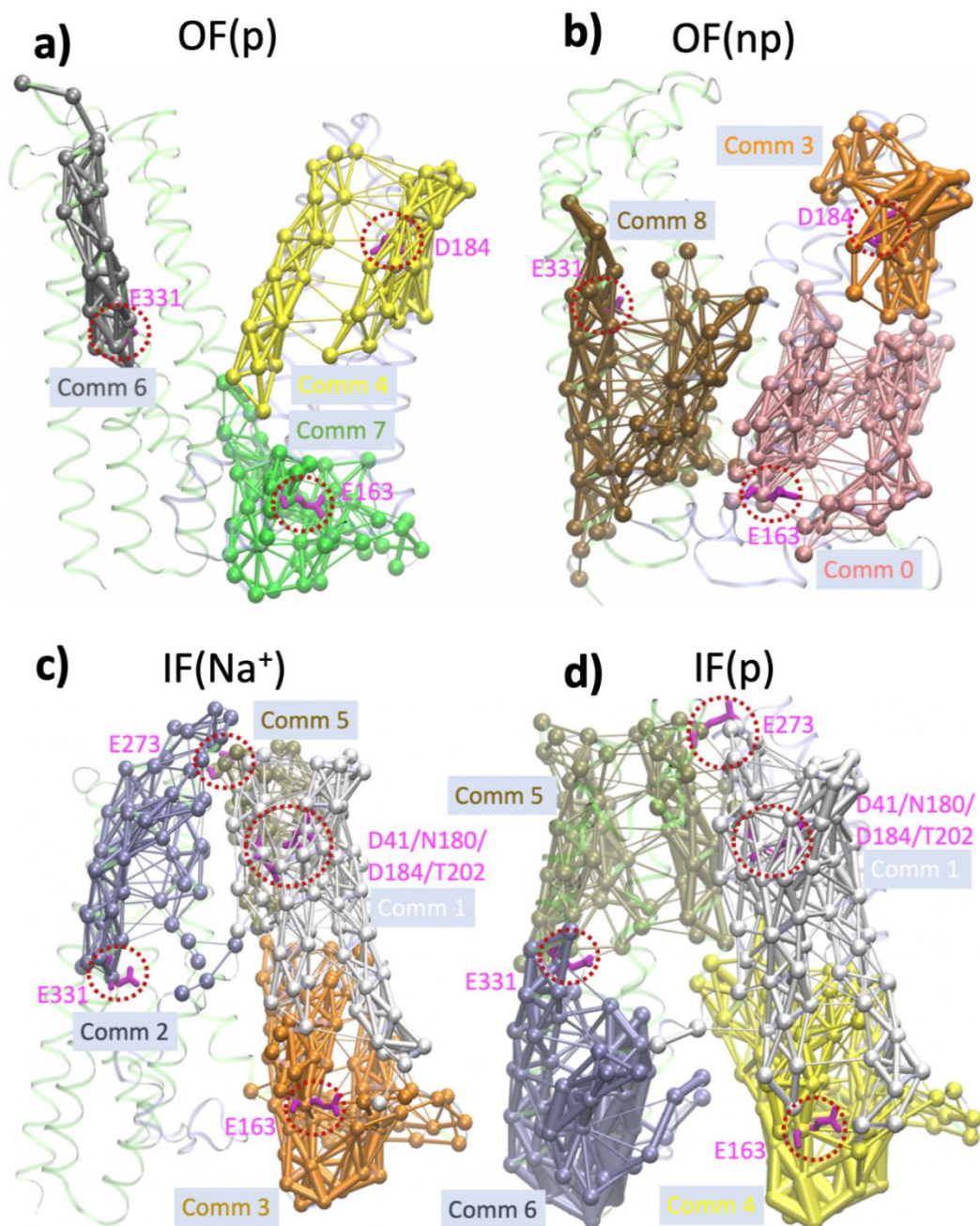
Figure 5.7 Communities that contain amino acids that are involved in structure conversion upon protonation or binding of Na^+. (a) OF(p), (b) OF(np), (c) IF(Na^+), and (d) IF(p) PfMATE systems.

5.7 Transfer Entropy and Information Flow for Structural Transitions

The large-scale structural transitions between IF and OF protein configurations require the transfer of information among amino acids. References[76,95] identify amino acids located at the protonation site as well as amino acids at the binding site for Na^+. In order to determine which of these amino acids are likely to generate the information that causes the structural transition, I performed Transfer Entropy (TE) calculations using the computational technique described in ref.[142]. A TE donor (leader, TE>0) is a source of structural information and can drive a structural transition, whereas a TE receptor (follower, TE<0) receives structural information. In addition, TE pathways in the protein involving several amino acids can also provide insight into the mechanistic details of structural transitions.

I am interested in determining how information is transferred through the protein to trigger the structural change. The OF(np) configuration is stable, but protonation of the OF configuration seems to cause the OF→IF transformation. Therefore, for comparison I calculated TE of amino acids in both the OF(np) and OF(p) states. Likewise, IF(np) is stable, but the binding of Na^+ causes the IF→OF transformation. For comparison, I calculate TE for amino acids in IF(np) and IF(Na^+). The TE is calculated from the trajectories of the 100 ns all-atom MD simulations for each of the four systems.

5.7.1 Protonation Induced Changes in TE

Protonation causes the structural transition from OF to IF. The amino acids around the protonation binding site were identified[76] as E163, D184, E331. Jagessar[95] proposes a mechanism for isomerization from the OF to IF state that is initiated specifically by E163 protonation in OF that loosens the IL6-7 loop. My TE results support this proposal, and

also provide additional insight concerning the effects of D184 and E331 on the structural configuration upon protonation. To determine how these amino acids communicate the difference between non-protonated versus protonated to other amino acids to generate the structural transition from OF to IF, I looked at two parameters from the Transfer Entropy calculations: the strength of an amino acid as a TE donor, and the importance of the TE protein pathway that starts from a specific amino acid. I specifically looked for trends in how these two parameters changed between the non-protonated outward facing OF(np) structure and the protonated OF(p) structure. With respect to E163, I found that in the OF(np) structure, the TE pathway that starts with E163 (Table 5.7) has a rank of 377 among all 451 TE pathways calculated, and the pathway does not include any amino acids in the IL6-7 loop. In addition, there are seven other OF(np) TE pathways that E163 is part of, but none of these pathways involve residues in the IL6-7 loop. Upon protonation of OF, the only TE pathway involving E163 is the pathway that starts with E163, which rises in importance to a rank of 328 and includes one residue (D222) on IL6-7. These TE results show that protonation leads to an increase in communication between E163 and the IL6-7 loop compared to OF(np).

Table 5.7 The amino acid TE pathways for the OF structure that start with E163, D184, and E331. The pathways (and ranks) are given for the OF(np) structure and the OF(p) structure.

E163	np	E163 → K7 → G379 → L375 → E273 → A355 → S354 → E131 → G130 → K129 → L116 → P117 → G164 → L11 → K7 (Rank 377)
	p	E163 → G162 → V9 → G8 → S352 → K4 → A128 → D222 → L348 → K4 (Rank 328)
D184	np	D184 → V182 → I181 → G289 → G130 → K129 → L116 → P117 → G164 → L11 → K7 → G379 → L375 → E273 → A355 → S354 → E131 → G130 (Rank 5)
	p	D184 → V182 → G178 → R13 → T5 → K4 → A128 → D222 → L348 → K4 (Rank 27)

E331	np	E331 → T5 → V174 → G176 → A40 → N167 → R161 → G162 → S104 → L229 → A230 → D89 → G8 → K7 → G379 → L375 → E273 → A355 → S354 → E131 → G130 → K129 → L116 → P117 → G164 → L11 → K7 (Rank 153)
	p	E331 → L332 → G289 → P234 → L229 → L12 → G8 → S352 → K4 → A128 → D222 → L348 → K4 (Rank 45)

Table 5.8 shows that protonation decreases the role of E163 in receiving structural transformation from other amino acids in the protein. In addition, Table 6 shows that upon protonation, E331 has the largest increase as a TE donor and its TE pathway rank increases in importance the most. This is consistent with the experimental work of ref.[76], which found that the OF→IF structural transition can be triggered by protonation of E331. In addition to the increase in donor strength of E331 when protonated, the TE pathway involving E331 also increases in importance upon protonation. Table 5.7 shows that the only OF pathway involving E331 contains two amino acids in the IL6-7 loop (L229, R230) and has a relatively low rank of 153. When protonated, E331 again participates in only one OF pathway and this pathway includes two different residues in the IL6-7 loop (D222, P234). However, the rank of this TE pathway is 45, significantly more important than in the OF(np) structure. This TE result on the importance of E331 upon protonation is also consistent with the result presented earlier that E331 undergoes a large change in RMSF upon protonation. For D184, transfer entropy shows no clear trends. Table 5.7 shows that D184's only TE pathway decreases in rank upon protonation, but protonation also changes the TE pathway to include an amino acid in IL6-7 (D222) whereas there is no IL6-7 residue in D184's OF(np) TE pathway. Table 6 shows that upon protonation, D184 loses strength as a TE donor in the OF structure.

Table 5.8 Change in Transfer Entropy parameters for the OF structure with no protonation (OF(np)) versus with protonation (OF(p)). The strength as a TE Donor is given as a normalized value. The rank of the primary information pathway that starts with a specific amino acid is given with respect to the 451 pathways generated by the TE algorithm of ref.[142]. A lower rank means a more important TE pathway.

	OF(np)		OF(p)	
	Donor Strength	Pathway Rank	Donor Strength	Pathway Rank
E163	-6.89	377	-3.63	328
D184	23.72	5	14.21	27
E331	1.60	153	7.26	45

5.7.2 Na$^+$ Induced Changes in TE

The binding of Na$^+$ causes the structural transition from IF to OF. The amino acids around the Na$^+$ binding site were identified[95] as D41, N180, D184, T202. To determine which of these amino acids are the likely source of information that is transferred to other amino acids to generate the structural transition from IF to OF, I looked at two parameters from the Transfer Entropy calculations: the strength of an amino acid as a TE donor, and the importance of the TE protein pathway that starts from a specific amino acid. I specifically looked for trends in how these two parameters changed when the IF structure had not bound Na$^+$ versus the IF structure with Na$^+$ bound near these four amino acids. The TE calculations show that the binding of Na$^+$ significantly increases the TE donor strength of T202, N180, and D184, but causes D41 to become more of a TE receptor. The trend of increasing importance as a TE donor for T202, N180, and D184 is shown in both the increase in TE donor value as well as an increase in the importance of each amino acid's primary pathway, as seen in Table 5.9 As with the RMSF results, the TE results show that the binding of Na$^+$ has the largest effect on the dynamics of N180.

Table 5.9 Change in Transfer Entropy parameters for the IF structure with no Na$^+$ versus with Na$^+$ bound. The strength as a TE Donor is given as a normalized value. The rank of the primary information pathway that starts with a specific amino acid is given with respect to the 451 pathways generated by the TE algorithm of ref.[142]. A lower rank means a more important TE pathway.

	IF Without Bound Na$^+$		IF With Bound Na$^+$	
	Donor Strength	Pathway Rank	Donor Strength	Pathway Rank
N180	4.09	378	15.24	50
D184	3.53	61	8.86	23
T202	-5.57	103	8.18	82
D41	-0.08	70	-6.61	132

6.   IN-SILICO SCREENING OF INHIBITORS TARGETING THE SARS-COV-2 ENVELOPE PROTEIN

In this project, I performed in silico drug screening using molecular docking with the US Food and Drug Administration (FDA)-approved and investigational drugs targeting a conformational ensemble of SARS-CoV-2 E protein (PDB ID: 7K3G)[43] by doing molecular dynamics (MD) simulation. These drugs can be repurposed as anti-viral drugs against SARS-CoV-2. The screening study identified 6 promising drugs that presented outstanding chemical and pharmacokinetic properties and stability inside the E-protein during MD simulation, thereby could be recommended for further experimental analyses and pre-clinical trials.
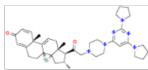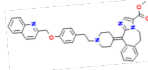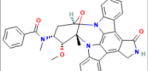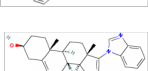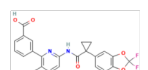
6.1   Molecular Docking Based Screening Results

I generated an ensemble of 100 PDBs of SARS-CoV-2 E protein using the molecular dynamics simulation trajectory. I took a list of about 3800 drug molecules that includes both FDA-approved and investigational drugs and performed a high-throughput in silico docking with them to target the E protein. The docking results were further sorted based on the binding energy generated by the docking tool AutoDock Vina 4.2[116], and was followed by further analysis- docking based interaction analysis, MD based analysis, pharmacokinetics studies etc.

Based on the binding energy score and the visualization of the docked complexes, I chose six best compounds from a total of 380000 hits as listed in Table 6.1. All the six compounds interact inside the transmembrane (TM) of the E protein. However, five of them, tirilazad, laniquidar, midostaurin, siramesine, and galeterone, bind to C-terminal domain, and one,

lumacaftor, fits in to the N terminal domain of the E protein. The C-terminal of the TM domain found to be an active site for the drug binding in previous studies[174,175], whereas Mandala et al[43] found that small-molecules/drugs have high-affinity to bind to N-terminal in order to occlude the E-protein in the presence of membrane bilayer. All the six compounds show hydrophobic interaction with E-protein, and two compounds show hydrogen bonds made with E-protein based on the result generated by a 2D ligand-protein interaction diagrams application LigPlot[+176]. Laniquidar and Lumacaftor have the highest number of hydrophobic interactions with five and four respectively, and others have only three hydrophobic interactions as shown in Figure 6.1. Only lumacaftor make hydrogen bonds with E-protein residue Thr9. According to binding energy, laniquidar and galeterone are the two highest ones with -11.5 kcal/mol each, next one is tirilazad with -11.3 kcal/mol, and midostaurin, siramesine, lumacaftor each has -11.2 kcal/mol as shown in Table 6.1. The binding sites or active sites of E-protein in the selected complexes are shown in Figure 6.2.

Table 6.1 Six best docking hits, their 2D chemical structures, binding energy, hydrogen bond, and hydrophobic interactions.

Compound	2D Chemical Structure	Binding Energy (kcal/mol)	Hydrogen Bond	Hydrophobic interaction
Tirilazad		-11.3	No	Leu28, Ala32, Cys44
Laniquidar		-11.5	No	Leu21, Val25, Leu28, Val29, Ala32
Midostaurin		-11.2	No	Leu28, Ala32, Ala36
Siramesine		-11.2	No	Leu28, Ala32, Ala36
Galeterone		-11.5	No	Leu28, Ala32, Thr35
Lumacaftor		-11.2	Yes	Thr11, Val14, Val17, Leu18
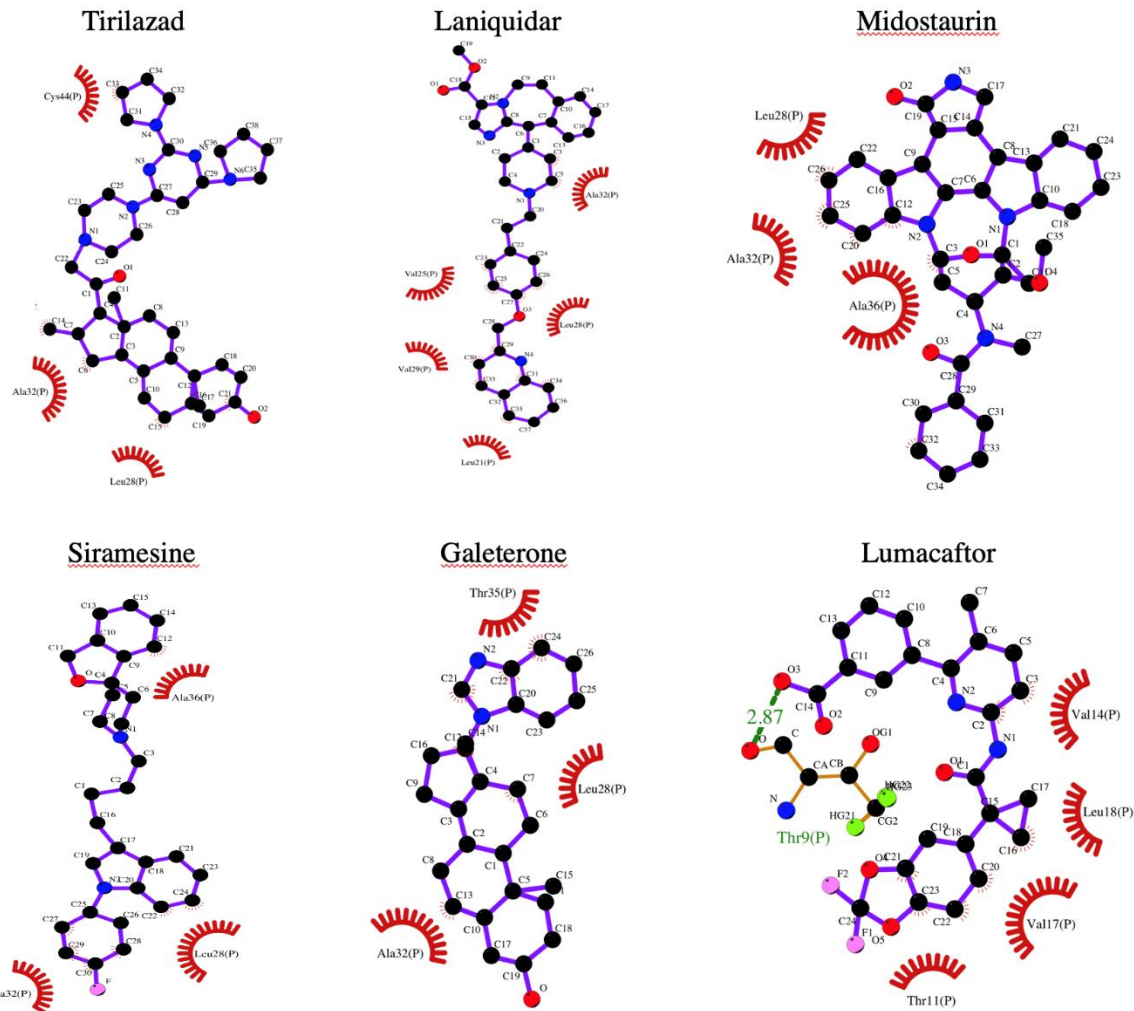
Figure 6.1 2D ligand-protein interaction diagrams of six compounds with highest binding energies obtained in docking calculation against SARS-CoV-2 E-protein. The interaction diagrams were made with LigPlot$^+$ application.
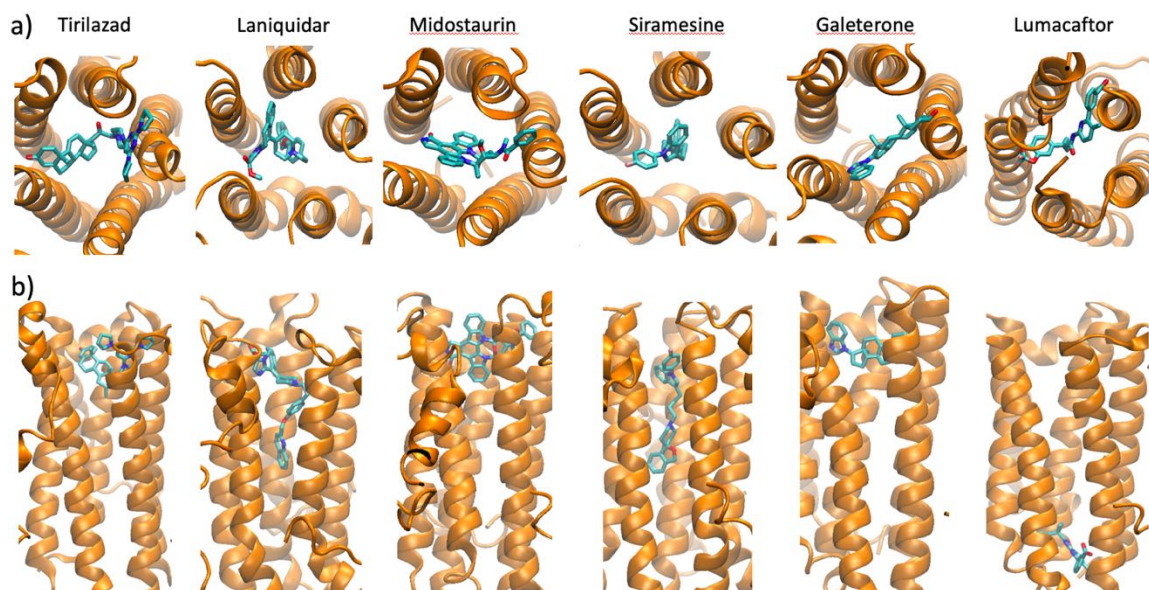
Figure 6.2 3D interaction diagram of six compounds. a) the top/bottom view of the drug-protein complexes and b) the side of the same complexes that show the exact location of the binding site in the transmembrane domain of the E-protein. The drugs are shown in stick representation.

6.2   Molecular Dynamics (MD) Simulation Results

Based on the docking results, the best fit six complexes of drug molecules and E-protein structures are now considered for MD simulations related analysis, that is helpful understanding the ligand-protein interaction, conformational dynamics of biological and chemical compounds. Six systems were made for tirilazad, laniquidar, midostaurin, siramesine, galeterone, and lumacaftor complexed with E-protein and embedded in ERGIC membrane bilayer using CHARMM-GUI website[87,108]. The systems were hydrated with a cubical TIP3 water box and ionized with a $Ca^{2+}Cl^{-}$ for neutralization. All other parameters were set to standard MD simulation parameters default in NAMD2.14 software[110,177]. The systems were minimized and equilibrated for a short period of time and was followed by a 100 ns MD production run with a 2 fs timestep and 300 K temperature.
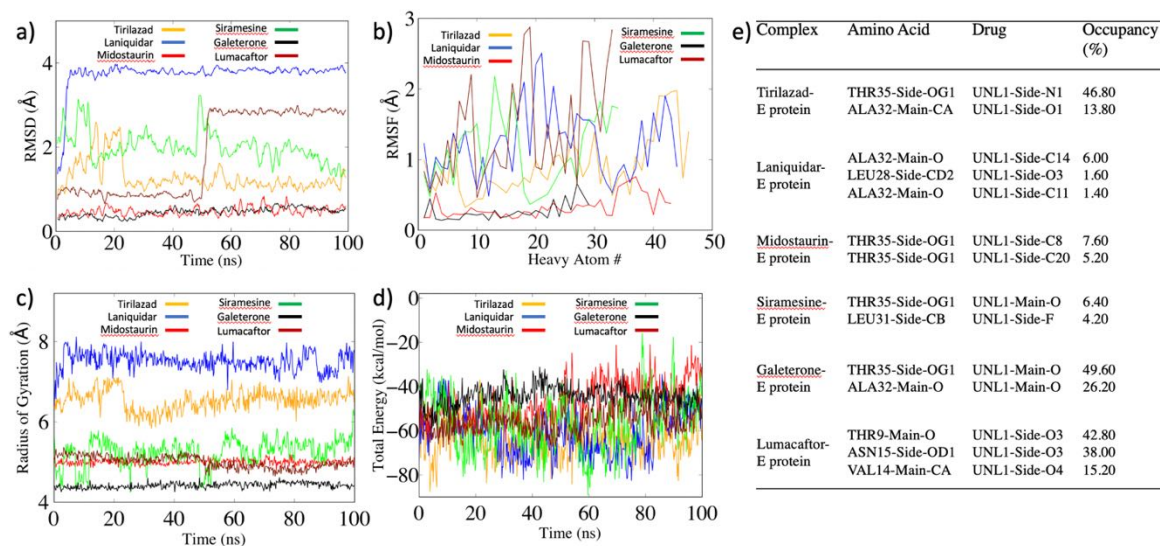
Figure 6.3 Results based on MD trajectory analysis. a) root mean square deviation (RMSD), b) root mean square fluctuation (RMSF), and c) radius of gyration. (RGYR) of the drug molecules. Panel d) and e) are showing NAMD energy and hydrogen bonds between drug compounds and E-proteins calculated with angle 30 degree and distance 3.5 Å as a cutoff.

The RMSD of the drug compound indicates the stability of the ligand inside the E-protein over time. Midostaurin and galeterone are showing the highest stability and experiencing a very little movement inside the E-protein. Tirilazad and siramesine are fluctuating moderately during the simulation. However, laniquidar at around 5 ns and lumacaftor at around 50 ns show a tremendous move and get settled immediately and remain in the stable state for rest of the simulation. The RMSF of the drug molecules show clear relation to the RMSD of them. Laniquidar and lumacaftor have the highest fluctuations in their heavy atoms, whereas galeterone and midostaurin heavy atoms are showing the lowest flexibility. As the RGYR of a chemical indicates the distribution of its atoms, it is good measure to see the degree of changes of the ligand atoms. I also calculated the RGYR for the ligands as shown in Figure 6.3(c), and all the drugs are showing a stable distribution of their atoms during simulations. The total energy and hydrogen bond were calculated with NAMD energy and Hydrogen bond plugin of VMD[94] depicted in Figure 6.3(d) and 6.3(e)

70

respectively. Both the calculations are showing a stable complex for E-protein and ligands. After confirming the stability of all the complexes, I performed the pharmacokinetics studies for them.

6.3   Pharmacokinetic Studies

Now that I have six stable ligands and E-protein complexes obtained from molecular docking and verified with MD simulation, the pharmacokinetics studies, which is important for safety assessment[178], of the drug molecules are being considered for further analysis. Lipinski's rule of 5[179], or Pfizer's rule of five is an investigative approach to identify the druglikeness of a chemical compound that has a definite pharmacological or biological activity. The rules were fixed depending on the remarks that oral drugs are usually relatively small with high lipophilicity[180]. The rule infers that compound having properties: molecular weight (< 500 Da), lipophilicity, LogP (< 5), hydrogen bond donor, HBD (< 5), and hydrogen bond acceptor, HBA (> 10), shows a better absorption and saturation property across a cell membrane. In addition to these properties, I also calculated % of absorption, AB%, manually with the following equation[181]:

$$AB\% = 109 - (0.345 \times TPSA)$$

The properties were assessed for all the six drug molecules as shown in Table 6.2 using an web-based application called molsoft[182] (https://molsoft.com/mprop/). The properties which are violated by the compounds are shown in red color in the table. Tirilazad, laniquidar, and midostaurin have a little high molecular weight, whereas others are well inside the limit. None of the compounds violate the hydrogen bond donor/acceptor criteria. However, the MolLogP property is the one violated by most the drugs but midostaurin.

The percentage of absorption is just fine for all the drugs and is indicating that when administered orally all the compounds possess a good property of bioavailability, distribution, and circulation[183]. While considering the druglikeness score, all the compounds are showing to be like a prior identified drug, although midostauring, galeterone, and lumacaftor are off by a fraction only and could less likely be a drug molecule[184]. However, overall, all the compounds show to be a good drug candidate for SARS-CoV-2 E protein.

Table 6.2 Lipinski's rule of 5 prediction of the selected compounds using molsoft web-based application.

Drugs	Molecular Weight (<500)	Hydrogen Bond Acceptor (<10)	Hydrogen Bond Donor (<5)	MolLogP (<5)	MolPSA/TPSA (≤140) A^2	Absorption Percentage (AB%) (>50%)	Drug likeness Score (>0)
Tirilazad	624.42	5	0	6.62	58.21	88.91	1.45
Laniquidar	584.28	6	0	6.06	51.47	91.24	1.37
Midostaurin	570.23	4	1	4.69	54.97	90.04	-0.01
Siramesine	454.24	2	0	6.37	13.95	104.19	1.18
Galeterone	388.25	2	1	5.19	26.65	99.81	-0.02
Lumacaftor	452.12	7	2	5.97	77.69	82.2	-0.52

I also calculated bioactivity of the chosen compounds using molinspiration website[185] against GPCR ligand, ion channel modulator, kinase inhibitor, nuclear receptor ligand, protease inhibitor, enzyme inhibitor as shown in Table 6.3, and the prediction read as active with bioactive score > 0, moderately active with bioactive score -5.0-0.0, and inactive with bioactive score < -5.0[185,186]. The study show that all the six compounds are active against GPCR ligand and kinase inhibitor with a positive bioactivity score. Only siramesine is showing active nature against ion channel modulator, whereas other five are moderately active against it. Siramesine, galeterone, lumacaftor are found to be active and

and tirilazad, laniquidar, and midostaurin are moderately active against nuclear receptor ligand. Except galeterone, all the compounds are calculated as active for protease inhibitor. The compounds are also active with positive bioactivity value except tirilazad and laniquidar against enzyme inhibitor.

Table 6.3. Bioactivity prediction of the selected drugs obtained using molinspiration website.

Drugs	GPCR Ligand	Ion Channel Modulator	Kinase Inhibitor	Nuclear Receptor Ligand	Protease Inhibitor	Enzyme Inhibitor
Tirilazad	0.01	-0.81	0.71	-0.02	0.03	-0.12
Laniquidar	0.32	-0.43	0.35	-0.34	0	-0.03
Midostaurin	0.2	-0.33	0.96	-0.45	0.18	0.39
Siramesine	0.78	0.22	0.29	0.21	0.17	0.22
Galeterone	0.22	-0.01	0.18	0.91	-0.08	0.86
Lumacaftor	0.21	-0.23	0.12	0.16	0.1	0.06

Developing a successful drug, there is a set of pharmacodynamic properties- appropriate drug uptake, reasonable drug distribution to corresponding tissues and organs, effective metabolization for reasonable activity, suitable elimination, and definite non-toxic manner, called ADMET properties, which are critical for manufacturing effective drugs[187-190]. The ADMET stands for absorption, distribution, metabolism, excretion, and toxicity. These properties were predicted for all the selected compounds using pkCSM webpage[191] among other web-based tools[192-196], and listed in Table 6.4. The measured parameters were human intestinal absorption (HIA), blood brain barrier (BBB) penetration, CYP substrate and inhibitor, total clearance, and rat oral acute toxicity (LD50) which represent absorption, distribution, metabolism, excretion, and toxicity respectively.

Table 6.4 Compounds' pharmacodynamics profile obtained using pkCSM website.

Compound	Absorption	Distribution	Metabolism	Excretion	Toxicity
	Human Intestinal Absorption (HIA) (% Absorbed)	Blood Brain Barrier (BBB) Penetration (logBB)	CYP Substrate/ Inhibitor	Total Clearance (log ml/min/kg)	Rat Oral Acute Toxicity LD50 (mol/kg)
Tirilazad	97.899	-0.534	Yes/No	0.116	2.873
Laniquidar	92.016	-0.505	Yes/Yes	0.935	2.672
Midostaurin	98.086	-0.802	Yes/No	-1.014	2.552
Siramesine	92.64	0.686	Yes/Yes	1.25	2.054
Galeterone	86.553	0.257	Yes/No	0.434	2.66
Lumacaftor	92.916	-0.368	Yes/Yes	-0.155	2.434

The calculated HIA property is indicating a higher absorption of all the compounds and seem to be highly absorbed in the intestinal cells[197]. Only siramesine is showing a better BBB penetration which is critical for the development of new drugs for brain diseases, and neuroimaging of brain[198]. The CYP450 stands for cytochrome P450 enzymes and are critical for drug biotransformation reactions in/on the smooth endoplasmic reticulum of different cells, e.g., liver. A CYP inhibitor impedes the biotransformation of drugs by CYP450 enzymes, thereby obstruct the metabolism of the drug. In this study, only tirilazad, midostaurin, and galeterone are found to be a CYP non-inhibitor. Except midostaurin and lumacaftor, other four compounds have a very good assessment of total clearance. The toxicity (LD50) values of all the compounds are high and indicating a nonlethal property of them, as lower the LD50 value higher is the toxicity.

7. CONCLUSIONS

I investigated the role of conformational changes of the SARS-CoV-2 Omicron variant's spike protein, hACE2 receptor protein, E protein, and the PfMATE transporter protein. Amino acid mutations that cause conformational changes in viral proteins can allow the virus to more efficiently bind to a host cell and allow the virus to evade existing therapeutics/vaccines.

The spike protein of SARS-CoV-2 requires the RBD of at least one chain in its trimer to be in the "up" conformation to scan the host cell receptor proteins and antibodies. Switching of the RBD conformation from the "down" conformation is a large structural change. The orientation of the sidechain of the amino acid(s) is also vital to find the proper match with the neighboring amino acids in the host cell's hACE2 receptor protein to form non-bonded interactions that allow viral material to enter(infect) the cell and cause disease in the host organism. Similarly, amino acids in the N-terminal (residues Glu8–Ile13) of the SARS-CoV-2 envelope protein show greater dynamic properties compared to the C-terminal and may play a critical role in binding Ca^{2+} and allowing the protein to pass out of the ERGIC lumen to the cytoplasm.

In the case of the PfMATE protein, both inward- and outward-facing structures undergo a huge structural change to switch to their isomer. These conformational changes in structure allow PfMATE to act as a transporter protein that expels antibiotic molecules before they can damage the bacterial cell. This results in drug resistance to medication, e.g., norfloxacin ,and thereby intensifies the economic and health-related burden of the patients.

In the Omicron project, my analyses show that the Omicron RBD shows a higher ACE2 binding affinity compared to the Delta RBD. Unlike other variants, Omicron has RBD mutations Y505H and S371L that lie in the region interacting with the surrounding spike protein domains in the closed-form spike trimer and this can affect the RBD opening. Increased ACE2 affinity and potentially easier and more efficient RBD opening, combined with antibody evasion due to mutation-induced antigenic shifts provide the Omicron strain with a significant increase in the probability of successful cellular attachment, and this may contribute to its dominance over Delta. Other factors such as ACE2 accessibility of the RBD due to spike mutations affecting the hinge flexibility[199-201], altered TMPRSS2 usage in cleaving S1/S2[202] as well as post-fusion conformational changes[203-205] may have effects on the transmissibility and severity of infection. While the Omicron RBD escapes most antibodies specific to other variants, it harbors sequences with significantly improved antigenicity compared to prior sequences. This suggests a possibility of superior neutralizing antibodies for Omicron and provides insights into vaccine design as well as a perspective on the future of SARS-CoV-2 persistence.

In addition to the Omicron study, I investigated the molecular level details of the structural transitions between the IF and OF conformations of PfMATE using all-atom molecular dynamics simulation techniques and thermodynamic and mechanistic analyses using Targeted Molecular Dynamics (TMD) computations. Calculations of potential of mean force (PMF) from the TMD simulations allowed us to track the change in free-energy as the system switches from OF to the more thermodynamically favorable IF upon protonation, and to track the change in free-energy as the system switches from IF to the thermodynamically more favorable OF upon the binding of Na^+. The atomic-level cause

of differences in thermodynamic stability and structural flexibility among the different PfMATE systems was traced to protonation-induced differences in hydrogen bonding on both the extracellular and intracellular sides of the PfMATE systems. The flexibility of specific amino acids changed noticeably upon protonation. Compared to their behavior in the OF(non-protonated) configuration, in OF(protonated) the residue D184 experienced a large decrease in RMSF of 35%, residue E331 had a 28% decrease in RMSF, and E163 had a decrease of 9%. Upon the binding of Na^+ in the IF structure, D41, N180, and D184 experience large decreases in RMSF of 35%, 27%, and 10%, respectively. These results imply that the structural transformations involve important amino acids becoming more rigid in their position, perhaps acting as fulcrums.

My Dynamical Network Analysis and Transfer Entropy Analysis showed how protonation could modify the cascading structural changes responsible for the transition between the IF and OF configurations. The stable OF(np) structure contains many inter-lobe and intra-lobe connections (26 total), whereas protonation changes the connectivity so that only a small number (8 total) of intra-lobe connections must be broken for OF(p) to convert. The IF(p) structure contains many inter-lobe connections that must be broken upon structural conversion. Consistent with Jagessar's proposal[27], E163 plays a crucial role in the structural transition from OF(p) to IF(p). Also, the charged state of several residues and ion binding near D41/N180/D184/T202 provides an easier conformational transition for the IF(Na^+) system by decreasing the node connections in communities. My transfer entropy calculations showed that protonation of OF allows E331 to take on more importance in affecting the structural arrangement of the amino acids in the protein. The

TE calculations also show that the binding of Na^+ significantly increases the TE donor strength of T202, N180, and D184 but causes D41 to become more of a TE receptor.

While finding the molecular level details of the dynamics of the SARS-CoV-2 Omicron RBD of the spike protein, RBD-hACE2 complex, and RBD-antibody, I also carried out research on finding therapeutics targeting the E protein. The SARS-CoV-2 E protein is a pentameric channel that is responsible for expelling out the Ca^{2+} from the ERGIC lumen to the cytoplasm of the host cell, and thereby starting the inflammation of the infected cells. It acts as a viroporin and takes part in modifying the cell membrane, resulting in the release of the viruses from the affected cells. The pathogenic activity of the CoVs can be reduced by introducing mutations in the E protein. Also, the E protein deletion causes the attenuation of the number of viruses as this action produces undeveloped and incompetent progeny viruses. Because of the multifunctionality of the E protein in CoVs, it becomes a potential drug target and can be a promising vaccine candidate against SARS-CoV-2. In this study, I used molecular docking methods to screen the best drug candidates for SARS-CoV-2 E protein from a list of 3800 FDA-approved and investigational drugs. The six docked drug-E protein complexes, tirlazad, laniquidar, midostaurin, siramesine, galeterone, and lumacaftor, were selected based on the binding energy-based docking score calculated with molecular docking software AutoDock Vina. The selected complexes were embedded in the ERGIC membrane and prepared for MD simulation by solvating and ionizing them. RMSD, RMSF, and RGYR calculations based on the MD simulation show that all the drugs bind well to the E protein. These results were also supported by NAMD energy and hydrogen bond calculation. Hydrophobic interactions of each of the selected drugs with the transmembrane part of the E protein add extra stability to the complexes.

To come up with a successful drug for a particular target, pharmacological studies of the drug molecules are essential to check the safety profile of the drug candidate. This study can help to avoid detrimental off-target activity profiles of drug targets. I assessed the Lipinski's rule of 5 to evaluate the druglikeness of the selected drugs and their pharmacodynamic property. Based on the Lipinski's rule of 5, all the compounds were promising drug candidates for the SARS-CoV-2 E protein. All the compounds seemed to be highly absorbed in the intestinal cells as per the HIA calculation. Only siramesine revealed a better BBB penetration. Tirilazad, midostaurin, and galeterone are found to be a CYP non-inhibitor. Four compounds- tirlazad, laniquidar, siramesine, and galeterone, showed a very good assessment of total clearance. All six compounds showed a nonlethal property as their measured toxicity (LD50) values were high. The study listed six drugs to inhibit the channel activity of SRAS-CoV-2 E viroporin. These recommended repurposed pharmaceuticals and nutraceuticals require further validation of their action in clinical settings. The compounds that perform well in clinical trial may be used against the covid-19 pandemic.

REFERENCES

1   Alberts, B. The cell as a collection of protein machines: preparing the next generation of molecular biologists. *cell* **92**, 291-294 (1998).

2   Flechsig, H. & Mikhailov, A. S. Simple mechanics of protein machines. *Journal of the Royal Society Interface* **16**, 20190244 (2019).

3   Rehman, I., Kerndt, C. C. & Botelho, S. Biochemistry, Tertiary Protein Structure. (2017).

4   Chaffey, N.   (Oxford University Press, 2003).

5   Bajaj, M. & Blundell, T. Evolution and the tertiary structure of proteins. *Annual review of biophysics and bioengineering* **13**, 453-492 (1984).

6   Branden, C. I. & Tooze, J. *Introduction to protein structure*. (Garland Science, 2012).

7   Degrado, W. F. Design of peptides and proteins. *Advances in protein chemistry* **39**, 51-124 (1988).

8   Ponomarenko, E. A. *et al.* The size of the human proteome: the width and depth. *International journal of analytical chemistry* **2016** (2016).

9   Genton, L. & Pichard, C. Protein catabolism and requirements in severe illness. *International Journal for Vitamin and Nutrition Research* **81**, 143 (2011).

10   Elango, R. & Ball, R. O. Protein and amino acid requirements during pregnancy. *Advances in nutrition* **7**, 839S-844S (2016).

11   Yeung, S. E., Hilkewich, L., Gillis, C., Heine, J. A. & Fenton, T. R. Protein intakes are associated with reduced length of stay: a comparison between Enhanced Recovery After Surgery (ERAS) and conventional care after elective colorectal surgery. *The American journal of clinical nutrition* **106**, 44-51 (2017).

12   Nowson, C. & O'Connell, S. Protein requirements and recommendations for older people: a review. *Nutrients* **7**, 6874-6899 (2015).

13   Mason, J. M. & Arndt, K. M. Coiled coil domains: stability, specificity, and biological implications. *Chembiochem* **5**, 170-176 (2004).

14   Khoshnoodi, J., Pedchenko, V. & Hudson, B. G. Mammalian collagen IV. *Microscopy research and technique* **71**, 357-370 (2008).

15   Muiznieks, L. D., Weiss, A. S. & Keeley, F. W. Structural disorder and dynamics of elastin. *Biochemistry and Cell Biology* **88**, 239-250 (2010).

16   Cooper, G. M.   (Sinauer Associates, 2000).

17   Cuesta, S. M., Rahman, S. A., Furnham, N. & Thornton, J. M. The classification and evolution of enzyme function. *Biophysical journal* **109**, 1082-1086 (2015).

18   Li, P., Yin, Y.-L., Li, D., Kim, S. W. & Wu, G. Amino acids and immune function. *British Journal of Nutrition* **98**, 237-252 (2007).

19   Forthal, D. N. Functions of antibodies. *Microbiology spectrum* **2**, 2.4. 21 (2014).

20   Sperelakis, N. *Cell physiology sourcebook: a molecular approach*.   (Elsevier, 2001).

21   Hamm, L. L., Nakhoul, N. & Hering-Smith, K. S. Acid-base homeostasis. *Clinical Journal of the American Society of Nephrology* **10**, 2232-2242 (2015).

22   Lamanda, A., Cheaib, Z., Turgut, M. D. & Lussi, A. Protein buffering in model systems and in whole human saliva. *PLoS One* **2**, e263 (2007).

23   Aoi, W. & Marunaka, Y. Importance of pH homeostasis in metabolic health and diseases: crucial role of membrane proton transport. *BioMed research international* **2014** (2014).

24   Rui, L. Energy metabolism in the liver. *Comprehensive physiology* **4**, 177 (2014).

25   Gerstein, M., Lesk, A. M. & Chothia, C. Structural mechanisms for domain movements in proteins. *Biochemistry* **33**, 6739-6749 (1994).

26   Perutz, M. F. Mechanisms of cooperativity and allosteric regulation in proteins. *Quarterly reviews of biophysics* **22**, 139-237 (1989).

27   Fischer, E. Einfluss der Configuration auf die Wirkung der Enzyme. *Berichte der deutschen chemischen Gesellschaft* **27**, 2985-2993 (1894).

28   Bacon, F. On the nature of allosteric transitions: a plausible model. *J Mol Biol* **12**, 88-118 (1965).

29   Rini, J. M., Schulze-Gahmen, U. & Wilson, I. A. Structural evidence for induced fit as a mechanism for antibody-antigen recognition. *Science* **255**, 959-965 (1992).

30   Williamson, J. R. Induced fit in RNA–protein recognition. *Nature structural biology* **7**, 834-837 (2000).

31 Hammes, G. G., Chang, Y.-C. & Oas, T. G. Conformational selection or induced fit: a flux description of reaction mechanism. *Proceedings of the National Academy of Sciences* **106**, 13737-13741 (2009).

32 Jones, E. Y., Fugger, L., Strominger, J. L. & Siebold, C. MHC class II proteins and disease: a structural perspective. *Nature Reviews Immunology* **6**, 271-282 (2006).

33 den Hollander, A. I., Roepman, R., Koenekoop, R. K. & Cremers, F. P. Leber congenital amaurosis: genes, proteins and disease mechanisms. *Progress in retinal and eye research* **27**, 391-419 (2008).

34 Taylor, J. P., Hardy, J. & Fischbeck, K. H. Toxic proteins in neurodegenerative disease. *science* **296**, 1991-1995 (2002).

35 Ranjan, P., Dey, A., Sharma, V. P. & Tiwari, N. K. in *Biomedical Applications of Natural Proteins*   101-113 (Springer, 2015).

36 Ciotti, M. *et al.* The COVID-19 pandemic. *Critical reviews in clinical laboratory sciences* **57**, 365-388 (2020).

37 Lone, S. A. & Ahmad, A. COVID-19 pandemic–an African perspective. *Emerging microbes & infections* **9**, 1300-1308 (2020).

38 Watkins, J. Vol. 368   (British Medical Journal Publishing Group, 2020).

39 Baral, P. *et al.* Mutation-induced changes in the receptor-binding interface of the SARS-CoV-2 Delta variant B. 1.617. 2 and implications for immune evasion. *Biochemical and biophysical research communications* **574**, 14-19 (2021).

40 Park, M. D. Immune evasion via SARS-CoV-2 ORF8 protein? *Nature Reviews Immunology* **20**, 408-408 (2020).

41 Lazarevic, I., Pravica, V., Miljanovic, D. & Cupic, M. Immune evasion of SARS-CoV-2 emerging variants: what have we learnt so far? *Viruses* **13**, 1192 (2021).

42 McCallum, M. *et al.* SARS-CoV-2 immune evasion by the B. 1.427/B. 1.429 variant of concern. *Science* **373**, 648-654 (2021).

43 Mandala, V. S. *et al.* Structure and drug binding of the SARS-CoV-2 envelope protein transmembrane domain in lipid bilayers. *Nature structural & molecular biology* **27**, 1202-1208 (2020).

44 Nieto-Torres, J. L. *et al.* Severe acute respiratory syndrome coronavirus E protein transports calcium ions and activates the NLRP3 inflammasome. *Virology* **485**, 330-339 (2015).

45   Schoeman, D. & Fielding, B. C. Coronavirus envelope protein: current knowledge. *Virology journal* **16**, 1-22 (2019).

46   Tanaka, Y. *et al.* Structural basis for the drug extrusion mechanism by a MATE multidrug transporter. *Nature* **496**, 247-251 (2013).

47   He, X. *et al.* Structure of a cation-bound multidrug and toxic compound extrusion transporter. *Nature* **467**, 991-994 (2010).

48   Kaatz, G. W., McAleese, F. & Seo, S. M. Multidrug resistance in Staphylococcus aureus due to overexpression of a novel multidrug and toxin extrusion (MATE) transport protein. *Antimicrobial agents and chemotherapy* **49**, 1857-1864 (2005).

49   McAleese, F. *et al.* A novel MATE family efflux pump contributes to the reduced susceptibility of laboratory-derived Staphylococcus aureus mutants to tigecycline. *Antimicrobial agents and chemotherapy* **49**, 1865-1871 (2005).

50   Chitsaz, M. & Brown, M. H. The role played by drug efflux pumps in bacterial multidrug resistance. *Essays in biochemistry* **61**, 127-139 (2017).

51   Cao, Y. *et al.* Potent neutralizing antibodies against SARS-CoV-2 identified by high-throughput single-cell sequencing of convalescent patients' B cells. *Cell* **182**, 73-84. e16 (2020).

52   Callaway, E. & Ledford, H. How bad is Omicron? What scientists know so far. *Nature* **600**, 197-199 (2021).

53   Hoffmann, M. *et al.* The Omicron variant is highly resistant against antibody-mediated neutralization: Implications for control of the COVID-19 pandemic. *Cell* **185**, 447-456. e411 (2022).

54   Mallapaty, S. *et al.* How COVID vaccines shaped 2021 in eight powerful charts. *Nature*, doi:10.1038/d41586-021-03686-x (2021).

55   Rattner, N. in *CNBC*   (2021).

56   Planas, D. *et al.* Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. *Nature*, doi:10.1038/s41586-021-03777-9 (2021).

57   Leung, N. H. Transmissibility and transmission of respiratory viruses. *Nature reviews microbiology* **19**, 528-545 (2021).

58   Cai, Y. *et al.* Structural basis for enhanced infectivity and immune evasion of SARS-CoV-2 variants. *Science*, doi:10.1126/science.abi9745 (2021).

59   Alizon, S. & Sofonea, M. T. SARS-CoV-2 virulence evolution: avirulence theory, immunity and trade-offs. *Journal of Evolutionary Biology* **34**, 1867-1877 (2021).

60   Du, D., van Veen, H. W., Murakami, S., Pos, K. M. & Luisi, B. F. Structure, mechanism and cooperation of bacterial multidrug transporters. *Current opinion in structural biology* **33**, 76-91 (2015).

61   Blair, J. M. A., Webber, M. A., Baylay, A. J., Ogbolu, D. O. & Piddock, L. J. V. Molecular mechanisms of antibiotic resistance. *Nature reviews microbiology* **13**, 42-51 (2015).

62   Higgins, C. F. Multiple molecular mechanisms for multidrug resistance transporters. *Nature* **446**, 749-757 (2007).

63   Fischbach, M. A. & Walsh, C. T. Antibiotics for emerging pathogens. *Science* **325**, 1089-1093 (2009).

64   Hall, W. *Superbugs: An arms race against bacteria*. (Harvard University Press, 2018).

65   Hassan, K. A. *et al.* Pacing across the membrane: the novel PACE family of efflux pumps is widespread in Gram-negative pathogens. *Research in microbiology* **169**, 450-454 (2018).

66   Hvorup, R. N. *et al.* The multidrug/oligosaccharidyl-lipid/polysaccharide (MOP) exporter superfamily. *European Journal of Biochemistry* **270**, 799-813 (2003).

67   Brown, M. H., Paulsen, I. T. & Skurray, R. A. The multidrug efflux protein NorM is a prototype of a new family of transporters. *Molecular microbiology* **31**, 394-395 (1999).

68   Omote, H., Hiasa, M., Matsumoto, T., Otsuka, M. & Moriyama, Y. The MATE proteins as fundamental transporters of metabolic and xenobiotic organic cations. *Trends in Pharmacological Sciences* **27**, 587-593 (2006).

69   He, G.-X. *et al.* An H+-coupled multidrug efflux pump, PmpM, a member of the MATE family of transporters, from Pseudomonas aeruginosa. *Journal of bacteriology* **186**, 262-265 (2004).

70   Jin, Y., Nair, A. & van Veen, H. W. Multidrug transport protein NorM from Vibrio cholerae simultaneously couples to sodium-and proton-motive force. *Journal of Biological Chemistry* **289**, 14624-14632 (2014).

71   Kuroda, T. & Tsuchiya, T. Multidrug efflux transporters in the MATE family. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* **1794**, 763-768 (2009).

72   Cdc. Coronavirus Disease 2019 (COVID-19). *Centers for Disease Control and Prevention* (2020).

73   Saito, A. *et al.* Enhanced fusogenicity and pathogenicity of SARS-CoV-2 Delta P681R mutation. *Nature* **602**, 300-306 (2022).

74   Zahradník, J. *et al.* SARS-CoV-2 variant prediction and antiviral drug design are enabled by RBD in vitro evolution. *Nature microbiology* **6**, 1188-1198 (2021).

75   Wise, J. Covid-19: The E484K mutation and the risks it poses. *BMJ* **372**, n359, doi:10.1136/bmj.n359 (2021).

76   Zakrzewska, S. *et al.* Inward-facing conformation of a multidrug resistance MATE family transporter. *Proceedings of the National Academy of Sciences* **116**, 12275-12284 (2019).

77   Lu, M. Structures of multidrug and toxic compound extrusion transporters and their mechanistic implications. *Channels* **10**, 88-100 (2016).

78   Pervushin, K. *et al.* Structure and inhibition of the SARS coronavirus envelope protein ion channel. *PLoS pathogens* **5**, e1000511 (2009).

79   Torres, J. *et al.* Model of a putative pore: the pentameric α-helical bundle of SARS coronavirus E protein in lipid bilayers. *Biophysical journal* **91**, 938-947 (2006).

80   Li, Y., Surya, W., Claudine, S. & Torres, J. Structure of a conserved Golgi complex-targeting signal in coronavirus envelope proteins. *Journal of Biological Chemistry* **289**, 12535-12549 (2014).

81   UniProt: the universal protein knowledgebase in 2021. *Nucleic acids research* **49**, D480-D489 (2021).

82   Li, W. *et al.* Bats are natural reservoirs of SARS-like coronaviruses. *Science* **310**, 676-679 (2005).

83   DeDiego, M. L. *et al.* A severe acute respiratory syndrome coronavirus that lacks the E gene is attenuated in vitro and in vivo. *Journal of virology* **81**, 1701-1713 (2007).

84   Nieto-Torres, J. L. *et al.* Severe acute respiratory syndrome coronavirus envelope protein ion channel activity promotes virus fitness and pathogenesis. *PLoS pathogens* **10**, e1004077 (2014).

85   Sztain, T. *et al.* A glycan gate controls opening of the SARS-CoV-2 spike protein. *Nat Chem* **13**, 963-968, doi:10.1038/s41557-021-00758-3 (2021).

86   Walls, A. C. *et al.* Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* **181**, 281-292. e286 (2020).

87   Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: a web-based graphical user interface for CHARMM. *Journal of computational chemistry* **29**, 1859-1865 (2008).

88   J. Lee, X. C., J.M. Swails, M.S. Yeom, P.K. Eastman, J.A. Lemkul, S. Wei, J. Buckner, J.C. Jeong, Y. Qi, S. Jo, V.S. Pande, D.A. Case, C.L. Brooks III, A.D. MacKerell Jr, J.B. Klauda, and W. Im. CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations using the CHARMM36 Additive Force Field. *J. Chem. Theory Comput* **12**, 405-413 (2016).

89   Mannar, D. *et al.* SARS-CoV-2 Omicron Variant: ACE2 Binding, Cryo-EM Structure of Spike Protein-ACE2 Complex and Antibody Evasion. *bioRxiv*, 2021.2012.2019.473380, doi:10.1101/2021.12.19.473380 (2021).

90   Wang, Y. *et al.* Structural basis for SARS-CoV-2 Delta variant recognition of ACE2 receptor and broadly neutralizing antibodies. *Nature Communications* **13**, 1-12 (2022).

91   Baral, P. *et al.* Mutation-induced changes in the receptor-binding interface of the SARS-CoV-2 Delta variant B.1.617.2 and implications for immune evasion. *Biochem Biophys Res Commun* **574**, 14-19, doi:10.1016/j.bbrc.2021.08.036 (2021).

92   Bhattarai, N., Baral, P., Gerstman, B. S. & Chapagain, P. P. Structural and Dynamical Differences in the Spike Protein RBD in the SARS-CoV-2 Variants B. 1.1. 7 and B. 1.351. *The Journal of Physical Chemistry B* (2021).

93   Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature protocols* **10**, 845 (2015).

94   Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *Journal of molecular graphics* **14**, 33-38 (1996).

95   Jagessar, K. L., Claxton, D. P., Stein, R. A. & McHaourab, H. S. Sequence and structural determinants of ligand-dependent alternating access of a MATE transporter. *Proceedings of the National Academy of Sciences* **117**, 4732-4740 (2020).

96   Fiala, G. & Stetter, K. O. Pyrococcus furiosus sp. nov. represents a novel genus of marine heterotrophic archaebacteria growing optimally at 100 C. *Archives of Microbiology* **145**, 56-61 (1986).

97   Valentine, D. L. Adaptations to energy stress dictate the ecology and evolution of the Archaea. *Nature reviews microbiology* **5**, 316-323 (2007).

98   De Rosa, M., Gambacorta, A. & Gliozzi, A. Structure, biosynthesis, and physicochemical properties of archaebacterial lipids. *Microbiological reviews* **50**, 70 (1986).

99   Balleza, D., Garcia-Arribas, A. B., Sot, J., Ruiz-Mirazo, K. & Goñi, F. M. Ether- versus ester-linked phospholipid bilayers containing either linear or branched apolar chains. *Biophysical journal* **107**, 1364-1374 (2014).

100   Damsté, J. S. S., Schouten, S., Hopmans, E. C., van Duin, A. C. T. & Geenevasen, J. A. J. Crenarchaeol the characteristic core glycerol dibiphytanyl glycerol tetraether membrane lipid of cosmopolitan pelagic crenarchaeota. *Journal of lipid research* **43**, 1641-1651 (2002).

101   Koga, Y. & Morii, H. Recent advances in structural research on ether lipids from archaea including comparative and physiological aspects. *Bioscience, biotechnology, and Biochemistry* **69**, 2019-2034 (2005).

102   Sprott, G. D., Agnew, B. J. & Patel, G. B. Structural features of ether lipids in the archaeobacterial thermophiles Pyrococcus furiosus, Methanopyrus kandleri, Methanothermus fervidus, and Sulfolobus acidocaldarius. *Canadian Journal of Microbiology* **43**, 467-476 (1997).

103   Berman, H. M. *et al.* The protein data bank. *Nucleic acids research* **28**, 235-242 (2000).

104   Rihn, S. J. *et al.* A plasmid DNA-launched SARS-CoV-2 reverse genetics system and coronavirus toolkit for COVID-19 research. *PLoS biology* **19**, e3001091 (2021).

105   Waterhouse, A. *et al.* SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic acids research* **46**, W296-W303 (2018).

106   Guex, N., Peitsch, M. C. & Schwede, T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. *Electrophoresis* **30**, S162-S173 (2009).

107   Jo, S., Kim, T. & Im, W. Automated builder and database of protein/membrane complexes for molecular dynamics simulations. *PloS one* **2**, e880 (2007).

108   Wu, E. L. *et al.*   (Wiley Online Library, 2014).

109   Mark, P. & Nilsson, L. Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *The Journal of Physical Chemistry A* **105**, 9954-9960 (2001).

110   Phillips, J. *et al.* and Schulten, K. 2005. *Scalable molecular dynamics with NAMD. Journal of Computational Chemistry* **26**, 1781-1802 (2005).

111   Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021).

112   Varadi, M. *et al.* AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic acids research* **50**, D439-D444 (2022).

113   Pokhrel, R., Chapagain, P. & Siltberg-Liberles, J. Potential RNA-dependent RNA polymerase inhibitors as prospective therapeutics against SARS-CoV-2. *Journal of medical microbiology* **69**, 864 (2020).

114   Pihan, E., Colliandre, L., Guichou, J.-F. & Douguet, D. e-Drug3D: 3D structure collections dedicated to drug repurposing and fragment-based drug design. *Bioinformatics* **28**, 1540-1541 (2012).

115   O'Boyle, N. M. *et al.* Open Babel: An open chemical toolbox. *Journal of cheminformatics* **3**, 1-14 (2011).

116   Trott, O. & Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry* **31**, 455-461 (2010).

117   Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *Journal of computational chemistry* **26**, 1781-1802 (2005).

118   Huang, J. & MacKerell Jr, A. D. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *Journal of computational chemistry* **34**, 2135-2145 (2013).

119   Huang, J. & MacKerell Jr, A. D. Force field development and simulations of intrinsically disordered proteins. *Current opinion in structural biology* **48**, 40-48 (2018).

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

121   Essmann, U. *et al.* A smooth particle mesh Ewald method. *The Journal of chemical physics* **103**, 8577-8593 (1995).

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

123   Vanommeslaeghe, K. *et al.* CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of computational chemistry* **31**, 671-690 (2010).

124   Best, R. B. *et al.* Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain χ1 and χ2 dihedral angles. *Journal of chemical theory and computation* **8**, 3257-3273 (2012).

125   Kern, N. R. *et al.* Lipid-linked oligosaccharides in membranes sample conformations that facilitate binding to oligosaccharyltransferase. *Biophysical journal* **107**, 1885-1895 (2014).

126   Klauda, J. B. *et al.* Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *The journal of physical chemistry B* **114**, 7830-7843 (2010).

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

128   Feller, S. E., Zhang, Y., Pastor, R. W. & Brooks, B. R. Constant pressure molecular dynamics simulation: the Langevin piston method. *The Journal of chemical physics* **103**, 4613-4621 (1995).

129   Wang, Q., Xue, T., Song, C., Wang, Y. & Chen, G. Study on the Application of the Combination of TMD Simulation and Umbrella Sampling in PMF Calculation for Molecular Conformational Transitions. *International journal of molecular sciences* **17**, 692 (2016).

130   Hénin, J. & Chipot, C. Overcoming free energy barriers using unconstrained molecular dynamics simulations. *The Journal of chemical physics* **121**, 2904-2914 (2004).

131   Pearlman, D. A. A comparison of alternative approaches to free energy calculations. *The Journal of Physical Chemistry* **98**, 1487-1493 (1994).

132   Smith, D. E. & Haymet, A. D. J. Free energy, entropy, and internal energy of hydrophobic interactions: Computer simulations. *The Journal of chemical physics* **98**, 6445-6454 (1993).

133   Tobias, D. J. & Brooks Iii, C. L. Calculation of free energy surfaces using the methods of thermodynamic perturbation theory. *Chemical physics letters* **142**, 472-476 (1987).

134   Gc, J. B., Gerstman, B. S. & Chapagain, P. P. The role of the interdomain interactions on RfaH dynamics and conformational transformation. *The journal of physical chemistry B* **119**, 12750-12759 (2015).

135   Sethi, A., Eargle, J., Black, A. A. & Luthey-Schulten, Z. Dynamical networks in tRNA: protein complexes. *Proceedings of the National Academy of Sciences* **106**, 6620-6625 (2009).

136   McClendon, C. L., Kornev, A. P., Gilson, M. K. & Taylor, S. S. Dynamic architecture of a protein kinase. *Proceedings of the National Academy of Sciences* **111**, E4623-E4631 (2014).

137   Girvan, M. & Newman, M. E. J. Community structure in social and biological networks. *Proceedings of the national academy of sciences* **99**, 7821-7826 (2002).

138   Glykos, N. M. Software news and updates carma: A molecular dynamics analysis program. *Journal of computational chemistry* **27**, 1765-1768 (2006).

139   Kamberaj, H. & van der Vaart, A. Extracting the causality of correlated motions from molecular dynamics simulations. *Biophysical Journal* **97**, 1747-1755 (2009).

140   Haliloglu, T., Bahar, I. & Erman, B. Gaussian dynamics of folded proteins. *Physical review letters* **79**, 3090 (1997).

141   Hacisuleyman, A. & Erman, B. Entropy Transfer and Allosteric Communication in Proteins. *J. Drug Des. Res* **4**, 1054 (2017).

142   Garcia Michel, L. R., Keirns, C. D., Ahlbrecht, B. C. & Barr, D. A. Calculating Transfer Entropy from Variance–Covariance Matrices Provides Insight into Allosteric Communication in ERK2. *Journal of chemical theory and computation* **17**, 3168-3177 (2021).

143   Bakan, A., Meireles, L. M. & Bahar, I. ProDy: protein dynamics inferred from theory and experiments. *Bioinformatics* **27**, 1575-1577 (2011).

144   Hacisuleyman, A. & Erman, B. Causality, transfer entropy, and allosteric communication landscapes in proteins with harmonic interactions. *Proteins: Structure, Function, and Bioinformatics* **85**, 1056-1064 (2017).

145   Turoňová, B. *et al.* In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by three hinges. *Science* **370**, 203-208 (2020).

146   Pierri, C. L. SARS-CoV-2 spike protein: flexibility as a new target for fighting infection. *Signal Transduction and Targeted Therapy* **5**, 1-3 (2020).

147   Wrapp, D. *et al.* Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **367**, 1260-1263 (2020).

148   Shang, J. *et al.* Structural basis of receptor recognition by SARS-CoV-2. *Nature* **581**, 221-224 (2020).

149 Wang, Q. *et al.* Structural and functional basis of SARS-CoV-2 entry by using human ACE2. *Cell* **181**, 894-904. e899 (2020).

150 Karathanou, K. *et al.* A graph-based approach identifies dynamic H-bond communication networks in spike protein S of SARS-CoV-2. *J Struct Biol* **212**, 107617, doi:10.1016/j.jsb.2020.107617 (2020).

151 Sztain, T. *et al.* A glycan gate controls opening of the SARS-CoV-2 spike protein. *Nature Chemistry* **13**, 963-968 (2021).

152 Baral, P., Pavadai, E., Gerstman, B. S. & Chapagain, P. P. In-silico identification of the vaccine candidate epitopes against the Lassa virus hemorrhagic fever. *Sci Rep* **10**, 7667, doi:10.1038/s41598-020-63640-1 (2020).

153 Irini A Doytchinova, D. R. F. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics* (2007).

154 Mercurio, I., Tragni, V., Busto, F., De Grassi, A. & Pierri, C. L. Protein structure analysis of the interactions between SARS-CoV-2 spike protein and the human ACE2 receptor: from conformational changes to novel neutralizing antibodies. *Cellular and Molecular Life Sciences* **78**, 1501-1522 (2021).

155 Lan, J. *et al.* Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* **581**, 215-220 (2020).

156 Yang, L. J. *et al.* Corilagin prevents SARS-CoV-2 infection by targeting RBD-ACE2 binding. *Phytomedicine* **87**, 153591 (2021).

157 Jawad, B., Adhikari, P., Podgornik, R. & Ching, W.-Y. Key interacting residues between RBD of SARS-CoV-2 and ACE2 receptor: Combination of molecular dynamics simulation and density functional calculation. *Journal of chemical information and modeling* **61**, 4425-4441 (2021).

158 Han, P. *et al.* Receptor binding and complex structures of human ACE2 to spike RBD from Omicron and Delta SARS-CoV-2. *Cell* (2022).

159 Kim, S. *et al.* Binding of Human ACE2 and RBD of Omicron Enhanced by Unique Interaction Patterns Among SARS-CoV-2 Variants of Concern. *bioRxiv* (2022).

160 Kim, S. *et al.* Differential Interactions between Human ACE2 and Spike RBD of SARS-CoV-2 Variants of Concern. *J Chem Theory Comput* **17**, 7972-7979, doi:10.1021/acs.jctc.1c00965 (2021).

161 Wu, L. *et al.* SARS-CoV-2 Omicron RBD shows weaker binding affinity than the currently dominant Delta variant to human ACE2. *Signal Transduct Target Ther* **7**, 8, doi:10.1038/s41392-021-00863-2 (2022).

162   Tragni, V. *et al.* Modeling SARS-CoV-2 spike/ACE2 protein–protein interactions for predicting the binding affinity of new spike variants for ACE2, and novel ACE2 structurally related human protein targets, for COVID-19 handling in the 3PM context. *EPMA Journal*, 1-27 (2022).

163   Mannar, D. *et al.* SARS-CoV-2 Omicron Variant: ACE2 Binding, Cryo-EM Structure of Spike Protein-ACE2 Complex and Antibody Evasion. *bioRxiv* (2021).

164   Benton, D. J. *et al.* Receptor binding and priming of the spike protein of SARS-CoV-2 for membrane fusion. *Nature* **588**, 327-330 (2020).

165   Genovese, L., Zaccaria, M., Farzan, M., Johnson, W. & Momeni, B. Investigating the mutational landscape of the SARS-CoV-2 Omicron variant via ab initio quantum mechanical modeling. *bioRxiv* (2021).

166   Lupala, C. S., Ye, Y., Chen, H., Su, X. D. & Liu, H. Mutations on RBD of SARS-CoV-2 Omicron variant result in stronger binding to human ACE2 receptor. *Biochem Biophys Res Commun* **590**, 34-41, doi:10.1016/j.bbrc.2021.12.079 (2021).

167   Huang, Y. *et al.* SARS-CoV-2 spike binding to ACE2 is stronger and longer ranged due to glycan interaction. *Biophysical journal* **121**, 79-90 (2022).

168   Marcelo C R Melo, R. C. B., Cesar de la Fuente-Nunez, Zaida Luthey-Schulten. Generalized correlation-based dynamical network analysis: a new high-performance approach for identifying allosteric communications in molecular dynamics trajectories. *J Chem Phys* **153(13)**, doi:https://doi.org/10.1063/5.0018980 (2020).

169   Smart, O. S., Neduvelil, J. G., Wang, X., Wallace, B. A. & Sansom, M. S. P. HOLE: a program for the analysis of the pore dimensions of ion channel structural models. *Journal of molecular graphics* **14**, 354-360 (1996).

170   Chen, H., Wu, Y. & Voth, G. A. Proton transport behavior through the influenza A M2 channel: insights from molecular simulation. *Biophysical journal* **93**, 3470-3479 (2007).

171   Sula, A. & Wallace, B. A. Interpreting the functional role of a novel interaction motif in prokaryotic sodium channels. *Journal of General Physiology* **149**, 613-622 (2017).

172   Laghaei, R., Kowallis, W., Evans, D. G. & Coalson, R. D. Calculation of iron transport through human H-chain ferritin. *The Journal of Physical Chemistry A* **118**, 7442-7453 (2014).

173   Darve, E. & Pohorille, A. Calculating free energies using average force. *The Journal of chemical physics* **115**, 9169-9183 (2001).

174   Borkotoky, S. & Banerjee, M. A computational prediction of SARS-CoV-2 structural protein inhibitors from Azadirachta indica (Neem). *Journal of Biomolecular Structure and Dynamics* **39**, 4111-4121 (2021).

175   Bhowmik, D. *et al.* Identification of potential inhibitors against SARS-CoV-2 by targeting proteins responsible for envelope formation and virion assembly using docking based virtual screening, and pharmacokinetics approaches. *Infection, Genetics and Evolution* **84**, 104451 (2020).

176   Laskowski, R. A. & Swindells, M. B.   (ACS Publications, 2011).

177   Kalé, L. *et al.* NAMD2: greater scalability for parallel molecular dynamics. *Journal of Computational Physics* **151**, 283-312 (1999).

178   Williams, P. D. The role of pharmacological profiling in safety assessment. *Regulatory toxicology and pharmacology* **12**, 238-252 (1990).

179   Lipinski, C. A. Lead-and drug-like compounds: the rule-of-five revolution. *Drug discovery today: Technologies* **1**, 337-341 (2004).

180   Lipinski, C. A., Lombardo, F., Dominy, B. W. & Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced drug delivery reviews* **23**, 3-25 (1997).

181   Zhao, Y. H. *et al.* Rate-limited steps of human oral absorption and QSAR studies. *Pharmaceutical research* **19**, 1446-1457 (2002).

182   Molsoft, L.   (2018).

183   Ghose, A. K., Viswanadhan, V. N. & Wendoloski, J. J. A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A qualitative and quantitative characterization of known drug databases. *Journal of combinatorial chemistry* **1**, 55-68 (1999).

184   Chandrasekaran, K. & Thilak Kumar, R. Molecular properties prediction, docking studies and antimicrobial screening of ornidazole and its derivatives. *J. Chem. Pharm. Res* **8**, 849-861 (2016).

185   Grob, S. *mib – Molinspiration molecular processing engine, version 2007.10, Molinspiration Cheminformatics*, <http://www.molinspiration.com>

186   Ungell, A.-L. In vitro absorption studies and their relevance to absorption from the GI tract. *Drug development and industrial pharmacy* **23**, 879-892 (1997).

187   Cumming, J. G., Davis, A. M., Muresan, S., Haeberlein, M. & Chen, H. Chemical predictive modelling to improve compound quality. *Nature reviews Drug discovery* **12**, 948-962 (2013).

188   Van De Waterbeemd, H. & Gifford, E. ADMET in silico modelling: towards prediction paradise? *Nature reviews Drug discovery* **2**, 192-204 (2003).

189   Hou, T. & Wang, J. Structure–ADME relationship: still a long way to go? *Expert Opinion on Drug Metabolism & Toxicology* **4**, 759-770 (2008).

190   Hodgson, J. ADMET—turning chemicals into drugs. *Nature biotechnology* **19**, 722-726 (2001).

191   Pires, D. E., Blundell, T. L. & Ascher, D. B. pkCSM: predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures. *Journal of medicinal chemistry* **58**, 4066-4072 (2015).

192   Xiong, G. *et al.* ADMETlab 2.0: an integrated online platform for accurate and comprehensive predictions of ADMET properties. *Nucleic Acids Research* **49**, W5-W14 (2021).

193   Daina, A., Michielin, O. & Zoete, V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Scientific reports* **7**, 1-13 (2017).

194   Cheng, F. *et al.*   (ACS Publications, 2012).

195   Schyman, P., Liu, R., Desai, V. & Wallqvist, A.   (2017).

196   Venkatraman, V. FP-ADMET: a compendium of fingerprint-based ADMET prediction models. *Journal of cheminformatics* **13**, 1-12 (2021).

197   Gautret, P. *et al.* Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial. *International journal of antimicrobial agents* **56**, 105949 (2020).

198   Pardridge, W. M. Drug transport across the blood–brain barrier. *Journal of cerebral blood flow & metabolism* **32**, 1959-1972 (2012).

199   Plante, J. A. *et al.* Spike mutation D614G alters SARS-CoV-2 fitness. *Nature* **592**, 116-121 (2021).

200   Daniloski, Z. *et al.* The Spike D614G mutation increases SARS-CoV-2 infection of multiple human cell types. *Elife* **10**, e65365 (2021).

201   Zhang, L. *et al.* The D614G mutation in the SARS-CoV-2 spike protein reduces S1 shedding and increases infectivity. bioRxiv. 2020. *Google Scholar*.

202   Meng, B. *et al.* Altered TMPRSS2 usage by SARS-CoV-2 Omicron impacts tropism and fusogenicity. *Nature*, 1-1 (2022).

203   Fan, X., Cao, D., Kong, L. & Zhang, X. Cryo-EM analysis of the post-fusion structure of the SARS-CoV spike glycoprotein. *Nature communications* **11**, 1-10 (2020).

204   Tai, L. *et al.* Nanometer-resolution in situ structure of the SARS-CoV-2 postfusion spike protein. *Proceedings of the National Academy of Sciences* **118** (2021).

205   Cai, Y. *et al.* Distinct conformational states of SARS-CoV-2 spike protein. *Science* **369**, 1586-1592 (2020).

VITA

MD LOKMAN HOSSEN

Born, Barishal, Bangladesh

2008-2012	B.Sc., Physics Jahangirnagar University Savar, Bangadesh
2012-2013	M.S., Physics Jahangirnagar University Savar, Bangadesh
2014-2017	Lecturer, Physics University of Barishal Barishal, Bangladesh
2017-2022	Doctoral Candidate Florida International University Miami, Florida Graduate Teaching Assistant Florida International University Miami, Florida

PUBLICATIONS AND PRESENTATIONS

1. Hossen, M. L., Baral, P., Sharma, T., Gerstman, B., & Chapagain, P. (2022). *Significance of the RBD mutations in the SARS-CoV-2 Omicron: From spike opening to antibody escape and cell attachment*. Physical Chemistry Chemical Physics, 24(16), 9123-9129.

2. Baral, P., Bhattarai, N., Hossen, M. L., Stebliankin, V., Gerstman, B. S., Narasimhan, G., & Chapagain, P. P. (2021). *Mutation-induced changes in the receptor-binding interface of the SARS-CoV-2 Delta variant B. 1.617. 2 and implications for immune evasion*. Biochemical and biophysical research communications, 574, 14-19.

3. Bhattarai, N., Pavadai, E., Pokhrel, R., Baral, P., Hossen, M. L., Stahelin, R. V., ... & Gerstman, B. S. (2022). *Ebola virus protein VP40 binding to Sec24c for transport to the plasma membrane*. Proteins: Structure, Function, and Bioinformatics, 90(2), 340-350.

4. Fouque, K. J. D., Sipe, S. N., Garabedian, A., Mejia, G., Su, L., Hossen, M. L., ... & Fernandez-Lima, F. (2022). *Exploring the Conformational and Binding Dynamics of HMGA2· DNA Complexes Using Trapped Ion Mobility Spectrometry-Mass Spectrometry*. Journal of the American Society for Mass Spectrometry.

5. Barrios, A., Diaz, M. M., Perezo, E., Hossen, M. L., Chapagain, P., & Moon, J. H. (2022). *Effects of Sidechain Isomerism on Polymer-based Non-Covalent Protein Delivery*. Chemical Communications.

6. Nader, D., Gressett, T. E., Hossen, M. L., Chapagain, P. P., Kerrigan, S. W., & Bix, G. (2022). A dual-receptor mechanism between integrins and ACE2 widens SARS-CoV-2 tissue tropism. bioRxiv.

7. Hossen, M. L., Chapagain, P. P., & Gerstman, B. (2020). *Calculating the Binding Free Energy Difference between Conformational Changes of AT-Rich DNA Sequences*. Biophysical Journal, 118(3), 218a.

8. Hossen, M. L., Gerstman, B. S., & Chapagain, P. P. (2022). *Alternating access switching in molecular dynamics simulations of PfMATE transporter*. Biophysical Journal, 121(3), 250a-251a.

9. Hossen, M. L., Bhattarai, N., Gerstman, B., & Chapagain, P. (2021). *Molecular Dynamics Investigations of Inward-facing and Outward-facing Isomerization of PfMATE*. Bulletin of the American Physical Society, 66.