Rita Isabel Carvalho Teixeira

Licenciatura em Bioquímica

The impact of SARS-CoV-2 Omicron variant on the receptor binding domain conformational dynamics and interaction with human ACE2

Dissertação para obtenção do Grau de Mestre em
Bioquímica para a Saúde

Orientador: Diana Lousa, PhD, ITQB-NOVA
Co-orientador: João B. Vicente, PhD, ITQB-NOVA

Novembro, 2022

Rita Isabel Carvalho Teixeira

Licenciatura em Bioquímica

The impact of SARS-CoV-2 Omicron variant on the receptor binding domain conformational dynamics and interaction with human ACE2

Dissertação para obtenção do Grau de Mestre em
Bioquímica para a Saúde

Orientador: Diana Lousa, PhD, ITQB-NOVA
Co-orientador: João B. Vicente, PhD, ITQB-NOVA

Júri:

Presidente: Prof. Doutora Maria Teresa Nunes Mangas Catarino
Arguente: Prof. Doutor António Manuel Simões Carrão Albuquerque Baptista
Vogais: Doutora Diana Andreia Pereira Lousa
Prof. Doutora Ana Maria Varela Coelho

Instituto de Tecnologia Química e Biológica António Xavier da Universidade Nova de Lisboa

Novembro, 2022

...

The impact of SARS-CoV-2 Omicron variant on the receptor binding domain conformational dynamics and interaction with human ACE2

Acknowledgments

First of all, I would like to thank my supervisors, Dr. Diana Lousa and Dr. João B. Vicente, and the head of the Protein Modeling Laboratory, Professor Cláudio Soares, for giving me the opportunity to develop this project, in collaboration with the Applied Biochemistry group. During the last year, they were really helpful, patient, and available for anything, even in reading this thesis until the last minute.

A special thanks to Dr. Diana Lousa for all her guidance and time spent teaching me, her attention to detail, and her care for her students. In the last year, mostly due to the emergence of the Omicron variant, we worked side-by-side to gather as much information as we could, and I couldn't be more grateful for working with such a great and respectful woman.

I would like to thank my honorary supervisor, Mariana Valério, for accompanying me from the beginning of this journey to the end, even though she is abroad. Most of this work couldn't be done without her guidance and time spent not only teaching me so much about computational biology but also giving advice and supporting me in all aspects It is because of her that I started to be more practical and searching for my answers by "googling it".

I could not forget about my other honorary supervisors, Professor Cláudio Soares, Dr. Manuel N. Melo, Dr. Carlos Cruz, Dr. Caio Souza, and, Dr. Luís Borges-Araújo, for all the shared knowledge and for all the support during this time.

I would also like to thank everyone in the Protein Modeling, Molecular Simulation, and Multiscale Modeling groups, especially Susana, João, Rodrigo, Gonçalo, Fernando, and Carolina, for their infinite patience, their friendship, and support in the toughest moments. Without them, this journey would not be the same.

To my dearest best friends for all the unconditional support and friendship, that listen to all my work problems and eared more about COVID-19 from me than from the news in the last year and a half.

Lastly, I would like to thank my family for always supporting me in every situation, especially my father that never stopped believing in me and deserves the world for being there for me in good and bad times. He is my "rock" and I will keep working hard every day to make his effort worthwhile.

Abstract

The COVID-19 pandemic caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has led to over 6.6 million deaths worldwide, as of 27 November 2022.

The SARS-CoV-2 infection mechanism involves the viral binding to the angiotensin-converting enzyme 2 (ACE2) host receptor through the Spike (S) protein receptor-binding domain (RBD). Thus, understanding the molecular details of binding is of pharmaceutical interest. Nevertheless, the unbound RBD presents conformational changes in the “ridge” region that are hypothesized to also affect receptor recognition.

The emergence of variants of concern (VOC) poses a great challenge due to the loss of natural and vaccine immunity. The rise of the Omicron VOC raised considerable global concern due to the amount of S protein substitutions, fifteen of which are located in the RBD. Later on, the Omicron variant has been divided into seven lineages, of which BA.1, BA.2, and BA.5 became the most concerning ones.

The main purpose of this thesis is to understand the impact of these Omicron subvariants on the binding to the human ACE2 (hACE2). To this end, long atomistic molecular dynamics (MD) simulations with Omicron RBDs bound to hACE2 were performed in order to study the direct role of substitutions in the RBD binding interface with the receptor. This allowed us to identify specific substitutions that may affect binding affinity via the establishment of new interprotein interactions.

A complementary study of the impact of Omicron subvariants on the RBD conformational dynamics was also performed by simulating the unbound BA.2 and BA.5 RBDs in solution. The results showed that both Omicron subvariants impact RBD conformational dynamics towards an efficient binding to hACE2. Additionally, the BA.2 subvariant presented an alternative “shrunken” conformation which could provide a fitness advantage by increasing hACE2 accessibility and by facilitating antibody escape.

Keywords: SARS-CoV-2, Receptor-binding domain (RBD), human Angiotensin-converting enzyme 2 (hACE2), Variant of Concern (VOC), Omicron, Molecular Dynamics (MD) simulation

Resumo

A pandemia COVID-19, causada pelo coronavírus da Síndrome Respiratória Aguda Grave 2 (SARS-CoV-2), foi responsável por mais de 6,6 milhões de mortes mundialmente, a 27 novembro de 2022.

O mecanismo de infeção do SARS-CoV-2 envolve a ligação do vírus ao receptor do hospedeiro, a enzima conversora de angiotensina 2 (ACE2), através do domínio de ligação ao recetor (RBD) da proteína Spike (S). Logo, o modo de ligação tem interesse farmacêutico. No entanto, o RBD livre apresenta alterações conformacionais na região de ligação ao receptor que poderá também afetar o reconhecimento do mesmo.

O aparecimento de variantes de preocupação (VOC) representa um desafio devido à perda da imunidade obtida naturalmente ou através da vacinação. O aparecimento da variante Omicron suscitou uma considerável preocupação globalmente devido à quantidade de substituições presentes na proteína S, quinze das quais estão localizadas no RBD. Mais tarde, a variante Omicron foi dividida em sete grandes linhagens, das quais BA.1, BA.2, e BA.5 tornaram-se as mais preocupantes.

O principal objetivo desta tese é compreender o impacto destas sub-variantes da Omicron na ligação à ACE2 humana (hACE2). Para tal, foram realizadas longas simulações de dinâmica molecular atomística (MD) com os RBDs das sub-variantes da Omicron ligados à hACE2 para estudar o papel direto das substituições na interface de ligação do RBD com o recetor. Isto permitiu identificar substituições específicas que podem afetar a afinidade de ligação através do estabelecimento de novas interações intermoleculares.

Foi também estudado o efeito das sub-variantes BA.2 e BA.5 na dinâmica conformacional do RBD através de simulações dos RBDs livres em solução. Os resultados mostraram que estas sub-variantes têm impacto nesta dinâmica de modo a ligar eficientemente à hACE2. Além disso, a sub-variante BA.2 apresentou uma conformação alternativa - "shrunken" – que poderia aumentar a acessibilidade à hACE2 e facilitar a fuga a anticorpos.

Palavras-chave: SARS-CoV-2, Domínio de ligação ao receptor (RBD), Enzima conversora da angiotensina-2 humana (hACE2), Variante de preocupação (VOC), Omicron, Simulação de Dinâmica Molecular (MD)

Index

Figure Index

Table Index

List of abbreviations

+ssRNA – Positive-sense single-stranded RNA

ACE2 – Angiotensin-converting Enzyme 2

AMBER – Assisted Model Building and Energy Refinement

Cα – Alpha carbons

CHARMM – Chemistry at Harvard Macromolecular Mechanics

CLD – Collectrin-like domain

COVID-19 – Coronavirus 2019 Disease

CoV – Coronavirus

Cryo-EM – Cryo-electron Microscopy

CT – Cytoplasmatic Tail

CTD – C-terminal Domain

DPP4 – Dipeptidyl Peptidase 4

E protein – Envelope protein

ER – Endoplasmic reticulum

ERGIC – Endoplasmic reticulum-Golgi intermediate compartment

FP – Fusion peptide

fs – femtosecond

GROMOS - Groningen Molecular Simulation

hCoV – Human Coronavirus

HR1/2 – Heptad Repeat 1/2

kb – kilobases

LJ – Leonard-Jones

μs – microsecond

MERS – Middle East Respiratory Syndrome

MD – Molecular Dynamics

MM – Molecular Mechanics

M protein – Membrane protein

nAb – Neutralizing antibody

N protein – Nucleocapsid protein

NP – Nucleoprotein

ns – nanosecond

NSP – Non-structural protein

NTD – N-terminal domain

ORFs – Open Reading Frames

PBC – Periodic Boundary Conditions

PCA – Principal Component Analysis

PCR – Polymerase Chain Reaction

PD – Protease catalytic domain

PDB – Protein Data Bank

PME – Particle Mesh Ewald

ps – picosecond

QM – Quantum Mechanics

RBD – Receptor Binding Domain

RBM – Receptor-binding Motif

RIN – Residue Interaction Network

RMSD – Root-mean-square Deviation

RMSF – Root-mean-square Fluctuations

RNA – Ribonucleic Acid

RTC – Replication-transcription complex

S1/2 – Subunit 1/2

SARS – Severe Acute Respiratory Syndrome

SARS-CoV-2 – Severe Acute Respiratory Syndrome Coronavirus 2

SASA – Solvent Accessible Surface Area

S protein – Spike Protein

SPR – Surface Plasmon Resonance

TIP3P – Transferable intermolecular potential Three-point

TM – Transmembrane Domain

TMPRSS2 – Transmembrane Protease Serine 2

VOC – Variants of Concern

VOI – Variants of Interest

VUM - Variants under monitoring

WHO – World Health Organization

1. Introduction

1.1 The COVID-19 Pandemic

In late 2019, an outbreak of pneumonia that began in the capital the Chinese province of Hubei, Wuhan, was related to a novel strain of Coronaviruses, designated initially as 2019-nCoV by the World Health Organization (WHO). Later, the International Committee on Taxonomy of Viruses officially renamed it the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). The WHO named the viral illness Coronavirus Disease 2019 (COVID-19), which is often characterized by several symptoms such as cough, fever, respiratory distress, fatigue/tiredness, pneumonia, and muscle pain. [2–4]

Following the exponential rise of infection cases with SARS-CoV-2 in China, the viral outbreak that raised international concern about public health rapidly escalated into a global pandemic at the beginning of 2020. During the past few years, the COVID-19 pandemic has had a devasting impact at several levels worldwide. Not only it has brought serious health crises, but it also resulted in significant social adversities due to the global community restrictions during several waves of the pandemic. Furthermore, the subsequent economic crisis generated a vulnerable condition in developing countries. According to WHO, by November 27th, 2022, over 630 million confirmed cases and more than 6.6 million deaths had been attributed to the COVID-19 pandemic globally (Figure **1.1**). [5,6]

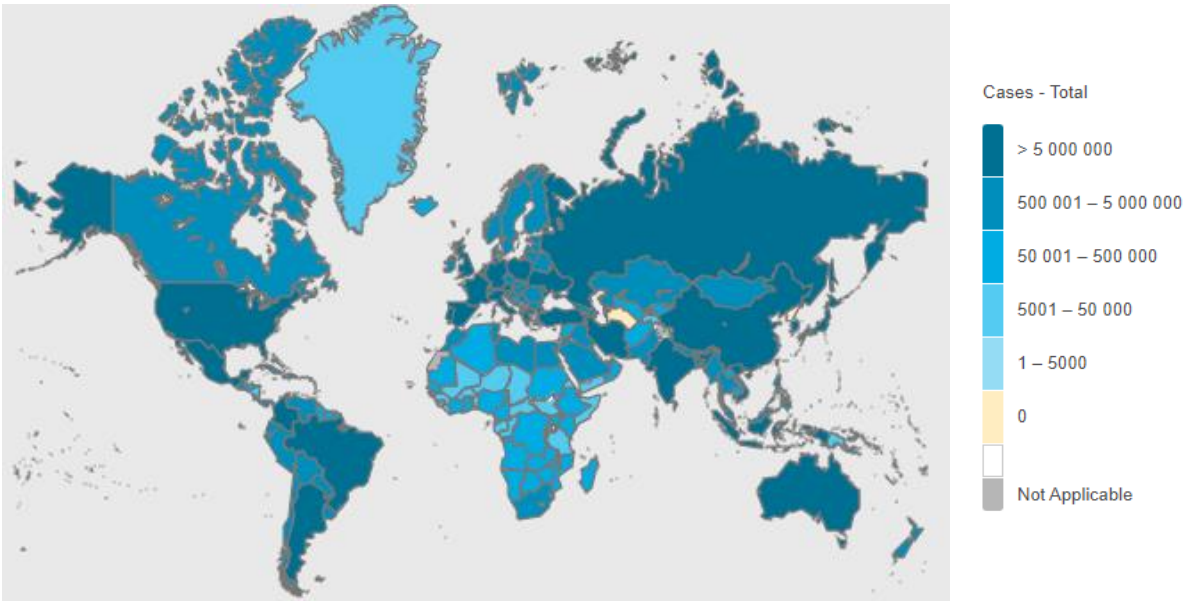


Figure 1.1 – World COVID-19 dashboard regarding the number of confirmed cases by November 27th, 2022. (Adapted from ref. 5)

Nevertheless, the collaborative efforts of the scientific community, industry, government, regulatory authorities, and foundations resulted in a quick and efficient reaction to the pandemic evolution. Due to decades of public and private sector investments in science and technology, it was possible to: develop vaccine rapid response technologies that allowed for the production of several vaccine candidates within months of the sequencing of SARS-CoV-2; to identify specific conformations of the Spike (S) glycoprotein, the main protein on the surface of coronaviruses, using prototype strain coronaviruses; and develop drugs for emerging coronaviruses.[7]

The vaccination has proven to induce an immune response that significantly reduces hospitalizations and deaths, even for SARS-CoV-2 variants that have emerged over time. As of November 27th, 2022, almost 13 billion doses have been administered globally, and more than 70% of the world population has received at least one dose of a COVID-19 vaccine.[8] Similarly, antibody treatments and developed small-molecule drugs are proving effective for preventing severe disease. Yet, the virus continues to spread across the globe, with possible emerging variants. In the long term, vaccines with broader and more durable protection and antibody antiviral therapies against infectious disease outbreaks are needed.[9]

1.2 Human Coronaviruses: outbreaks and similar biological features

Coronaviruses (CoVs) belong to a large family of enveloped viruses, *Coronaviridae*, with a positive-sense single-stranded RNA (+ssRNA) genome of 26-32 kilobases (kb). The organization of the viral genome is quite similar, consisting of two open reading frames (ORFs) 1a/b, at the 5' end, encoding for 16 non-structural proteins, and other ORFs at the 3' end encoding four major structural proteins: the S glycoprotein, that mediates the binding to the host cell receptors; the envelope (E) protein, which enables the virion assembly and release; the membrane (M) protein, responsible for the virion shape; and the nucleocapsid (N) protein, involved in RNA genome packing. When observed under electron microscopy, the Spike projections from the viral membrane provide a crown-like appearance to CoVs, thus the term 'coronavirus'. [10–12]

SARS-CoV-2, the virus responsible for COVID-19, is a virion with 60-140 nm in diameter, comprising a +ssRNA 29.9 kb RNA genome. The viral genome possesses two ORFs, ORF1a and ORF1b, which encode for the non-structural proteins (NSPs 1-16), and accounts for 67% of the total genome. The remaining genomic region contains the ORFs that encode the structural proteins. Additionally, the genome encodes for other accessory proteins (Figure **1.2**). [10–12]

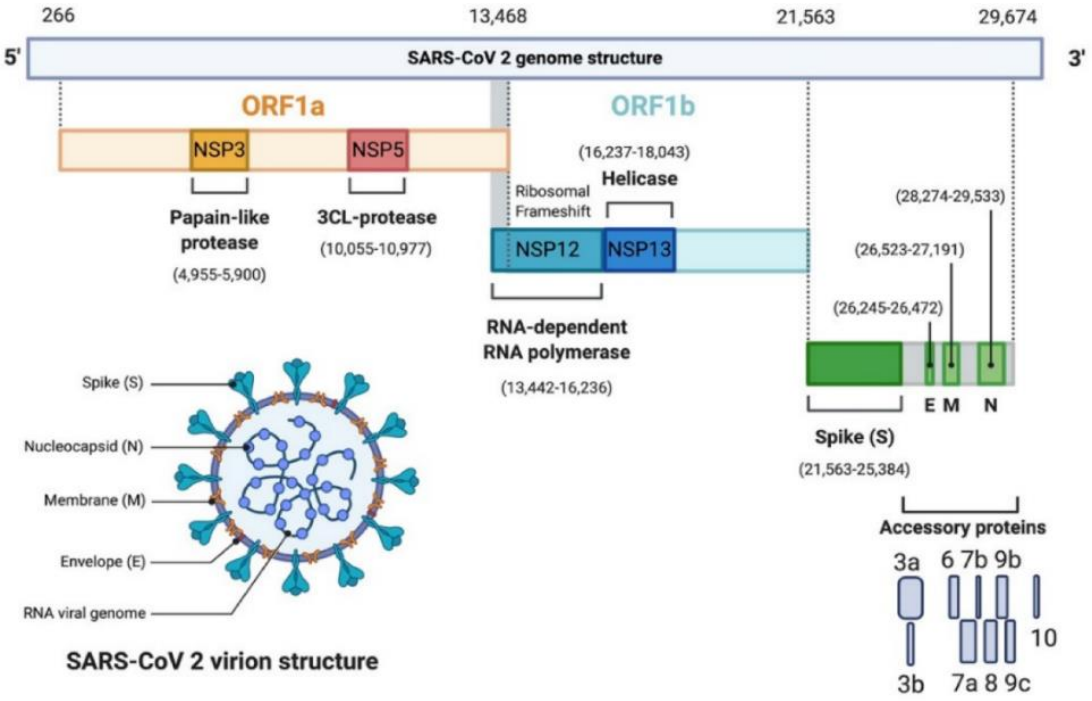


Figure 1.2 - Genomic organization of SARS-CoV-2. The genome consists of two genes: ORF1a (orange) and ORF1b, that encode for the 16 non-structural proteins (NSPs); the structural gene encodes for the structural proteins: spike (S), envelope (E), membrane (M), and nucleocapsid (N) (green); and the accessory protein-encoding genes. The scheme with SARS-CoV-2 structure is also represented. (Adapted from ref. 12)

Phylogenetically, these viruses are categorized into four genera: alpha-CoV, beta-CoV, gamma-CoV, and delta-CoV. The group of the beta-CoV genus has been subdivided into four lineages: Embecovirus (A), Sarbecovirus (B), Merbecovirus (C), and Nobecovirus (D) (Figure **1.3A**).[13]

Since the 1960s, CoVs are known to cause mild to moderate upper-respiratory tract illness in humans. Nowadays, out of seven types of human coronaviruses (hCoVs), three are highly pathogenic: SARS-CoV, which in 2002 caused an outbreak of severe acute respiratory syndrome (SARS), MERS-CoV, which emerged in 2012 to cause the Middle East respiratory syndrome (MERS), and SARS-CoV-2, the viral agent responsible for the current COVID-19 pandemic.[14,15]

Despite the differences in terms of the severe respiratory symptoms and mortality caused by these three viruses, they share common molecular features that enable findings obtained with specific viruses to be extrapolated to others and to future emerging strains.[7] According to genomic analysis, these coronaviruses belong to the group of the beta-CoV of the family *Coronaviridae*. SARS-CoV belongs to lineage B, while MERS-CoV is included in lineage C.[13] Sequence alignments revealed that SARS-CoV-2 is more closely related to SARS-CoV (identity ~80%) than MERS-CoV (identity ~50%), falling within lineage B (Figure **1.3B**).[16]

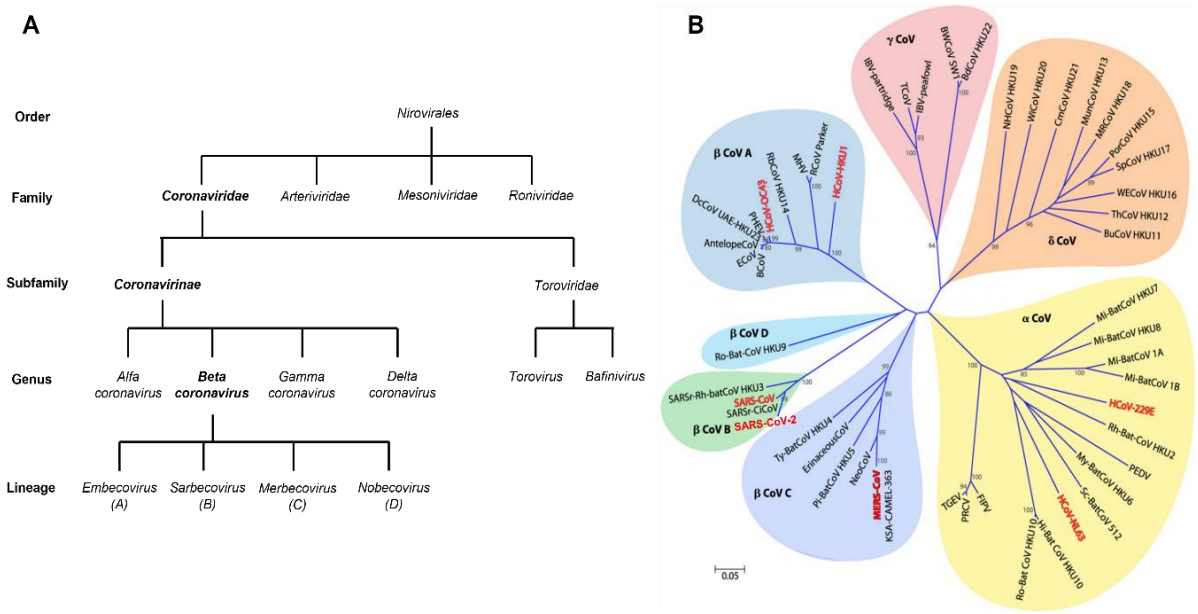


Figure 1.3 – (A) Taxonomy of *Coronavidae* according to the International Committee on Taxonomy of Viruses, and **(B)** phylogenetic tree of coronaviruses, wherein human coronaviruses (hCoVs) are marked in red, including SARS-CoV, MERS-CoV, and SARS-CoV-2. (Adapted from ref. 13).

These beta-CoVs are all zoonotic agents that emerged from natural animal reservoirs posited to be bats.[17] SARS-CoV was detected in China's wet markets. A CoV strain isolated from wild palm civets purchased in those markets shared 99.8% identity with SARS-CoV. Thus, palm civets were considered to be an intermediate host, causing the human-to-human transmission[18]. However, this situation is yet unclear.

As for MERS-CoV, there are indications that bats are likely the source of the virus. An RNA fragment isolated from bat stool demonstrated 100% nucleotide identity with a MERS-CoV-infected patient in the same region. Additionally, bats showed the ability to replicate the virus without developing symptoms, which suggested that bats were an ideal MERS-CoV reservoir. The dromedary camel was thought to be an intermediate reservoir since a 99.2–99.5% genomic identity of MERS-CoV isolated from dromedary camels and humans was subsequently confirmed.[12,18]

Similar to SARS-CoV, the emergence of SARS-CoV-2 may be linked to wet markets. SARS-CoV and Bat-CoV share a strong genomic similarity (96.2%). Therefore, it was suggested that bat was also the natural reservoir for this CoV. Furthermore, the pangolin-CoV and SARS-CoV-2 share 91.02% genomic similarity. After these findings, the question of which is the natural reservoir for SARS-CoV-2 remains. [12,18,19]

Regarding the mode of transmission among humans, the main transmission route is close contact by inhalation of droplets and aerosols or by contact with contaminated surfaces. Of the Three, SARS-CoV-2 is the one with the highest transmissibility, followed by SARS-CoV and MERS-CoV.[18,20]

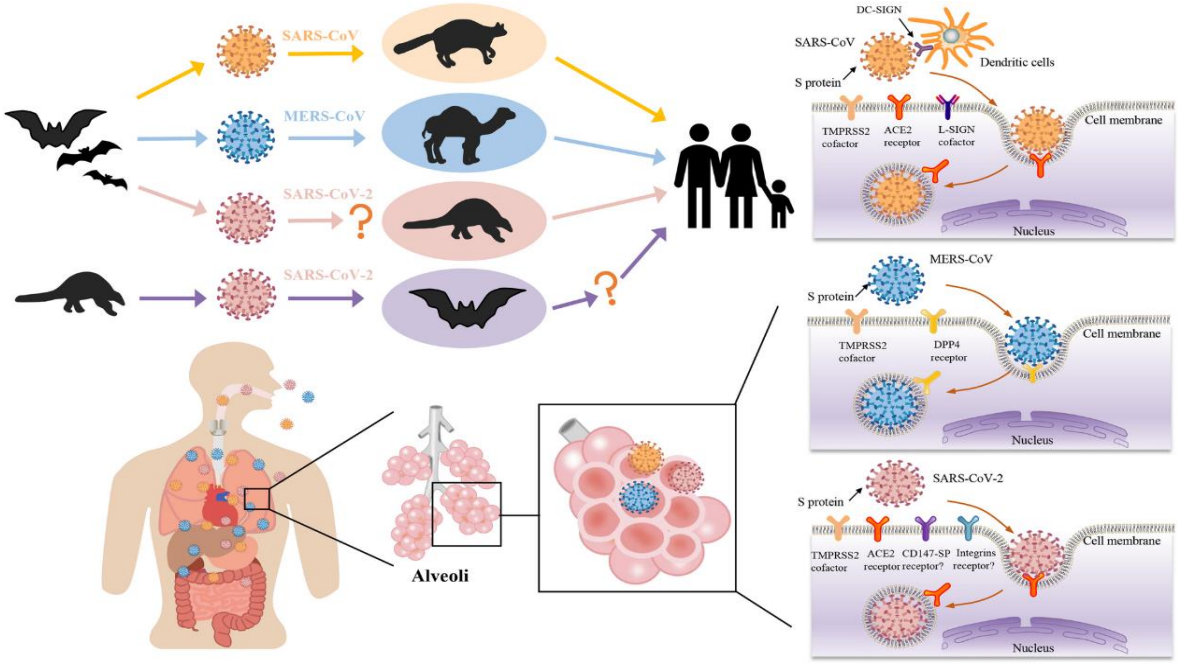


Figure 1.4 – Schematic of the potential animal reservoirs, and the biodistribution and host cell receptors of SARS-CoV, MERS-CoV, and SARS-CoV-2. (Adapted from ref. 18)

Considering the entrance mode of these viruses, they all rely on a host-cell receptor. For both SARS-CoV and SARS-CoV-2, the dominant receptor is the angiotensin-converting enzyme 2 (ACE2), whereas for MERS-CoV, the obligate receptor is the dipeptidyl peptidase-4 (DPP4) (Figure **1.4**). After cell infection, the incubation period is similar for the three viruses (4 to 6 days). Moreover, similar symptoms begin to occur, such as fever, sore throat, cough, dyspnea, fatigue, diarrhea, and headache. The diagnosis is normally performed through polymerase chain reaction (PCR) with viral RNA from clinical samples due to its sensitivity and specificity. [15,18,19]

1.3 SARS-CoV-2: Mechanism of transmission and pathogenesis

The transmission of SARS-CoV-2 mainly occurs by person-to-person airborne transmission, via droplets or contacts. The entrance of SARS-CoV-2 into host cells is a critical multi-step mechanism for viral infection, mediated by the S protein. The initial step is the recognition of its receptors on the surface of host cells, which in this case is the ACE2, the same receptor for the SARS-CoV S protein. The S protein, which is assembled as a homotrimer, comprises two non-covalently bound subunits in each monomer. The S1 domain contains the receptor-binding domain (RBD), which attaches to the host cell receptor, whereas the S2 subunit anchors the S protein to the host membrane and mediates the fusion of the viral host cell membranes. [21,22]

Following the host cell receptor engagement, the determinant step for the infection process is the activation of the S protein. This activation requires two cleavage events of the S protein, carried out by different host proteases: one at the junction site of S1 and S2 subunits, comprising a multibasic sequence – S1/S2, and the other at a site internal to the S2 subunit –

S2'. The S1/S2 boundary site is cleaved by a furin protease during virus maturation in the infected cells. [21–23]

After binding to the ACE2 receptor, the S protein S1 subunit experiences conformational changes that allow the S2' cleavage site to become exposed to target-cell proteases. Depending on the SARS-CoV-2 entry mode, the S2′ site is cleaved by either transmembrane protease serine 2 (TMPRSS2) at the cell surface, or by cathepsins in the endosomes following endocytosis mediated by ACE2. In the presence of TMPRSS2, S2' cleavage occurs at the target-cell surface. Although SARS-CoV-2 prefers the cell surface entry, if the target cell expresses insufficient TMPRSS2 or the virus-ACE2 complex does not encounter this protease, this complex will be internalized via clathrin-mediated endocytosis, where cleavage of the S2' site is carried out by cathepsins, requiring an acidic environment for their activity (Figure **1.5**). [23,24]

In both pathways, the fusion peptide (FP) is exposed and the dissociation of S1 from the S2 subunit leads to a conformational change of the latter in a way that allows FP to be inserted into the host membrane and to form a fusion pore. Once the fusion process between the viral and the cellular membrane is completed, viral genetic material is released into the host cells and replicated. [23]

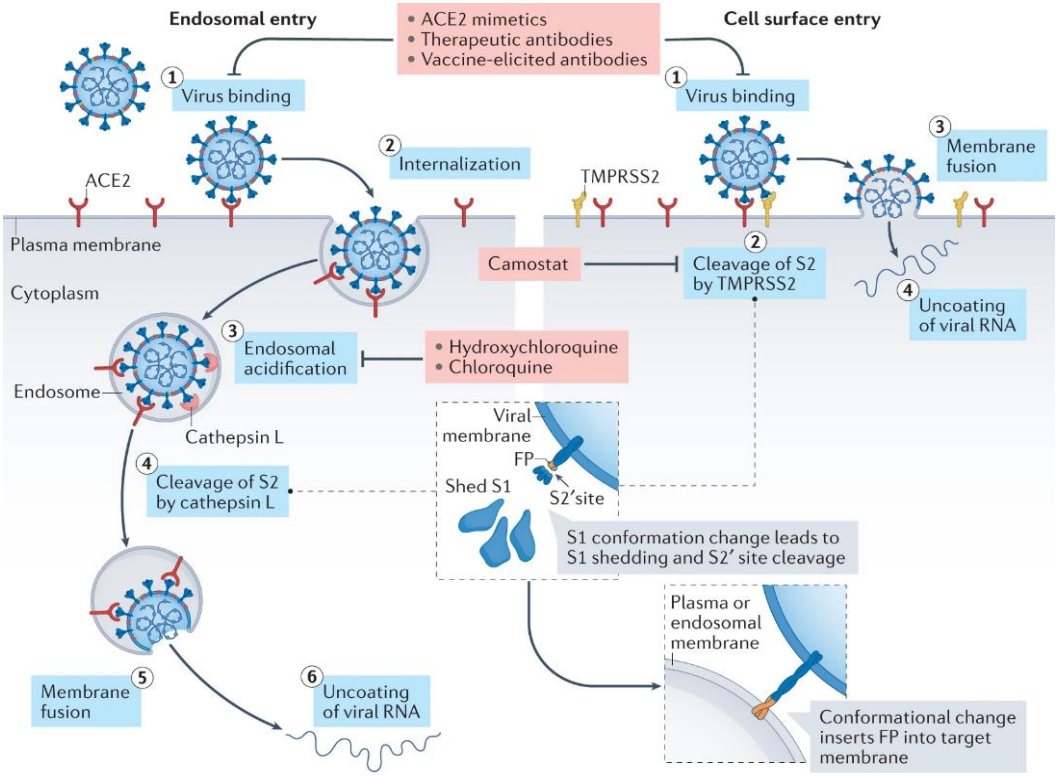


Figure 1.5 – Scheme of the two entry pathways of SARS-CoV-2 into host cells. After ACE2 binding (**1**), the virus can enter through the endosomal entry (left), where the virus-ACE2 complex is internalized via endocytosis (**2**) followed by the cathepsin-mediated cleavage of the S2' site, which requires an acidic environment for their activity (**3** and **4**). The cell surface entry (right) occurs in presence of TMPRSS2 that cleave the S2' site (**2**). Both entry modes allow the virus to initiate the fusion process between the viral and the cellular membranes, inserting the viral RNA into host cells and replicating. (Adapted from ref. 23)

The viral replication starts with the translation of the replicase-polymerase gene and assembly of the replication-transcription complex (RTC). Afterwards, intermediates of negative-sense RNA are produced to function as templates for the synthesis of the genomic regions that will encode structural proteins. Following the translation of these proteins, M, S, and E proteins are released in the endoplasmic reticulum (ER) membrane and then, transported to the endoplasmic reticulum-Golgi intermediate compartment (ERGIC). Meanwhile, the N protein will complex with newly synthesized genomic RNA in the cytoplasm, generating nucleoproteins (NPs). In ERGIC, the new virions are assembled to finally exit the host cell (Figure **1.6**), leading to transmission and viral spreading to other organs. [25,26]

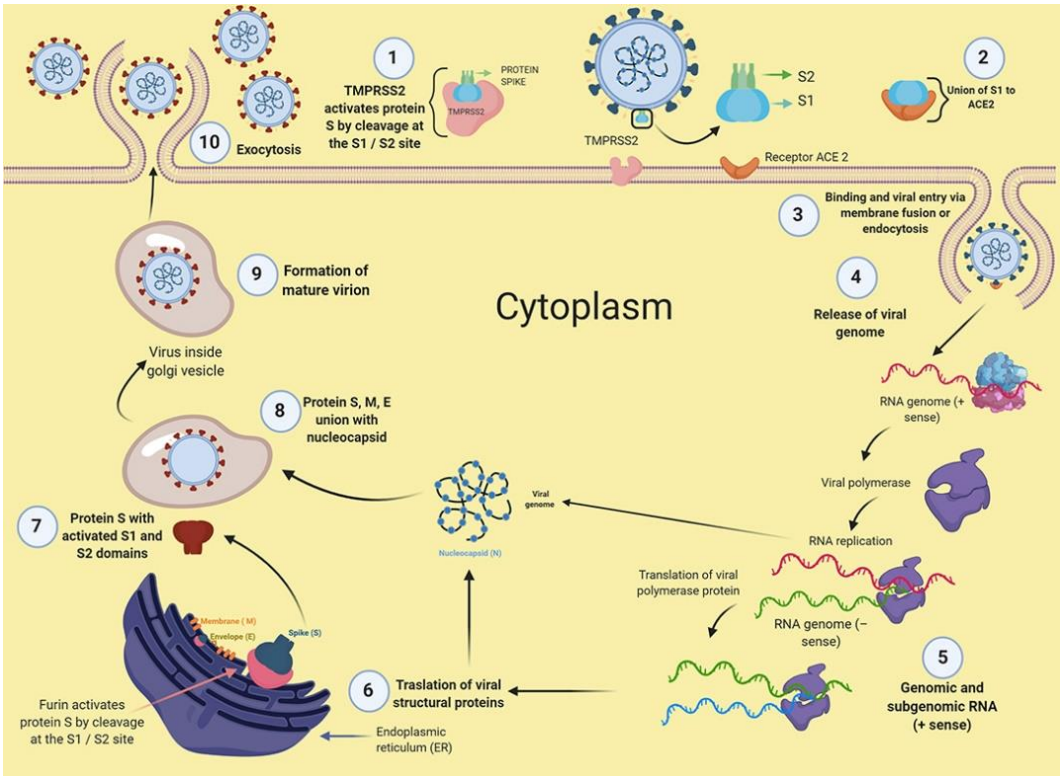


Figure 1.6 – Representation of the SARS-CoV-2 infection process, including the entry into host cells (**1** and **2**), membrane fusion (**3**), viral genome insertion into host cytoplasm (**4**), replication (**5**), translation of the structural proteins (**6**), which are assembled, in the ERGIC (**8**), to form newly mature virions (**9**). Finally, there are released from the host cell (**10**). (Adapted from ref. 26)

1.3.1 Spike Glycoprotein

The SARS-CoV-2 S glycoprotein plays an important role in viral infection, mediating the viral attachment to the ACE2 receptor, receptor engagement, catalytic cleavage of glycoproteins, and host membrane fusion. Structurally, the S glycoprotein, displayed on the viral surface, is a heavily glycosylated trimer, each monomer being constituted by 1273 amino acid residues. Since the initial outbreak of the COVID-19, advances in the structural biology field allowed the determination of a large number of structures of the SARS-CoV-2 S protein fragments, including pre-fusion and post-fusion states.[27]

In the pre-fusion conformation, the S1 subunit located at the S protein surface (14-685 residues) contains two main regions: the N-terminal domain (NTD) and the RBD. The S2 subunit (686-1273 residues) is organized into an N-terminal hydrophobic FP, two heptad repeats (HR1 and HR2), a transmembrane domain (TM), and a cytoplasmatic tail (CT). (Figure **1.7A**)[27,28]

Moreover, the three RBDs of the S protein trimer were found to adopt two different conformations. When all RBDs are in a "down" position, the S protein trimer conformation is designated "closed" conformation, wherein the RBD binding surface is covered by the adjacent protomer, thus inaccessible by the receptor (Figure **1.7B**). Interprotomer interactions take place in the "closed" conformation of the S trimer through the RBD, which is packed against the other two and one NTD from the neighboring protomer, and through the S2, involving helical contacts formed by the upstream and central helices from each subunit.[28,29]

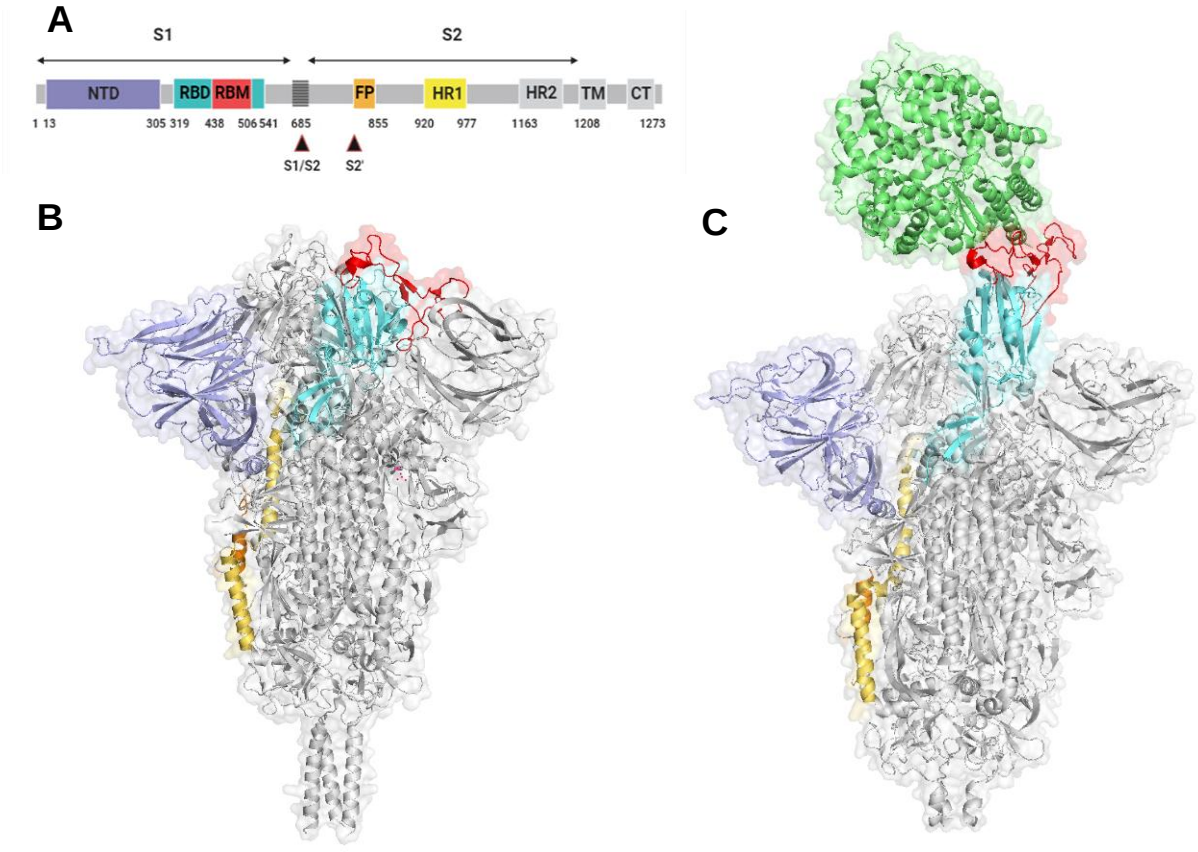


Figure 1.7 – **(A)** Scheme of the SARS-CoV-2 S protein primary structure. NTD, N-terminal domain; RBD, receptor-binding domain; RBM, receptor-binding motif; FP, fusion peptide; HR1, heptad repeat 1; HR2, heptad repeat 2; TM, transmembrane domain; CT, cytoplasmic tail. Both S1/S2 and S2' cleavage sites are represented with arrows. **(B)** Overall structures of the SARS-CoV-2 S protein trimer in the closed conformation, with all RBDs in the "down" conformation (PDB ID 6XR8); and **(C)** open, with one "up" RBD bound to the ACE2 receptor in green (PDB ID 7DF4). One of the monomers in each structure is represented with different colors to highlight the main regions: NTD in light blue; RBD in cyan; RBM in red; FP in orange; HR1 in yellow; HR2, TM, and CT in gray, as well as the remaining monomers. Images created in Biorender.com and PyMOL[30].

On the other hand, the S protein with one single RBD with the binding interface fully exposed, i.e., the “up” position, adopts an “open” conformation, representing a functional state that can bind ACE2 (Figure **1.7C**). Additionally, in this conformation, the “up” position of the RBD abolishes the interactions with the S2 of the neighboring protomer. This event allows the dissociation of the S1 and facilitates the conformational rearrangements of the S2 subunit to mediate viral fusion.[28,29]

Later, the S glycoprotein passes through a conformational transition that provides sufficient energy to overcome the natural repulsion between host and viral membranes. The receptor engagement triggers the S protein structural organization, in which the S1 subunit dissociates from S2, while S2 passes through refolding events towards a stable and elongated trimer, called the post-fusion state, allowing the process of infection to follow its course. [28,29]

Another key structural feature of the S protein is its extensive glycosylation. The surface of the S protein trimer displays 66 heterogeneous *N*-linked glycans, which may play an important role in protein folding and viral pathogenesis acting as a glycan shield (Figure **1.8A**). In each S protomer, 14 *N*-linked glycosylation sites are complex-type glycans, whereas the remaining eight sites are dominated by oligomannose-type glycans. Studies have suggested that carbohydrate moieties could be immunogenic, emphasizing the need for immunogens to display the glycans required for the recognition of neutralizing antibodies. [28,31,32]

It has also been demonstrated that the glycan shield of the S protein may alter the conformations of certain domains. For example, the *N*-linked glycosylation sites at residues N165 and N234 can alter the conformational dynamics of the RBD (“up”/”down”), affecting the interactions with ACE2. [33] Moreover the *N*-glycan located at the RBD, N343, was found to have a gating role, facilitating the exposure of the RBD.[34] The *N*-linked glycans near the RBD are represented in Figure **1.8B**.

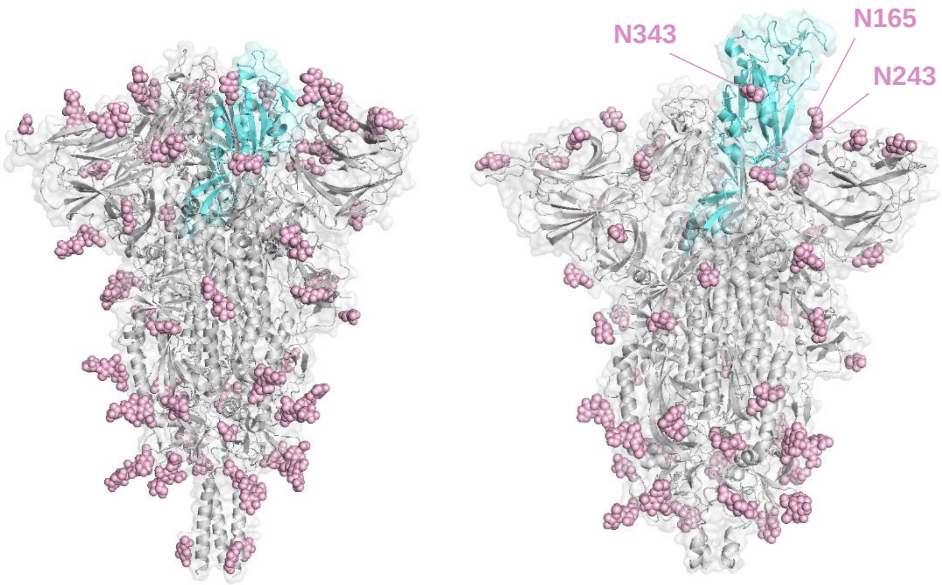


Figure 1.8 – SARS-CoV-2 S protein *N*-linked glycans, represented as light pink spheres in both “closed” (PDB ID 6XR8) **(A)** and “open” (PDB ID 7DF4) **(B)** conformations of the S trimer. In **(B)**, the *N*-glycosylation sites near the RBD (cyan) are labeled. Image created with PyMOL[30].

1.3.2 Spike RBD binds to the ACE2 host receptor

As stated before, studies have established that ACE2 is the major host-receptor for SARS-CoV-2, similar to SARS-CoV. In humans, the ACE2 (hACE2) is an 805-amino acid type I, transmembrane endothelium-bound metallo-carboxypeptidase, homologous to the ACE protein, well known in the renin-angiotensin system. This receptor is mainly expressed in endothelial cells of several organs, including the cardiovascular system, renal tubular epithelium, and alveolar epithelial type II cells in the lungs and brain. [35,36]

The SARS-CoV and SARS-CoV-2 RBDs not only share ~74% sequence identity but also exhibit similar interaction profiles with hACE2, although SARS-CoV-2 RBD presents higher binding affinity compared to SARS-CoV, due to substitutions of key interacting residues. This may contribute to more efficient transmission of SARS-CoV-2. [27,37]

The structure of hACE2 comprises two major domains: one on the N-terminus - protease catalytic domain (PD); and the other at the C-terminus - the collectrin-like domain (CLD) which contains a neck domain, a linker, a transmembrane domain (TM), and a cytoplasmic tail. hACE2 dimerization is mainly mediated by the neck domain (Figure **1.9A**). The SARS-CoV-2 S protein RBD initiates direct binding with each PD of the hACE2 host receptor for virus entry. The hACE2 dimer can accommodate two S proteins trimers, each one through a monomer of hACE2. (Figure **1.9B**). [35,36,38,39]

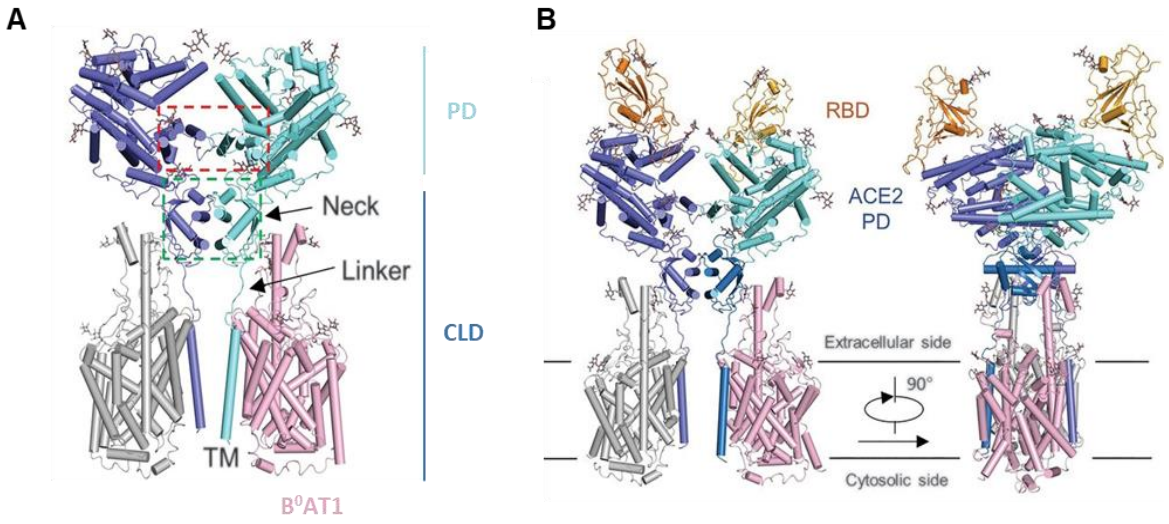


Figure 1.9 – (A) Overall structure of the hACE2-B^{0}AT1 (PDB ID 6M18). hACE2 is divided into two domains: N-terminal catalytic protease domain (PD), responsible for the binding to the S protein trimer; and a C-terminal collectrin-like domain, CLD, which comprises a neck domain, a linker, a transmembrane sequence (TM), and a cytoplasmic tail (CT), which is not represented here. The hACE2 (in blue) dimerizes mainly through the neck domain. hACE2 protomers are shown as cylindrical blue helices and loops, whereas the B^{0}AT1 protomers are colored gray and pink. **(B)** Structure of RBD-hACE2- B^{0}AT1 (PDB ID 6M17) demonstrates that the dimeric hACE2 accommodates two S protein trimers. RBD is represented in orange and yellow. The glycans are shown as sticks. (Adapted from ref. 38)

Moreover, the hACE2 contains seven *N*-linked glycans at the N-terminus of the PD (at residues N53, N90, N103, N322, N432, N546, and N690), and *O*-linked glycosylation sites at S155 and T730. The *N*-glycosylation at residues N90, N322, and N546 of hACE2 is particularly important in the binding of hACE2 to the RBD of the S protein. The substitution of the residue N90 improves the binding affinity to the S protein, indicating that N90 glycosylation may protect host cells against viral infection. [32,40]

RBD-hACE2 Binding Interface

Understanding the receptor recognition mechanism of SARS-CoV-2 is important, because not only does it affect the infectivity and pathogenesis of the virus[11], but also blocking the interaction of the RBD with hACE2 receptor has significant therapeutic potential.[41–43]

Several structures have revealed details about the interactions between the SARS-CoV-2 RBD and its receptor hACE2. The RBD structure is composed of two subdomains: a core and an external region (figure **1.10**). The RBD core is characterized by a five-stranded antiparallel β-sheet structure (β1-β3-β7-β4-β2), stabilized by three pairs of disulfide bonds (C336-C371, C379-C432, and C391-C525), as well as small loops and helices. The receptor-binding motif (RBM, residues 438-506) is composed by an extended loop on the edge of the RBD core that produces a slightly concave surface for receptor binding, with a ridge region (residues 471-491) generated by a loop stabilized by another disulfide bond (C480-C488). The RBM region also has two short β-sheets (β5 and β6).[1]

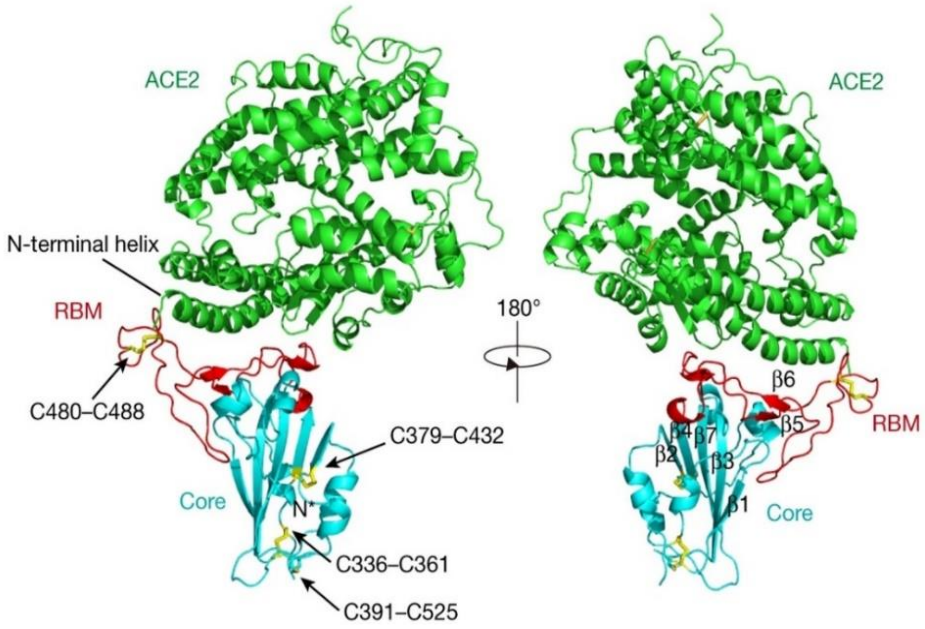


Figure 1.10 – Overall structure of SARS-CoV-2 S protein RBD bound to hACE2 receptor, where the extended insertion (RBM) binds to the hACE2 N-terminal helix. hACE2, RBM and RBD core are colored in green, red, and cyan, respectively. Disulfide bonds in the SARS-CoV-2 RBD are shown as sticks and indicated by arrows. (Adapted from ref. 1).

The interface between SARS-CoV-2 RBD and hACE2 receptor is mainly composed of both hydrophilic and hydrophobic interactions, mediated by a total of 17 residues from the RBD and 20 residues of hACE2. The principal residues mainly involved in the interaction with the S protein are located at the N-terminal helix of the hACE2 PD (helix α1). At the N-terminus of the α1, the residues Y41, Q42, K353, and R357 form a hydrogen bond network with the RBD residues Q498, T500, and N501. In the middle, RBD residues K417 and Y453 interact with D30 and H34 of hACE2, respectively. At the α1 C terminus, Q474 of the RBD interacts with Q24 of hACE2 through a hydrogen bond, while F486 of the RBD interacts with Q24, L79, Y83, and M82 of hACE2 through van der Waals forces. Additionally, the helix α2 and a loop that connects antiparallel β-sheets (β3 and β4) of the PD hACE2 also interacts with the RBD (figure **1.11**).[38]

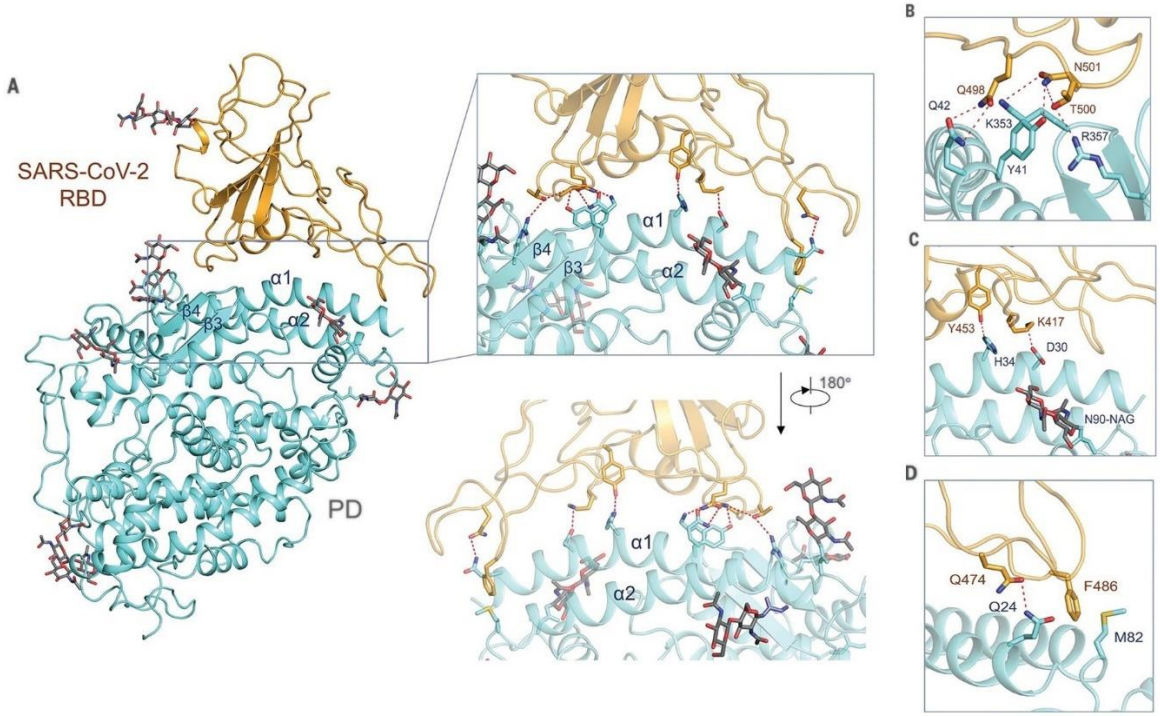


Figure 1.11 – Representation of the interactions between SARS-CoV-2 S RBD and ACE2. **(A)** The ACE2 PD helix α1 is mainly involved in the interaction with the S protein. α2 helix and the loop that connects β3 and β4 also contribute to the interaction. **(B-D)** Detailed representation of the complex interface. Dashed lines indicate polar interactions. (Adapted from ref. 38).

RBD is a privileged target for neutralizing antibodies

Most of the most powerful neutralizing antibodies (nAbs) elicited either by vaccination or natural infection that have been identified so far bind the viral S protein's RBD, recognizing one or both “up” and “down” conformations, although NTD nAbs have been described.[1,44–48]

Depending on the binding region, the epitopes in the RBD can be classified into different classes. Although different classification systems have been proposed, the most commonly referenced is from Barnes *et al.*, in which the antibodies are grouped into four classes based on their binding mode to the RBD. [45,47,48]

Antibodies from class I and class II compete with the hACE2 receptor for the binding site. While class I antibodies recognize only the “up” RBD conformation, class II antibodies are capable of binding to the RBD in both “up” and “down” conformations. Class III antibodies binding epitope is located in the opposite region of class I and II binding epitope, near the *N*-glycosylation site at residue 343. Finally, class IV antibodies target a highly conserved epitope on the inner region of the “up” RBDs, which usually presents low neutralizing efficacy (Figure **1.12A**). Examples of nAbs of each class are represented in figure **1.12B**.[45–49]

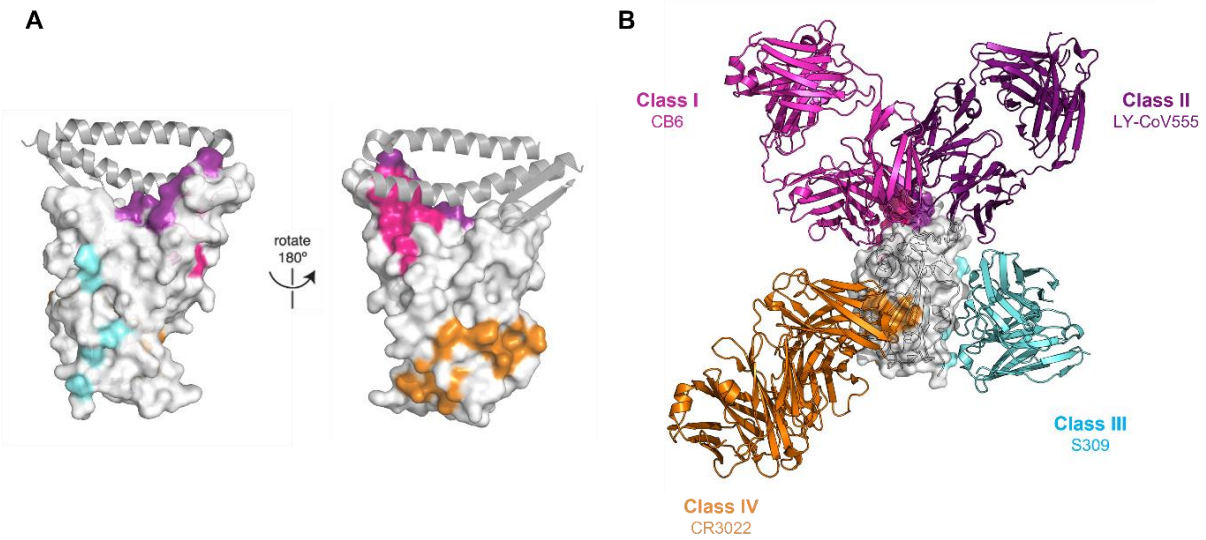


Figure 1.12 – RBD-directed neutralizing antibodies (nAbs) are classified into four classes depending on the binding region to the S protein RBD. **(A)** Colored surfaces on the RBD depict the epitopes. The class I epitope is colored in pink; class II in purple; class III in cyan and class IV in orange. ACE2-binding region is shown as a gray cartoon. **(B)** For each class, one representative nAb bound to the RBD monomer is shown: class I, CB6 (PDB ID 7C01); class II, LY-CoV555 (PDB ID 7KMG); class III, S309 (PDB ID 7TLY); class IV, CR3022 (PDB ID 7JN5). Image created in PyMOL[30] using the ACE2-bound RBD structure from PDB ID 6M0J.

Understanding the characteristics of these widely nAbs might contribute in the development of novel antibody therapies as well as the rational design of new vaccines.[47]

1.3.3 SARS-CoV-2 Variants: Evolution of the S protein

Since the beginning of the COVID-19 pandemic, SARS-CoV-2 has been evolving through the acquisition of genomic mutations, resulting in the emergence of several distinct variants. According to health agencies, namely WHO, SARS-CoV-2 variants are characterized based on evidence of transmissibility, virulence, or escape from immune response or therapeutics. In order to prioritize global monitoring and research, and to inform the ongoing response to the pandemic, these variants are classified into Three categories: variants of concern (VOC), variants of interest (VOI), or variants under monitoring (VUM). [50,51]

A VOC is characterized by increased transmissibility, with a more severe disease outcome; a considerable decrease in neutralization by antibodies produced during prior infection or vaccination, as well as reduced therapeutic effectiveness. On the other hand, a VOI is defined as a genetic variant associated with altered receptor binding, reduced antibody neutralization, reduced treatment efficacy, potential diagnostic impact, or predicted increase in transmissibility or disease severity. A VOC or a VOI may be reclassified once it has been definitively shown that it no longer significantly increases the danger to global public health compared to other circulating variants. Additionally, a variant with genetic alterations, which may affect certain characteristics, but its phenotypic or epidemiological impact is currently unclear can be classified as a VUM.[50]

Previous variants were classified as VOI, such as Epsilon, Zeta, Theta, Iota, Kappa, Lambda, and Mu, and others as VOC, including Alpha, Beta, Gamma, and Delta that currently are not circulating. In November 2021, the SARS-CoV-2 Omicron variant have largely replaced other co-circulating VOCs. Currently, this is the dominant variant circulating globally, especially its subvariant, BA.5. [50,52,53]

These strains contained several mutations at the encoding genes for different proteins, most of them leading to amino acid substitutions in the S protein. This has resulted in changes in viral behavior and pathogenesis as the viral entry in host cells depends on the engagement of this protein to the hACE2 receptor. The transmissibility of each SARS-CoV-2 variant has been associated with the structure stability of the S protein and its binding affinity to the host receptor. Another feature that can be changed in these variants is the cell fusion potential or fusogenicity due to alterations located at the S1/S2 or furin cleavage sites. Furthermore, alterations in the S protein, especially in the RBD, may lead to evasion from antibodies derived from natural infection or vaccination. [52,54–56]

The comprehension of the extend of the SARS-CoV-2 mutations and their potential effect on immune evasion is crucial to evaluate the spread and the sensitivity of the SARS-CoV-2 variants to therapeutic antibodies. [56,57]

The summary of the previous and current circulating VOCs is represented below (Table **1.1**), indicating when these variants have emerged, the substitutions located at the S protein NTD, RBD, furin site, and remaining sites, and principal features. [23,50,58,59]

Table 1.1 – Previous (1) and current (2) SARS-CoV-2 variants of concern (VOC), according to WHO, and their date of appearance, the substitutions, deletions (Δ), and insertions (ins) on the S protein NTD, RBD, furin site and other sites, and principal features. Information adapted from [23,50,58,59]

WHO label (PANGO lineage)	Date of appearance	S protein substitutions				Features
		NTD	RBD	Furin site	Other	
(B.1)[1]	February 2020	None	None	None	D614G	Possible ancestor of emerging variants
Alpha (B.1.1.7)[1]	September 2020	Δ69–70 Δ144	N501Y	P681H	A570D D614G T716I S982A D1118H	Increased transmissibility; modest Ab escape
Beta (B.1.351)[1]	August 2020	D80A D215G Δ241–243	K417N E484K N501Y	None	D614G A701V	Increased transmissibility; modest Ab escape
Gamma (P.1)[1]	July 2020	L18F T20N P26S D138Y R190S	K417T E484K N501Y	None	D614G H655Y T1027I	Increased transmissibility; modest Ab escape
Delta (B.1.617.2)[1]	October 2020	T19R Δ156–157 R158G	L452R T478K	P681R	D614G D950N	Increase transmissibility; modest Ab escape; preliminary studies suggest increased S1–S2 cleavage

Omicron (B.1.1.529)[2]	November 2021	T19I [b,d-f] Δ24-26[b,f] A27S [b,d-f] A67V[a] Δ69–70[a,c,f] T95I[a,c] G142D Δ143-145 [a,c] N211I[a,c] L212V [a,c] V213R [a,c] V213G [b,d-f] ins214EP[a] ΔR216[c] R216E[a]	G339D R346T[e,f] S371L[a] S371F[b-f] S373P S375F T376A[b,d-f] D405N[b-e] R408S[b,d-f] K417N N440K K444T[e] G446S[a] L452R[d-f] N460K [e] S477N T478K E484A F486V[d-f] Q493R[a,b] G496S[a] Q498R N501Y Y505H	H655Y I666V[e] N679K P681H	T547K[a] D614G E619Q[e] N764K D796Y N856K[a] Q954H N969K L981F[a]	Predominant variant (BA.5 lineage with ~90% worldwide prevalence, as of October 2022); most transmissible[1]

1.3.4 Impact of VOC RBDs in hACE2 binding and immune evasion

As mentioned before, the binding of SARS-CoV-2 RBD to the hACE2 receptor is a crucial step for viral entry. The substitutions on this region may affect the interaction with hACE2, enabling stronger affinity and binding capacity to this receptor, leading to higher transmissibility. Furthermore, the RBD is a well-known target of neutralizing antibodies (nAbs), thus some substitutions are also associated with immune evasion. Therefore, this could have a substantial impact on the efficacy of existing vaccines and therapeutics.

Several investigations have been carried out either experimentally or utilizing computational biology methods to comprehend the extent of the SARS-CoV-2 VOC RBDs in hACE2 binding and antibody escape. Among all VOCs, some substitutions became prevalent in this region, which may imply their importance in viral fitness. The frequency of RBD substitutions in previously and currently circulating VOCs is represented in figure **1.13**.

In this thesis, the significant substitutions in this region are described in more detail.

Table 1.1 describes the substitutions of the SARS-CoV-2 Omicron lineages a) BA.1, b) BA.2, c) BA.3, d) BA.4/B.5, e) BQ.1 (and BQ.1.1), and f) BF.7. The S proteins of lineages BA.4 and BA.5 are very similar.

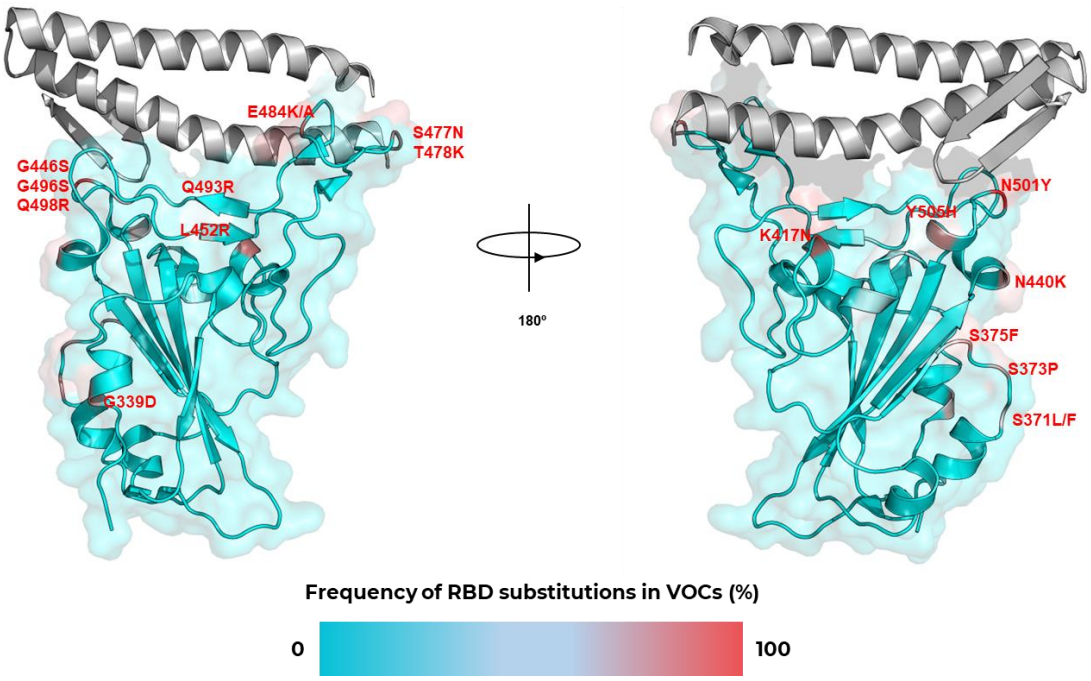


Figure 1.13 – Representation of the frequency of RBD substitutions in previous circulating SARS-CoV-2 VOCs (Alpha, Beta, Gamma, and Delta) and current circulating VOC (Omicron, including its subvariants). The RBD is colored according to the frequency (%), from cyan (0%) to red (100%). hACE2 binding region is represented in gray cartoon. The residues are labeled in red. Image created in PyMOL[30] using the hACE2-bound RBD structure from PDB ID 6M0J.

Previous circulating VOCs

In late 2020, new VOCs had been identified in the United Kingdom (Alpha), South Africa (Beta), Brazil (Gamma), and India (Delta).

The Alpha variant was responsible for the increased number of daily infections and mortality rate in England during September 2020, due to its high transmissibility compared to the non-VOC. This variant was also found to be resistant to neutralization through nAbs. [60]

Compared to the wt, the Alpha S protein comprises seven amino acid substitutions, including a change from asparagine to tyrosine at residue 501 (N501Y) located at the RBM. Studies have shown that this RBD substitution results in a stronger binding to the hACE2 receptor compared to wt RBD. While wt N501 formed a single hydrogen bond with hACE2 Y41, the aromatic side chain of Y501 resulted in favorable van der Waals interaction, and the hydroxyl group may interact with K353 of hACE2. Using Monte Carlo sampling and molecular dynamics (MD) simulations, this substitution was found to be responsible for the improved electrostatic interactions at the interface as well as the formation of a hydrogen bond between RBD residue T500 and residue D355 of hACE2 near the substitution site. [59,61–63]

Moreover, N501Y was identified in Beta and Gamma variants, both associated with re-infections and escape from vaccine-induced sera and natural sera. Two additional substitutions are located at the Beta and Gamma RBDs: E484K, K417N in Beta or K417T in Gamma. The E484K substitution has no significant effect on the affinity for the receptor. The loss of a salt bridge with hACE2 lysine 31 is replaced by a new interaction with neighboring E35 in the same hACE2 alpha-helix. However, it has a significant effect on antibody recognition, leading to immune evasion, possible reinfection, and reduced efficacy of vaccines and therapies.[59,62,64]

In these two strains, the E484K/N501Y variant is of particular concern as it combines high affinity for the hACE2 receptor and the immune evasion properties, therefore presenting fitness advantage. Moreover, the two substitutions, K417N and K417T are associated with conformational changes in S protein and confer additional immune evasion, although these destabilize the RBD-hACE2 complex due to the loss of a salt bridge across the interface that was observed in wt strain. [59,64]

The Delta variant first emerged in India in October 2020, and rapidly became dominant there, causing the second wave of infections in this country. Later, it spread around the globe, replacing the Alpha variant as the dominant variant. This VOC contains eight substitutions affecting the S protein, two of which are present in the RBD, L452R, and T478K. The first one was detected in earlier VOIs, such as Epsilon and Kappa, and is associated with increased infectivity, higher transmission (~20% more transmissible), and reduced neutralization by specific therapeutic antibodies. [59,65]

The presence of the T478K and L452R substitutions on the RBD could result in increased hACE2 binding as seen with other strains that contain such substitutions, resulting in better transmissibility. The improved interactions of K478 observed with hACE2 (residues P84 and M82) might explain why the Delta variant had higher infectivity.[66] Other studies illustrated that strains that contain the L452R substitution can escape host immune response and found that this substitution decreases response to vaccines (Pfizer-BioNTech and Moderna) compared to the original strain.[64,65,67] Additionally, evidences shows that substitutions located at L452 facilitate escape from class II and class III antibodies. [68]

Omicron variant: the current circulating VOC and its subvariants

Until November 2021, the Delta variant was the dominant strain worldwide. Nevertheless, on November 26, 2021, WHO identified a new VOC, B.1.1.529, labeling it Omicron, since it carries several mutations that may influence its behavior and might impact how rapidly it spreads or the severity of the disease it causes. The Omicron S protein contains about thirty amino acid substitutions, fifteen of which occur in the RBD (Table **1.1**), thus, raising concerns about transmissibility and antibody evasion. Among these substitutions, some are shared with previous VOCs, such as Alpha (N501Y), Beta, Gamma (E484A and K417N), and Delta (T478K and L452R, in BA.5 lineage). [69–71]

Since its emergence, the Omicron variant has been divided into seven major sublineages, BA.1, BA.2, BA.3, BA.4, BA.5, and recently BQ.1 and BF.7, as of October 25, 2022. These Omicron sublineages continue to undermine immunity induced by previous infections and/or vaccines.[72,73] The first BA.1 Omicron led to an enormous wave of COVID-19 infections and demonstrated a decreased response from the 2 mRNA vaccine doses, although this was recovered with the administration of vaccine booster doses.[71,74] Moreover, BA.1 sublineage was found to alter the entry pathway and showed reduced replication in lower airway epithelial cells, i.e., in lung tissues, however, increased its replication capability in the nasopharyngeal tract. These features may have decreased pathogenicity *in vivo*. The following Omicron subvariants have been maintaining these features.[75–78]

The substitutions in the BA.1 S protein are listed in Table **1.1**. Of the RBD substitutions in BA.1 S protein, eight are located at the hACE2-recognizing interface, namely K417N, G446S, E484A, Q493R, G496S, Q498R, N501Y, and Y505H. Studies found that substitutions at residues 493,496,498 and 501 may restore hACE2 binding affinity loss due to the K417N and Y505H substitutions Another substitution, S477N, is found to expand the binding surface area and was earlier described to afford resistance to multiple nAbs.[79] The substitution T478K, observed earlier in Delta, does not contribute to the hACE2 interaction.[70] Additionally, a computational study observed that S371L-S737P-S375F substitutions may lead to local conformational changes, contributing to the escape of class IV nAbs. [71,80]

Later, we witnessed the emergence of BA.2 that quickly replaced BA.1, due to its higher immune evasion and transmissibility, which led to reinfection cases from patients previously infected with BA.1. These two lineages share twenty-one S substitutions identified in the probable common ancestor, although BA.2 S protein presents seven additional substitutions compared to BA.1, where four occur in the RBD (S371F, T376A, D405N, and R408S). Also, in both residues 446 and 496 of the RBD there is a “reversion” from serine to glycine (S446G and S496G) since before Omicron all previous VOCs contain wt G446 and G496.[78,81,82]

One of the subvariants of BA.2, BA.2.75.2, showed a strongly reduced neutralization capacity and increased fusogenicity due to a substitution in the residue 486 from phenylalanine to serine (F486S). [83]

From that point forward, several subvariants quickly appeared, of which BA.4 and BA.5 began to expand worldwide and, currently BA.5 is the dominant one with ~90% prevalence. This is due to its substantial immune escape compared to the previous subvariants. It is noteworthy that these two subvariants contain identical S proteins, therefore are often referred to as BA.4/BA.5. These Omicron subvariants present two new substitutions in the RBD, L452R, already described, and F486V. F486 is a key residue for the binding to the hACE2 receptor and certain nAbs, forming hydrophobic contacts with hACE2 residues, M86 and Y83. In BA.4/BA.5, F486V facilitates escape from certain class I and II antibodies but compromises the binding affinity to the hACE2 receptor. Furthermore, in contrast with the earlier BA.1 and BA.2 subvariants, there is a “reversion” from arginine to glutamine at residue 493 (R493Q) which is hypothesized to restore the receptor affinity, increasing viral fitness. [68,78,83]

Later on, new subvariants emerged from BA.4/BA.5, such as BF.7 and BQ.1, which have revealed a strong neutralization resistance, as well as altered the fusogenicity of the S protein. Regarding alterations in the S protein RBD, these subvariants share one newly substitution, R346T, while BQ.1 contains two unique substitutions, N460K and K444T. R346 and K444 are located outside of the binding interface with hACE2 and within the class III nAb epitope, thus the substitutions at these sites may interfere with antibody recognition. A recent study found that the novel substitutions present in these subvariants, specifically N460K, R346T, and K444T, increased resistance to neutralization of antibodies elicited by immunization with 3 mRNA dosages and/or infection with BA.1 and BA.4/BA.5. Additionally, N460K exhibits enhanced fusogenicity and S proteolytic processing in BQ.1 subvariant. [78,83]

These findings together shed light on the impact of Omicron subvariants' substitutions in the S protein on the binding to hACE2 receptor and nAbs, which underpin the increased transmissibility and immune escape. However, the impact of RBD substitutions in emerging subvariants, and in particular effect on its conformational dynamics, remains to be further explored.

1.4 RBD conformational dynamics

According to the experimental structural data that is currently available, the RBD core is quite stable, whereas the RBM appears to be dynamic and not as structurally defined until attached to hACE2 or antibody fragments. [84]

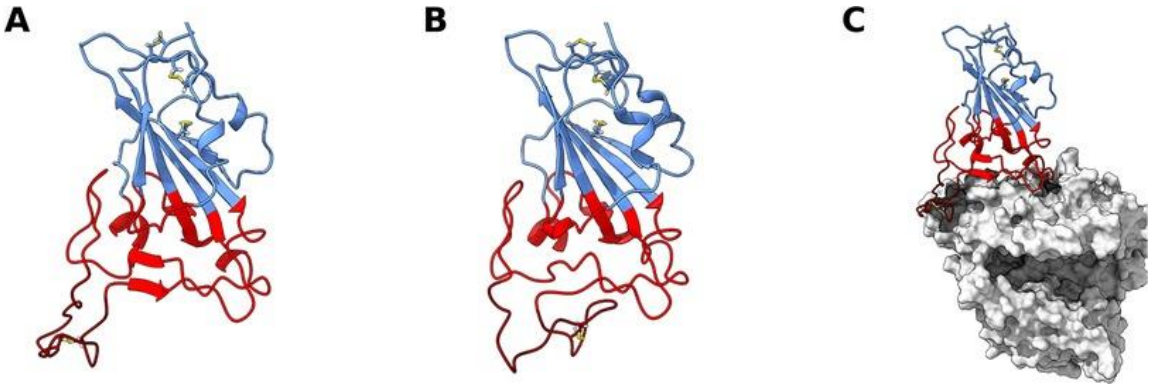
The majority of MD simulations studies either focused on the RBD complexed with these proteins or relied on short simulations of the free RBD, but did not examine the RBM dynamics. Therefore, the conformational dynamics of this region when unbound was not clear, although this might have an impact on receptor recognition and engagement, as well as give important information for the production of novel pharmaceuticals.

In order to address this, Valério M., Borges-Araújo L. *et. al* performed microsecond (μs) atomistic MD simulations with the unbound wt RBD, with particular detail on the conformational changes that occur in the RBM region. From the full trajectory analysis, it was possible to observe a high flexibility in this region, in which the RBD assumes two different conformations in equilibrium: one similar to the one present when the RBD is bound to the hACE2 receptor - "open" conformation – and another where the RBM "ridge" hides the hACE2 binding interface – "closed" conformation (Figure **1.14**). It is noteworthy that, in this study, the "open"/"closed" dynamic is not referred to the "up" and "down" S trimer conformational states which control the RBD exposure. Rather it is used as a measure of the RBM exposure, particularly when the S trimer adopts the "open" conformational state and the RBD is fully exposed to the solvent. [85]

A closer analysis of the intramolecular interactions formed in both "open" and "closed" conformations was performed through residue interaction networks (RIN). This highlighted the key residues that control this dynamics. In the "open" conformation, pi-stacking interactions between Y489, F456 and Y473 and a hydrogen bond between Y489 and Y473 stabilize the

“ridge”. Another two hydrogen bonds are formed by Y453 and Q493, stabilizing the formation of a small beta-sheet.[85]

The “closed” conformation is stabilized by the formation of new interactions, replacing the others present in the “open” state. There are two pi-stacking interactions established between residues F456 and Y421, and between Y489 and F486. Additionally, Y473 forms a hydrogen bond with the backbone of Y451. A triple interaction between E484, K417 and R403 was observed to be important for the closing of the loop, as well as the formation of three hydrogen bonds (C480-S494-G482-Q493). [85]



.

Figure 1.14 – SARS-CoV-2 wt RBD in the “open” **(A)** and “closed” **(B)** conformations, and the structure of wt RBD complexed with hACE2 **(C)**. The RBD core is represented in blue, the RBM region in red, and the “ridge” in dark red. hACE2 is represented in gray. Disulfide bonds are represented in yellow sticks. (Adapted from ref. 85)

1.4.1 SARS-CoV-2 variants impact RBD dynamics

Not only it is important to understand the direct impact of the variants on the hACE2 binding through the study of the alterations on the VOC RBDs-hACE2 interface but also their indirect effects characterized by the conformational dynamics of the RBD. Therefore, Valério M., Borges-Araújo L. *et al* also performed MD simulations of free VOC RBDs, except for the Gamma variant since it is similar to the Beta variant (Table **1.1**). [85]

Alpha and Beta variants share the N501Y substitution in the RBD, as mentioned earlier. The Beta variant presents the two additional K417N and E484K substitutions. In this study, it was possible to observe that the RBDs from both variants also demonstrated an “open”/“closed” dynamic, although this equilibrium was observed to shift towards the “open” state by ~20%. Thus, the Alpha and Beta RBD substitutions might increase the accessibility of the binding region to hACE2, leading to a better RBD-hACE2 binding. [85]

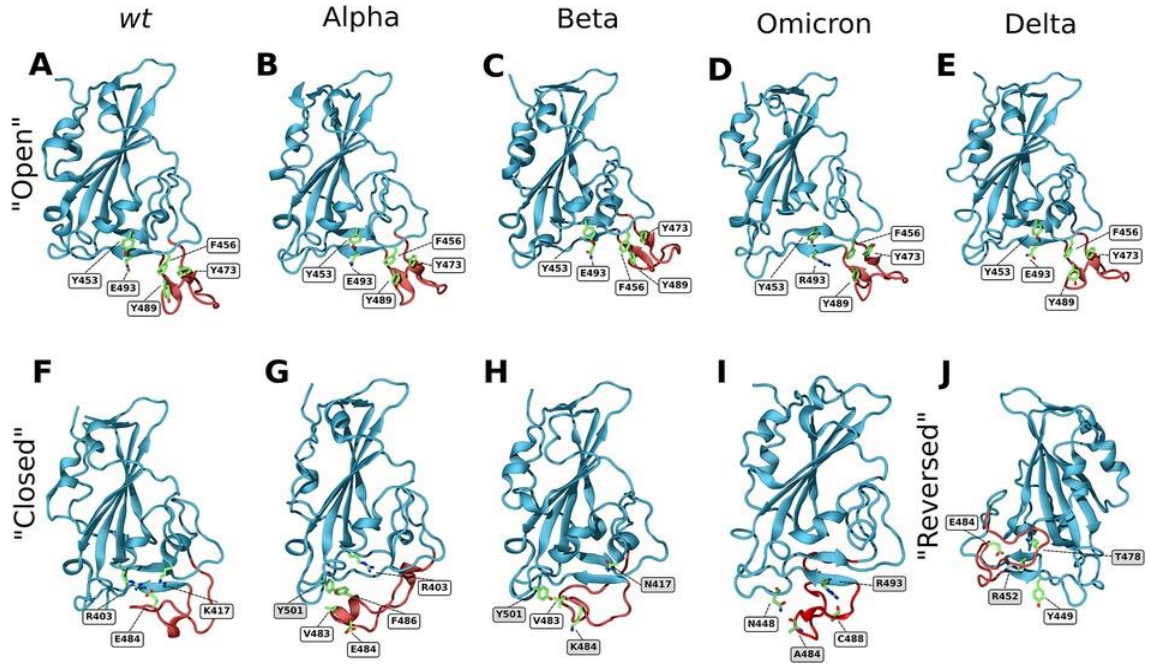


Figure 1.15 – SARS-CoV-2 RBD intramolecular interactions which stabilize the "open", "closed" and "reversed" RBD conformations for the wt (**A**, **F**), Alpha (**B**, **G**), Beta (**C**, **H**), Omicron (**D**, **I**) and Delta (**E**, **J**) variants. The "ridge" region (residues 471-491) is colored in red and the key residues are in green and labeled. The shaded labels indicate RBD substitutions relative to wt. (Adapted from ref. 85)

In the Alpha variant, through the analysis of the intramolecular interactions, it was observed that the N501Y substitution did not alter the main contacts in the wt RBD "open" conformation. However, the interactions that stabilized the wt "closed" state differ for this variant. The E484-R403 salt bridge was replaced with hydrophobic interactions between the substituted Y501, V483, and F486. The lack of the salt bridge turns the "closed" conformation less stable, thus the decrease in percentage of this state (Figure **1.15B**, **G**). [85]

In the Beta RBD, the E484K substitution had a significant impact on the destabilization of the "closed" conformation since the E484K substitution prevented the salt bridge described above. Nevertheless, the "closed" state was reached because of the hydrophobic interaction between the V483 and the substituted Y501. (Figure **1.15C, H**). [85]

Unlike the previous variants, Delta RBD contains two newly substitutions, L452R and T478K, and demonstrated a distinct conformational dynamic. MD simulations showed a prevalence of the "open" conformation observed in wt, Alpha, and Beta variants, although there was no evidence of a "closed" conformation. Instead, Delta RBD presented an alternative "open" state, the so-called "reversed" conformation, where the "ridge" region is folded backwards. This configuration may lead to an increase in viral fitness due to an open hACE2 binding surface and may affect antibody recognition. One of the main factors for this appears to be a hydrogen bond between the substituted R452 and Y449, destabilizing one of the beta sheets of the RBM and allowing the "ridge" to move closer to the core. Other transient interactions are observed to stabilize this conformation, such as the ones between G476, and S477 and the substituted K478 with R346, F347, and N354 (Figure **1.15E, J**). [85]

This work also provided information about Omicron BA.1 subvariant. MD simulations performed with BA.1 RBD showed that the “open” conformation accounted for 96% of simulation time, which likely increases accessibility for the RBM to hACE2 and facilitates binding. The “open” conformation for the BA.1 RBD is stabilized by similar interactions observed for the other variants (Figure **1.15D**). However, the “closed” conformation is quite different, in which two transient hydrogen bonds stabilize this state, one between the sidechain amide of N448 and the backbone of the substituted A484, and the other between the sidechain of Q493R and the backbone of C488. As for the Beta variant, the E484A substitution prevents the salt bridge interaction with R403. Moreover, hydrophobic interactions between Y501, V483, and F486 observed in the closed conformations of other variants are absent in the Omicron BA.1 (Figure **1.15I**). Thus, substitutions in Omicron BA.1 RBM inhibit the closed state, which may facilitate the RBD-hACE2 binding. [85]

Since the substitutions most relevant for the “open” / ”closed” dynamics of the loop are shared between all Omicron sublineages - E484A, Q493, N501Y, and K417N -, the RBD dynamics of the other Omicron subvariants were hypothesized to be similar to that of BA.1.[85] Nevertheless, it is important to comprehend how the Omicron variant evolved over time regarding the hACE2 binding and antibody escape. Thus, investigating the predominant Omicron sublineages, BA.2 and BA.5, is proposed in this work.

1.5 Objectives

Since the SARS-CoV-2 mechanism of infection is initiated with the binding of the virus to the hACE2 host receptor through the S protein RBD, it is crucial to describe the binding mode between the viral Spike and the host receptor.

With the evolution of SARS-CoV-2, new VOCs have emerged, containing substitutions in the RBD that are associated with increased transmissibility and/or the ability to escape the immune response elicited by vaccination or infection. Moreover, the Omicron variant raised considerable concern due to the presence of over thirty amino acid substitutions, fifteen of which occur in the RBD. Nowadays, this VOC is predominant worldwide and has evolved into seven sublineages.

The main purpose of this thesis is to characterize the direct and the indirect impact of the RBDs substitutions on its dynamics and interaction with human hACE2, with a major focus on the Omicron subvariants, BA.1, BA.2, and BA.5. To achieve this, this work was divided into two parts.

Firstly, long atomistic MD simulations of Omicron RBDs-hACE2 complexes are going to be performed to investigate the direct role of significant substitutions in the interface between the RBD and the receptor, using wt as control.

On the other hand, Omicron BA.1 subvariant was already considered to impact RBD conformational dynamics towards an efficient binding to hACE2.[85] In order to study the structural dynamics of the remaining subvariants, BA.2 and BA.5, and the consequent impact on receptor recognition, atomistic MD simulations with the unbound Omicron RBDs in solution are also proposed.

2. Materials and Methods

2.1 Molecular Modeling

Molecular modeling is defined as the study of the molecular structure and function using a simplified description of a system or process, designed to facilitate calculations and predictions - a “model”. Thus, it is possible to translate essential physical principles to predict the behavior of molecules and molecular systems, from simple solids and liquids to complex systems, and even biological matter. [86,87]

Given the advancement of technology, hand-written calculations performed to obtain a basic model were substituted by a powerful set of computer techniques that have revolutionized molecular modeling. The variety of models and systems to which molecular modeling can be applied have increased in such a way that nowadays most calculations cannot be performed without computers. [87]

Considering biological systems, predicting molecular properties can be challenging because of their complexity. Although experimental approaches can be applied to study some molecular features, detailed characterization of the conformations over time cannot be obtained. Therefore, computational studies based on molecular models have started to have an important role in biological chemistry, biophysics, and biology. Computer simulations are performed in order to complement the information that is obtained experimentally, including the conformational distribution or atomistic-detailed protein-protein interactions.[88]

Depending on the type of question to be addressed, different molecular models may vary on the level of detail that is used to represent the system (Table **2.1**).[88]

Table 2.1 – Examples of methods used in molecular modeling. Adapted from ref. 88.

Methods	Degrees of freedom	Time scale
Quantum dynamics	Atoms, nuclei, electrons	Picosecond (ps)
Quantum Mechanics (ab initio)	Atoms, nuclei, electrons	N/A
Classical Statistical Mechanics (MD, MM, force fields)	Atoms, solvent	Microsecond (μs)
Statistical Methods (database analysis)	Group of atoms, amino acid residues, bases	N/A
Continuum Methods (hydrodynamics and electrostatics)	Electrical continuum, velocity continuum, etc.	Supramolecular

2.2 Molecular Mechanics

Molecular modeling studies of biological molecules commonly require the evaluation of potential energies for alternate conformations. Quantum mechanics (QM) is the most accurate method, wherein the electronic degrees of freedom are represented explicitly in the calculation. Nevertheless, most systems of biological relevance contain a large number of particles that would lead to extremely time-consuming calculations if they were treated at the quantum mechanical level. [87,89]

Therefore, to describe molecular systems, such as solvated proteins, classical molecular mechanics (MM) energy functions are generally employed. In the MM method, the energy of the system is simplified to a function of nuclear position and velocities, due to the Born-Oppenheimer approximation, which states that the electronic and nuclear wave functions can be treated separately, since the electrons are much lighter and move faster than nuclei. Thus, one can analyze the degrees of freedom of nuclei while considering electrons in a particular state, e.g, the ground state. Furthermore, the MM model considers the atomic nuclei as classical particles, implying that their positions and velocities can be calculated by integrating Newton's laws of motion. Therefore, the computational speed required to perform energy calculations can be achieved due to this simplification. [87,90]

2.2.1 Force fields

In MM methods, potential energy functions or force fields are used to describe the intra- and intermolecular interactions between the atoms of a system. The potential energy (V) is a function of the nuclear atomic coordinates (r) of the particles, which compose the system. This function is defined as the sum of the terms that accounts for covalent and noncovalent interactions between atoms [87], represented by the expression in equation **2.1**:

$$V(r) = V_{bonds} + V_{angles} + V_{dihedrals} + V_{VDW} + V_{electrostatic} \qquad (\mathbf{2.1})$$

The first three terms correspond to bonded contributions including bond-stretching (V_{bonds}), bond angle bending (V_{angles}), bond torsion ($V_{dihedrals}$), whereas the last two terms refer to the nonbonded contributions, such as the van der Waals interactions (V_{VDW}) represented as the Lennard-Jones term, and the electrostatic Coulomb interactions ($V_{electrostatic}$) (Figure **2.1**).[87,88,91]

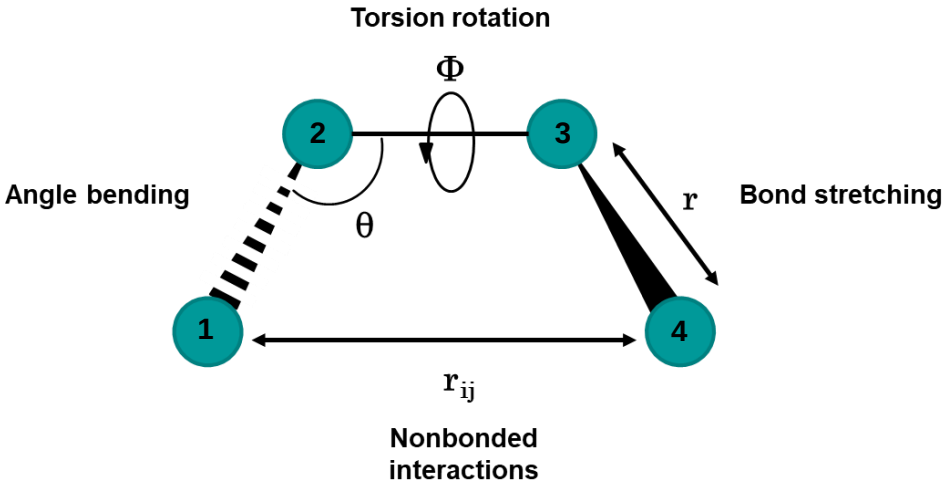


Figure 2.1 – Schematic representation of the bonded forces acting on a molecule: bond stretching, angle bending, and torsional rotations. Meanwhile, electrostatic and van der Waals terms describe non-bonded interactions. Image created using Biorender.com.

Because of the development of the MM method, the first force fields appeared in the 1960s in order to study small organic molecules. Over the years, the focus of the research extended to comprise considerably more complex systems, thus leading to the development of more widely applicable force fields. Nowadays, there are a large number of force field types, and the selection will depend on the particular problem being addressed. Popular force fields employed in simulations of biomolecules are Chemistry at HARvard Macromolecular Mechanics (CHARMM), Assisted Model Building with Energy Refinement (AMBER), and GROningen MOlecular Simulation (GROMOS).[91]

In this work, the force field applied in the simulation studies was the AMBER, more precisely Amber14sb, which is commonly used in simulations of nucleic acids and proteins.[92,93] The energy terms of the AMBER force field are given by the following equations[94,95]:

$$V_{bonds} = \sum_{bonds} 2k_b(r - r_0)^2 \qquad (2.2)$$

$$V_{angles} = \sum_{angles} 2k_\Theta(\Theta - \Theta_0)^2 \qquad (2.3)$$

$$V_{dihedrals} = \sum_{dihedrals} \frac{V_n}{2}(1 + \cos(n\Phi - \gamma)) \qquad (2.4)$$

$$V_{VDW} = \sum_{i<j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}}\right] \qquad (2.5)$$

$$V_{electrostatic} = \sum_{i<j} \frac{q_i q_j}{\varepsilon R_{ij}} \qquad (2.6)$$

Bonded interactions

The interactions between atoms within molecules are described by the bonded terms, which are represented in equations **2.2**, **2.3**, and **2.4**.

Equation **2.2** represents the bond potential, which is used to model the covalent bond interactions. Bond stretching is approximated by a simple harmonic function that describes the oscillation about a reference bond length, r_0, with a force constant, k_b, (in kJ/mol/Å^2). [91,94–96]

The angle potential describes the angle bending energy. The angle bending between two covalent bonds is also modeled by a harmonic potential function, which is increased as the angle Θ deviates from the reference value, r_0, with a force constant, k_b (in kJ/mol/radian^2), as represented in equation **2.3**. [91,94–96]

The dihedral or torsion angle (Φ) is defined as the rotation of a bond, leading to the deviation of the relative positions of atoms 1 and 4 (Figure **2.1**). In equation **2.4**, t term is described by a sinusoidal function, and includes parameters for the torsion barrier, V_n (in kJ/mol), in which the magnitude represents the height of the potential barrier; the periodicity, n, indicating the number of cycles per 360º rotation about the dihedral; and the phase, γ (in degrees), that dictates the location of the maxima in the dihedral energy surface.[91,95,96]

In certain force fields, the dihedral potential may be divided into two terms, one for the proper dihedral angles and another for the improper dihedral angles. While the proper dihedrals are modeled by a periodic function as described in equation **2.4**, the improper dihedral term is a harmonic potential (equation **2.7**), which penalizes deviations from the improper reference dihedral angle (ω_0), according to the force constant (k_ω). This last term is used to maintain the chirality of tetrahedral centers and the geometry of planar groups.[91,96]

$$V_{improper\ dihedrals} = \sum_{impropers} k_\omega(\omega - \omega_0)^2 \qquad (\mathbf{2.7})$$

However, in the case of the AMBER force field, planarity is handled through the same formulation as the torsion term, where the phase is only 180º and the periodicity is 2 (equation **2.8**). Thus, it is possible to avoid the distortion of planar structures. [95]

$$V_{improper\ dihedrals} = \sum_{impropers} \frac{V_2}{2}(1 + \cos(2\Phi - \pi)) \qquad (\mathbf{2.8})$$

Non-bonded interactions

The non-bonded energy terms account for the potential energy of the non-bonded interactions formed by atoms within the molecule, separated by more than two or three covalent bonds, or atoms in different molecules. These terms account for the van der Waals (non-electrostatic) and the coulombic (electrostatic) interactions.[87,96]

The contribution of the non-bonded energy terms is considered important in simulations of biological molecules since a large number of non-bonded interactions occur between atoms.[96]

Non-electrostatic potential energy (equation **2.5**) is described with the Lennard-Jones (LJ) 12-6 term, represented by two parameters: a repulsive term that varies with distance raised to the 12th power and an attractive term that varies with distance raised to the 6th power. R_{ij} is the distance between atoms *i* and *j*.[91,95,96]

The electrostatic term (equation **2.6**) accounts for the contribution of electrostatic interactions. These interactions are treated using Coulomb's law between pairs of atomic partial charges, q_i and q_j, and varies linearly with the distance between atoms *i* and *j*, R_{ij}, taking into account for the appropriate dielectric constant (ε). [91,95,96]

2.2.2 Energy Minimization

The way in which the potential energy changes with the coordinates can be designated as the potential energy surface. In MM, the most relevant regions of this surface correspond to the energy minimum points, which are considered to be the stable states of the system. The energy surface may contain a large number of local minima, and the lowest energy minimum being called the global energy minimum.[87]

To identify the conformations of the system that correspond to minima on the energy surface, several minimization algorithms have been formulated. Nevertheless, some factors should be considered when choosing the appropriate one for a certain problem. The ideal minimization algorithm is the one that provides an efficient minimization using the fewest computational resources.[87]

The minimization method employed in this work was the steepest descent algorithm. This is a first-order derivative minimization algorithm, and it moves the system in the direction parallel to the net force, descending in a straight line on the energy surface of the system. [87]

In each iteration, the forces of all the atoms are calculated and the atomic positions ($\boldsymbol{r}_n$) are updated by the following expression:

$$\boldsymbol{r}_{n+1} = \boldsymbol{r}_n + \frac{\boldsymbol{F}_n}{\max(|\boldsymbol{F}_n|)} \cdot h_n \qquad (\mathbf{2.9})$$

where h_n is the maximum displacement, $\boldsymbol{F}_n$ is the force, and the notation $\max(|\boldsymbol{F}_n|)$ refers to the largest scaler force on any atom. This process finishes after a fixed number of iteration steps or when the gradient of the energy is lower than the predefined value. [87]

2.3 Molecular dynamics simulations

The dynamic behavior of molecules and complex systems can be described by an ensemble that represents the successive configurations of the system, as well as the probabilities of occurrence. Molecular dynamics (MD) simulation is a technique to produce a dynamical trajectory for a system, and sample the set of configurations, with the correct distribution.[87]

MD simulations provide a methodology for atomic-scale modeling. This method is powerful and widely used to predict the static and dynamic properties of substances directly from the analysis of interactions between particles over time. [96]

In MD, the atoms are considered classical particles, and the trajectory, which specifies the way positions and velocities of the particles in the system vary with time, is obtained by integrating Newton's second law of motion (equation **2.10**). This states that the force applied on a particle i ($\boldsymbol{F}_i$) is directly proportional to its acceleration ($\boldsymbol{a}_i$) times the mass (m_i).[87,96]

$$\boldsymbol{F}_i = m_i \, \boldsymbol{a}_i \qquad (\mathbf{2.10})$$

The force $\boldsymbol{F}_i$ exerted on each particle at a certain moment t is given by the negative derivative of the potential energy (V), which is modeled by the force field, with respect to the particle's position ($\boldsymbol{r}_i$),[96] represented in equation **2.11**:

$$\boldsymbol{F}_i = -\frac{dV}{d\boldsymbol{r}_i} \qquad (\mathbf{2.11})$$

Given that the acceleration of a particle i ($\boldsymbol{a}_i$) is the first derivative of its velocity ($\boldsymbol{v}_i$), and the second derivative of its position ($\boldsymbol{r}_i$), the force $\boldsymbol{F}_i$ can be used to compute the new atomic positions according to Newton's equation of motion [96]:

$$\boldsymbol{F}_i = m_i \, \boldsymbol{a}_i = m_i \frac{d\boldsymbol{v}_i}{dt} = m_i \frac{d\boldsymbol{r}_i^2}{dt^2} \qquad (\mathbf{2.12})$$

2.3.1 Integration algorithms

The Newton's equation of motion can be integrated using an appropriate integration algorithm to update the positions and velocities over time. The leap-frog algorithm is one of several algorithms used for this, wherein the velocities are evaluated in-between the position ($\boldsymbol{r}$) evaluation and vice versa,[96] as described by the following equations:

$$\boldsymbol{r}_{n+1} = \boldsymbol{r}_n + \boldsymbol{v}_{n+1/2}\Delta t \qquad (\mathbf{2.13})$$

$$\boldsymbol{v}_{n+1/2} = \boldsymbol{v}_{n-1/2} + \frac{\boldsymbol{F}_n}{m}\Delta t \qquad (\mathbf{2.14})$$

where $\boldsymbol{v}_{n+1/2}$ represents the velocity at the mid-step time ($t \pm 1/2\Delta t$).

The atomic position $\boldsymbol{r_n}$, at a time t, is used to compute the current force $\boldsymbol{F_n}$. Using this force and the previous mid-step velocity $\boldsymbol{v}_{n-1/2}$ it is possible to calculate the next mid-step velocity $\boldsymbol{v}_{n+1/2}$, allowing the determination of the next position $\boldsymbol{r}_{n+1}$.[96] Then, the velocity at a time t can be calculated using equation **2.15**:

$$\boldsymbol{v}(t) = \frac{1}{2}\left(\boldsymbol{v}_{n+\frac{1}{2}} + \boldsymbol{v}_{n-\frac{1}{2}}\right) \qquad (\mathbf{2.15})$$

The velocities "leap-frog" over the positions to give their values at $t + 1/2\Delta t$, and the positions "leap over" the velocities to calculate their new values at $t + \Delta t$, hence the name[96].

The size of the time interval should take into account the properties of the system and it must be considerably smaller than the characteristic time of the fastest motion studied. On the other hand, the time step should be as large as possible to speed up the simulation. In order to increase the integration time step, bond lengths are constrained to their reference values, using constraint algorithms. In this thesis, LINCS was used to remove the fast vibrations of the system, enabling the use of larger time steps.[96,97]

2.3.2 Periodic Boundary Conditions

In MD simulations, one problem to overcome is the system boundaries when working with a finite system. To investigate the properties of a complex biological system, we would need to perform a simulation in a box that must be sufficiently large in a way to have fewer particles colliding with the box walls. However, this is extremely expensive to compute. On the other hand, using a smaller "box", most of the particles of the system would be near the edge of the simulation box, experiencing different forces from those on the particles near the center.[98,99]

In order to minimize the effects of boundaries and avoid the usage of a large simulation box, periodic boundary conditions (PBC) are commonly applied in MD simulations, enabling the approximation of an infinite system by using a small number of particles. Using PBC, the simulation box is replicated throughout space forming an infinite lattice, where molecules can move between the adjacent box replicates, as it is represented in figure **2.2**. As a result, the surrounding atoms affect each atom in the central cell, without the boundary effects. In reality, the simulation does not use replicate boxes; rather, mathematical operations are used to simulate this periodicity. [100]

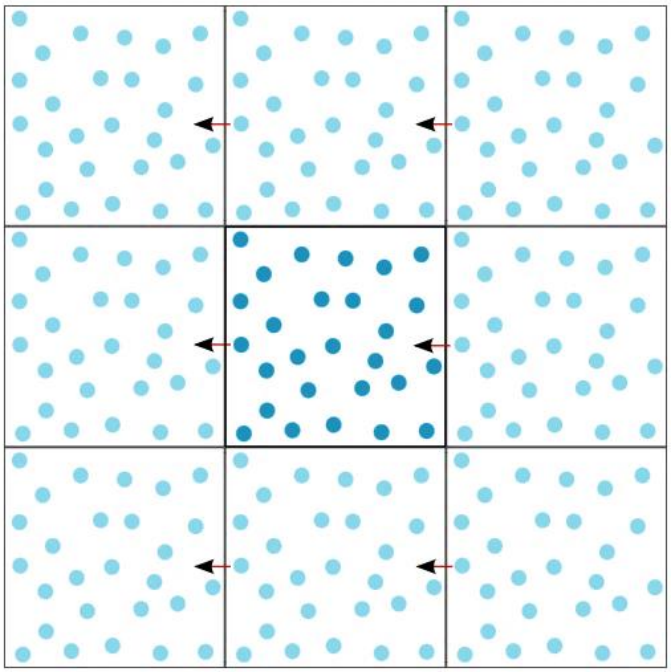


Figure 2.2 – Representation of a two-dimensional slice through a small portion of the system (the central box where the atoms are represented in dark blue) and the copies (atoms in light blue). Each copy is identical at the atomic level and each atom is subjected to the same time development as its image in all the other copies. When the dark blue atom leaves the central cell, its image will enter from an adjacent copy, which is represented by the vector displacements in the figure, which will cause the density to remain constant. (Adapted from ref. 100)

There are several shapes for simulation boxes that allow the use of PBC, and the choice of the ideal one will depend on the geometry of the system (figure **2.3A**). For example, the cubic box is the simplest and most commonly used periodic system. However, it is not the ideal shape to simulate an approximately spherical molecule. This would be inefficient since it requires more computer time to simulate a large number of solvent molecules surrounding the molecule to have a minimum distance between macromolecular images represented in figure **2.3B**.

The truncated octahedron and the rhombic dodecahedron are usually approximately spherical alternatives, thus being more suitable for simulations of spherical molecules, requiring fewer solvent molecules. In this work, the simulation studies are performed using a rhombic dodecahedron-shaped box.

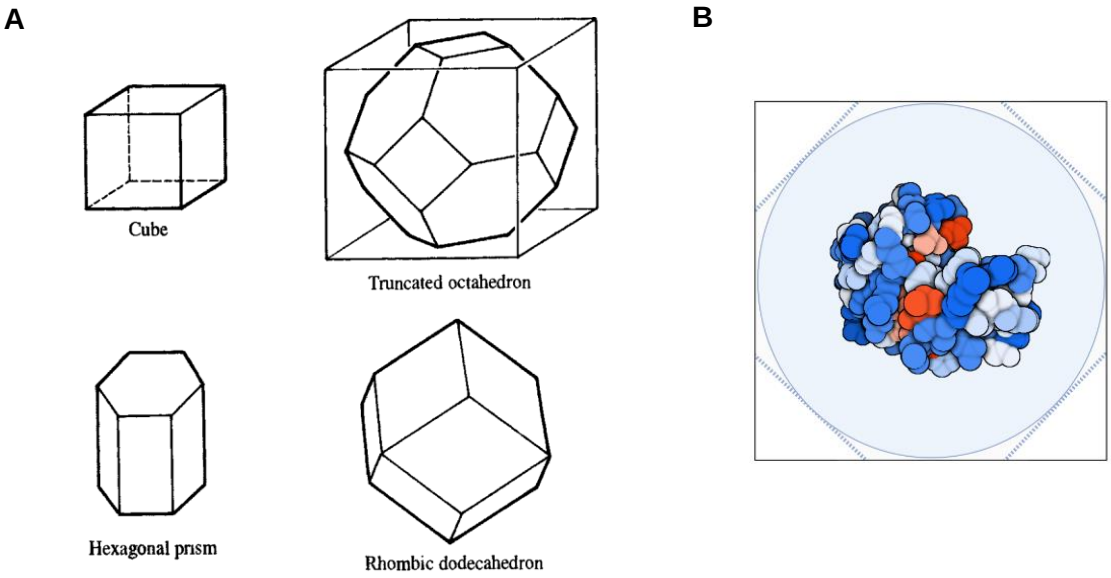


Figure 2.3 – (A) Possible shapes for simulation boxes used with periodic boundary conditions (PBC). Adapted from ref. 87 **(B)** Representation of a spherical molecule (PDB ID 1LYZ) inside of a cubic-shaped box. In this case, a large number of solvent molecules are needed to surround the molecule to have a minimum distance between the particle's images. Image created using Biorender.com.

When applying PBC, to avoid short-range interactions between a molecule and its images, a "non-bonded cutoff" is used so that each atom is close to only one image of every other atom in the system, and may not exceed half the shortest box vector - the "minimum image convention". [87]

2.3.3 MD simulations at constant temperature and pressure

MD simulation methods aim to simulate bulk properties, including the number of particles (N), the temperature (T), and the volume (V) of the system. In order to obtain thermodynamic averages, different configurations are sampled and then an average of the set of all configurations is calculated. This can also be called "ensemble average", since ideally it takes into account every possible microscopic state of the system. [96]

MD were traditionally performed under conditions of constant N, V, and energy (E), thus the time averages obtained are equivalent to microcanonical ensemble averages - the constant NVE ensemble. Other ensembles may be used to perform simulations, such as the canonical NVT ensemble, where the temperature (T) of the system is kept constant, as well as the values of N and V. [87,96]

Considering that most real systems are not thermally insulated from the exterior environment, simulations with constant N, T and pressure (P) - isothermal-isobaric ensemble - or with constant V, T, and chemical properties - grand canonical ensemble - are more realistic.[96,100]

In the canonical NVT ensemble, the probability distribution of the microscopic states of the system is computed by the Boltzmann distribution, represented in equation **2.16**:

$$P_i(N,V,T) = \frac{e^{-\frac{Ei(N,V)}{kT}}}{Q(N,V,T)} \qquad (\mathbf{2.16})$$

where E_i is the energy state of i, k is the Boltzmann constant and T the absolute temperature.

It is often scientifically more relevant to perform biomolecular MD simulations at a constant temperature. This allows the comparison of the simulation results with *in vitro* and *in vivo* experiments performed at a given temperature. Furthermore, it enables the study of the influence of the variation of this parameter on the simulation system.

Temperature is a thermodynamic property determined by the average kinetic energy of the system throughout the simulation time. In MD, to control the temperature, the system is usually coupled to an external bath that can provide or remove energy from the system. The Berendsen bath[97] is one example of a temperature coupling algorithm, wherein the rate of the temperature change is proportional to the difference of the temperature of the system (T) from the bath's temperature (T_{bath}):

$$\frac{dT(t)}{dt} = \frac{T_{bath} - T(t)}{\tau_T} \qquad (\mathbf{2.17})$$

The coupling parameter (τ_T) determines how the system temperature and the bath temperature are coupled, and it is inversely proportional to the coupling strength. In practice, the velocities are multiplied by a time-dependent scaling factor (λ) at each time step[97], given by:

$$\lambda = \left[1 + \frac{\Delta t}{\tau_T}\left\{\frac{T_{bath}}{T(t)} - 1\right\}\right]^{\frac{1}{2}} \qquad (\mathbf{2.18})$$

In some circumstances, using the NPT ensemble to perform MD simulations may be advantageous. Similar to temperature, the pressure can be controlled by coupling the system to a pressure bath. The Berendsen pressure algorithm, for example, scales the box coordinates and vectors at each time step for the pressure (P) to relax toward a reference point (P_{bath})[87,97]:

$$\frac{dP(t)}{dt} = \frac{P_{bath} - P(t)}{\tau_P} \qquad (\mathbf{2.19})$$

This algorithm uses a scaling factor (μ) to rescale the vectors and the box coordinates, given by the following equation[97]:

$$\mu = 1 - \frac{\beta \Delta t}{3\tau_P}\left(P_{bath} - P\right) \qquad (\mathbf{2.20})$$

where β is the isothermal compressibility of the system.

2.4 Preparation of the simulation systems

In this thesis, simulations of the SARS-CoV-2 RBDs of Omicron subvariants, BA.1, BA.2, and BA.5 were performed in complex with the hACE2. For the simulation of the free RBDs in solution, only BA.2 and BA.5 were studied since BA.1 was already studied (table **2.2**).

Table 2.2 - The systems simulated for each of the Omicron RBD, in complex with the hACE2 or free in water.

Omicron Subvariants	Simulated systems
BA.1	**RBD in complex with hACE2 protein**
BA.2	
BA.5	
BA.2	**RBD free in water**
BA.5	

The crystal structure of the wt SARS-CoV-2 S protein RBD bound to hACE2 (PDB ID 6M0J)[1] was used as a starting structure for atomistic MD simulations of the Omicron RBDs-hACE2 complexes to maintain consistency with previous simulations performed in our lab [64,85]. The structure contains two protein chains, the hACE2 (chain A) and the RBD (chain E). The chloride and zinc ions located in hACE2 were kept.

To be consistent with the published experimental findings about the Omicron BA.1 subvariant, a few adjustments were performed in both 6M0J chains based on the cryo-electron microscopy (cryo-EM) structure of the SARS-CoV-2 Omicron (BA.1) S protein bound to hACE2 (RBD-hACE2 refinement) (PDB ID 7T9L) [71,80]. Firstly, the RBD substitutions of the Omicron subvariants, BA.1, BA.2, and BA.5, were introduced using the PyMOL[30] mutagenesis wizard. Then, after the alignment of both structures (6M0J and 7T9L), the side chains of the residues from both hACE2 and RBD were rotated to obtain the interactions observed in 7T9L. hACE2 residues, H34 and K31, were rotated, respectively, for H34 to interact with the backbone of hACE2 S494, and to avoid clashes. The side chain of the substituted RBD R493, in BA.1 and BA.2, was also rotated to interact with hACE2 E35, while R498 was slightly deviated to prevent clashes.

For the simulations of the unbound Omicron RBDs, the initial used structures were the Omicron RBDs generated by substituting the appropriated residues in the wt RBD from the 6M0J structure, which corresponds to the hACE2 bound conformation of the RBD, using PyMOL. The hACE2 chain was then removed to perform simulations of the isolated RBDs.

Although the surface of the S protein is highly glycosylated and hACE2 contains three glycans at the interface with S RBD, *N*-glycans were not included in either of the simulation systems studied in this thesis. On the one hand, the RBD itself, as mentioned before, contains only an *N*-glycosylation site (N343) far from the region of interest, the RBM. Furthermore, taking into account the reported glycan shield that impacts the down-to-up conformational change in the S trimer, when one of the RBDs is in the "up" conformation it presents a fully accessible RBM and the S protein glycans no longer play a relevant role. On the other hand, the glycans would introduce degrees of freedom that could complicate sampling, especially in the case of the RBD-hACE2 systems. Therefore, we opted for a reductionist approach by simulating both systems without glycans, although we are investigating this matter in a separate work.

2.4.1 Setup for MD simulations of the Omicron RBD-hACE2 complexes

MD simulations of the Omicron subvariants' RBDs bound to hACE2 were performed with the GROMACS 2020.3 package[101], using the Amber14sb[95] force field, as mentioned earlier.

The Omicron RBD-hACE2 complexes were solvated with water in a rhombic dodecahedron-shaped box, using the transferable intermolecular potential three-point (TIP3P) water model to treat water molecules[102]. The minimum distance between the complex and the box walls was set to 1.2 nm. Chloride (Cl^-) and sodium (Na^+) ions were added as required to neutralize the system and to have a specific ionic strength of 0.1 M.

The systems were energy-minimized using the steepest descent method for a maximum of 50000 steps using position restraints on the heteroatom positions by restraining them to the crystallographic coordinates with a force constant of 1000 kJ/mol/nm in the X, Y, and Z directions.

Afterwards, the system was initialized through five stages of 100 picoseconds (ps) each. In the first four steps, all heavy atoms were restrained using a force constant of 1000 kJ/mol/nm, and at the final step, only the alpha carbons (Cα) were position-restrained using the same force constant.

In the first stage, the initial velocities were generated and the Berendsen temperature coupling algorithm was used to maintain the simulation at 300 K, using a temperature coupling constant of 0.01 ps, without pressure coupling. Temperature coupling was applied to the protein together with ions and to the solvent atoms separately. In the second stage, the Berendsen temperature algorithm continued to be applied, although with a coupling constant of 0.1 ps. The third stage kept the same temperature coupling parameters, but pressure coupling was introduced with the Berendsen pressure coupling algorithm, in which a pressure coupling constant of 5.0 ps was applied isotropically. In the fourth stage, the temperature algorithm was switched to the V-rescale[103], with a temperature coupling constant of 0.1 ps. The pressure algorithm was also changed to Parrinello-Rahman[104], with a pressure coupling constant of 5.0 ps. The last stage comprises the same temperature and pressure settings as the fourth stage, but position restraints are only applied to the Cα atoms.

During the simulation, the equations of motion were integrated using a timestep of 2 femtoseconds (fs). The temperature was maintained at 300 K, using the V-rescale algorithm with a time constant of 0.1 ps, and the pressure was maintained at 1 bar using the Parrinello-Rahman pressure coupling algorithm, with a time constant of 5.0 ps, applied isotropically. Long-range electrostatic interactions were treated with the Particle Mesh Ewald (PME)[105] method, using a grid spacing of 0.12 nm, with a cubic interpolation. The neighbor list was updated every twenty steps, and the cutoff scheme used was Verlet with 0.8 nm as the real space cut-off radius. The LINCS algorithm was used to constrain bonds containing hydrogens.

Finally, for the RBD-hACE2 complexes, five replicates of each system were simulated for a total time of 2 μs. Only 500 nanoseconds (ns) of these simulations were used for analysis, and the remaining simulation trajectory being considered equilibration time.

2.4.2 Setup for MD simulation of the unbound Omicron RBDs

Atomistic MD simulations of the unbound RBDs of the Omicron subvariants, BA.2 and BA.5, were performed under the same conditions as those performed with the complexes.

The preparation of the system as well as the initialization stages were performed with the same protocol. The difference is in the starting structure, where each RBD structure was obtained by removing the hACE2 chain from the original structure of the complex (PDB ID 6M0J). The water molecules bound to the RBD were maintained and this system was inserted in a rhombic dodecahedral box with a minimum distance between the protein and box walls of 1.2 nm, that was filled with water and ions (following the protocol described above to reach neutralization and an ion strength of 0.1 M).

Simulations of each system were performed for 7 μs in five replicates. The first 3 μs of all simulations were considered as equilibration time and the remaining trajectories were used for analysis.

2.5 Analysis of MD simulations

For the visualization and rendering of simulation snapshots, VMD[106] and PyMOL molecular graphic viewers were used. The analyses of the simulations were performed using GROMACS tools, as well as the MDAnalysis[107] package. All the plots were generated by the Matplotlib[108] library of Python.

2.5.1 Principal Component Analysis (PCA) and Energy Surface Landscape

Principal component analysis (PCA) is a dimensionality-reduction method used to condense a large set of correlated variables into a more manageable number of variables that still retain the majority of the variation in the data set to improve data analysis and visualization. This is accomplished by transforming the original set of variables into a new uncorrelated set of variables, the principal components (PCs), which are ordered decreasing by the variance they are able to capture. Thus, the first PC accounts for the most variance, while each subsequent PC accounts for less and less variance. By selecting the first PCs it is possible to reduce the data dimensionality.[109–111]

In an MD simulation with atomic-level detail, exploring every conformation of a protein is challenging. PCA of MD simulation is a popular method to account for the largest fluctuations of a protein on a reduced-dimensional free energy landscape.[111]

In this thesis, PCA was applied to the 3*N*-dimensional space of conformational coordinates obtained from MD simulations of the unbound Omicron RBDs, BA.2 and BA.5, *N* being the number of RBD residues. Before, each conformation was translationally and rotationally fitted to the wt RBD (PDB ID 6M0J) core Cα atoms. Using MDAnalysis, the PCs were previously determined from the trajectories of the wt and the other VOCs[85], considering only the coordinates of the RBD's Cα atoms. The two most representative PCs were selected so that RBD structures for each simulation frame could be projected as points in this two-dimensional space, simplifying the representation of each RBD conformational space. For both Omicron RBDs, BA.2 and BA.5, we projected the trajectories into this conformational space separately.

The computation of the two-dimensional density and energy landscape in the resulting PC space was performed. The probability density function for each trajectory projection was estimated using a Gaussian kernel estimator implemented in LandscapeTools' *get_density* program[110,112]. This procedure defines a probability density function, $P(r)$, with the values of $P(r)$ being stored for the position of each data point and the bins of a two-dimensional uniform grid with a mesh size of 0.5 Å. The energy is then calculated from $P(r)$, assuming a Boltzmann distribution and setting the energy value to 0 at the point located at the highest density (P_{max})[110]:

$$E(r) = -RTln\left(\frac{P(r)}{P_{max}}\right) \qquad (\mathbf{2.21})$$

where the P_{max} is the maximum of the probability density function, $P(r)$, R is the Boltzmann constant and T is the temperature to be considered in the Boltzmann distribution.

The analysis of the energy surface landscapes was achieved by determining the energy minima and corresponding basins. The basins were defined as the set of all conformations whose steepest descent path along the energy surface leads to a particular minimum[110]. Here, the steepest descent paths for each grid cell were computed, with each conformation inheriting the path of its corresponding grid cell. Landscape regions with E > 6 k_BT were discarded, resulting in the final set of basins for each data set.

2.5.2 Residue interaction networks in Protein Molecular Dynamics

Due to the data-intensive nature of the method, extracting relevant information from MD simulations is a complex and time-consuming procedure requiring several computational tools to perform this analysis. Of all the techniques, the residue interaction networks (RINs) have proven to aid the study of protein structures. A RIN is a graphical representation wherein nodes represent amino acid residues and the interactions between those residues are illustrated by edges.[113]

There are several approaches to studying protein structures as a network. The most common is by representing a single structure neglecting the dynamic properties of each amino-acid residue concerning the rest, thus called static-RINs. More recently, the dynamic-RINs were developed by several authors, where the RIN-based methodology was applied to MD simulations, taking into account the protein dynamics and the analysis of averages from trajectories. The software Residue Interaction Networks in Protein Molecular Dynamics (RIP-MD) was designed as an intuitive visualization tool of pairwise residue interactions and is used to generate both static and dynamic RINs.[113]

The workflow of RIP-MD software begins by giving an MD trajectory or a static protein structure (PDB file) and the input parameters defining each interaction. RIP-MD defines several types of interactions between the residues in a protein structure.[113] In this thesis, the study of inter and intramolecular interactions was based on the parameters defined for hydrogen bonds, salt bridges, and pi-pi (π-π) interactions (Table **2.3**).

Table 2.3 – Definition of each interaction type in RIP-MD (Adapted from ref. 113). (*) If His presents a protonated nitrogen atom it is considered as a cation. If it is not protonated it is considered a pi-system.

Interaction type	Mathematical Formulation	Input Parameters
Hydrogen Bonds	Dist (donor, acceptor) ≤ d θ(C−H,acceptor) ≥ a	d = 3 Å a = 120°
Salt Bridges	Contacts between NH/NZ groups of Arg/Lys and OE*/OD* in Asp/Glu ≤ d	d ≤ 6 Å
π–π interactions*	Distance between aromatic rings ≤ d	d ≤ 6 Å

After the pre-processing step, the software searches for interactions between all atoms in each frame from the trajectory or the single structure from PDB. Then, RIP-MD generates the output files defining the RINs which can be further characterized in network visualization tools such as Cytoscape.[114]

RIN Analysis of RBDs-hACE2 Complexes

In order to determine the residues that contribute for the interaction between the RBDs (wt, Omicron BA.1, Omicron BA.2 and Omicron BA.5) and hACE2, and the prevalence of these interactions, RIN analysis of these intermolecular interactions through all the simulations were performed.

Firstly, the equilibration time was eliminated from all replicates of the simulations with the RBDs-hACE2 complexes.

Since the region of study is the interface between each RBD and the hACE2 protein, we only used the interface structures of each complex as an input for RIP-MD. To obtain these, MDAnalysis was used to identify the RBD residues that are in contact with the hACE2 through the simulation, applying a distance cutoff value of 6.0 Å.

Then, the remaining trajectories and the structures of each RBD-hACE2 interface were used as inputs for RIP-MD software, as well as the standard parameters defining the interactions of interest, hydrogen bonds, salt bridges and pi-pi interactions, found in table **2.3**.

Once the interactions were determined, the intermolecular interaction networks were treated using a Python script created to organize the nodes and edges information (i.e. residues and interactions, respectively) and visualized using Cytoscape. Only the interactions that were present in over 30% of the simulation frames were selected.

RIN Analysis of RBD MD simulations

To characterize the interactions observed in the basins obtained from the PCA analysis of the RBD simulations, pairwise intramolecular residue interactions were analyzed for the 5000 lowest energy conformations obtained for the most populated “open”, “closed”, and “shrunken” conformation basins of the energy surface landscapes of each RBD variant, using RIP-MD. The interaction networks were visualized using Cytoscape. Only the interactions that were present in over 50% of the simulation frames were selected. Additionally, the selected intramolecular interactions are those that are established by RBM residues or others around this region.

3. Results and Discussion

3.1 MD simulations of the Omicron RBD-hACE2 complexes

In this thesis, the main goal is to provide a detailed characterization of the interaction of the Omicron variant with the hACE2 receptor. Therefore, atomistic MD simulations with the RBDs of the three Omicron subvariants of concern, BA.1, BA.2, and BA.5, complexed with the hACE2 protein were performed in solution (figure **3.1B**), performing in total three simulations with five replicates each. The wt RBD-hACE2 simulations, previously performed by the Protein Modeling group [64], were used as control. Notice that these simulations were performed under the same conditions as for the Omicron subvariants.

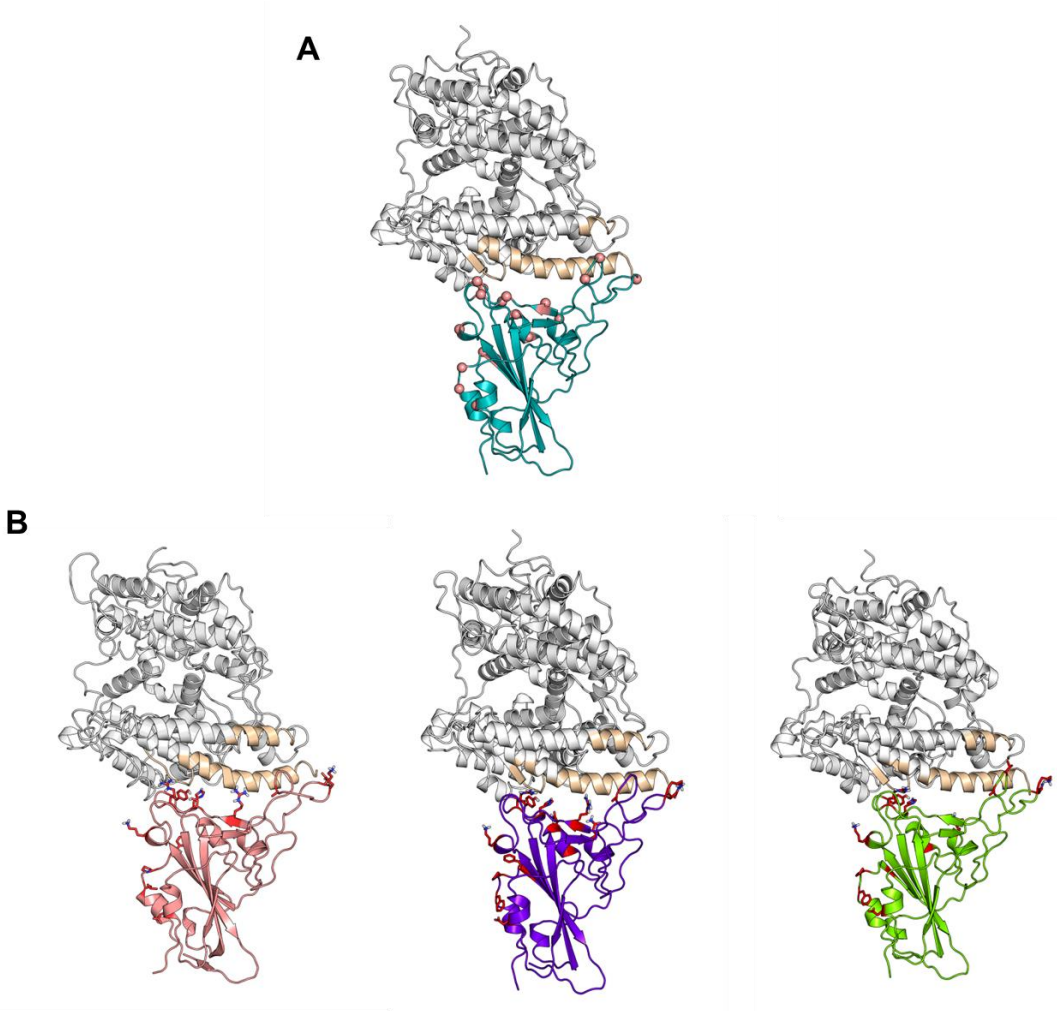


Figure 3.1 - Representation of the RBDs complexed with the hACE2 protein. **(A)** wt-RBD hACE2 complex with the residues that were substituted represented as pink spheres. **(B)** The substitutions of each Omicron subvariant are highlighted in red. Each RBD is represented in cartoon with different colors: wt in blue, Omicron subvariant BA.1 in orange, Omicron subvariant BA.2 in purple, and Omicron subvariant BA.5 in green. hACE2 interface region with the RBDs are highlighted in the structure in light-yellow.

3.1.1 Root-mean square deviation (RMSD)

The root-mean square deviation (RMSD) was calculated for wt and the simulated Omicron RBD-hACE2 complexes, using MDAnalysis. This measure assesses the similarity between each of the simulated complex with a reference structure that in this case was the X-ray structure of the wt RBD-hACE2 complex (PDB ID 6M0J). Lower RMSD values indicate that it deviates less from the original conformation.

It is noteworthy that this system contains two proteins with different dynamics. Regarding the RBD chain, it can be divided into two regions: a rigid core, and a more flexible one, the RBM. In the case of hACE2, its structure has been described as being composed by two subdomains, a C-terminal subdomain (CTD, residues 103-289, 398-416 and 431-615) and an N-terminal subdomain (NTD, residues 19-102, 290-397 and 417-430), which can be observed in figure **3.2**. This nomenclature should not be confused with the domains defined in the structure of Spike, described in the Introduction.

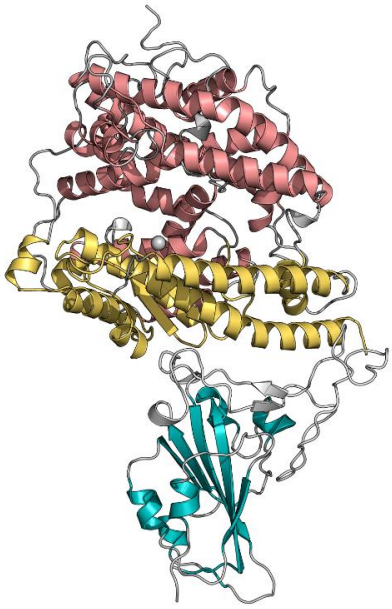


Figure 3.2 - Representation of RBD-hACE2 (PDB ID 6M0J), highlighting the RBD core, colored cyan, RBM, colored gray, N-terminal subdomain (NTD) of hACE2, colored yellow, and the C-terminal subdomain (CTD) of hACE2, colored in pink. The zinc ion of the catalytic site is represented with a gray sphere.

The Protein Modeling group had performed simulations with the free hACE2 in solution, which allowed us to observe a displacement of the NTD in relation to the CTD. Visual inspection of the trajectories obtained in the simulations of the complexes in water confirmed this dynamic behavior. To understand whether this conformational dynamic occurs when hACE2 is bound to the RBD, it was performed a measure of the distance between the two hACE2 subdomains' helices in function of the simulation time for all replicates of each system, using the GROMACS tool *gmx distance*. The results obtained (Supplementary Figure **S1**) showed that CTD moves with respect to the NTD subdomain that interacts with the RBD.

Additionally, to capture this dynamics using the RMSD calculation, the structure was superimposed by fitting only the Cα atoms of the regular structure of the RBD core and of the hACE2 NTD. The RMSD was calculated separately for each region of the complex: the RBM, the RBD core, and for the subdomains of hACE2, NTD, and CTD. Observing the average Cα RMSD of all replicates for each system (Supplementary Figure **S2-S5**), we could confirm the effect of the displacement of the CTD subdomain on the hACE2 RMSD.

Since the main aim was to analyze the complex interface that comprises the hACE2 NTD bound to the RBD, we focused our analysis on the RMSD calculated for the Cα atoms of the RBD and hACE2 NTD subdomain and excluded the hACE2 CTD from the calculations to remove the noise added by the motion of the CTD relative to the NTD (figure **3.3**).

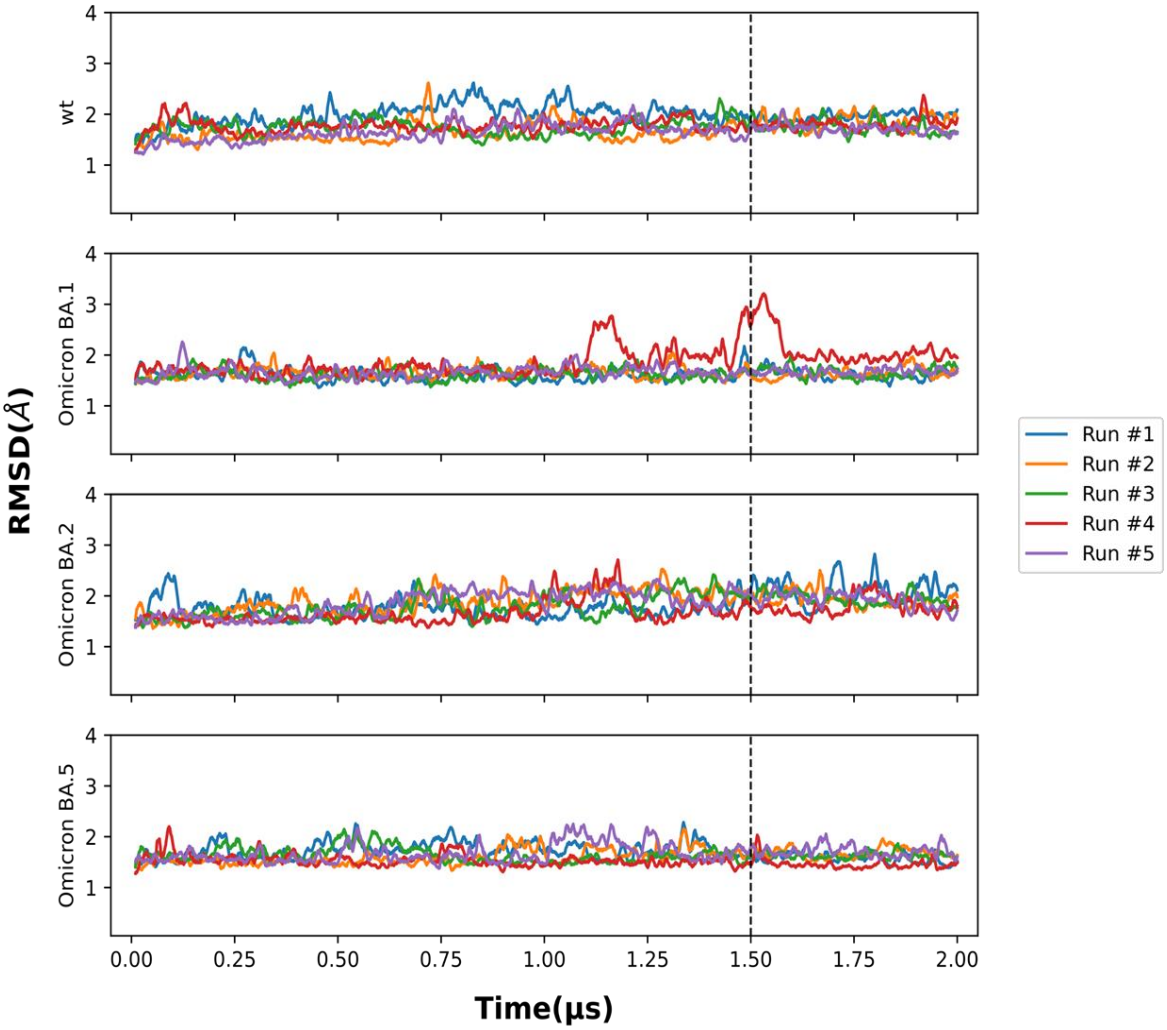


Figure 3.3 – Moving average of the root mean square deviation (RMSD) of the Cα atoms of the RBD and hACE2 NTD subdomain. Data are shown for the five replicas for each system tested. RBD Core and NTD Cα were fitted against the X-ray structure from PDB ID: 6M0J. The moving average was calculated using the neighboring 50 frames. The first 1.5 μs of simulation were considered for equilibration (black dashed line) and the remaining frames were used for further analysis.

By analyzing the overall RMSD values over the simulation trajectory in figure **3.3**, wt and Omicron BA.5 interfaces have stabilized at 1.5 μs. For the BA.1 RBD-hACE2 interface, although four replicates present quite stable RMSD values through time, large fluctuations can be observed in replicate 4 (in red) between 1.1-1.6 μs. As for the BA.2 complex, all the replicates are reasonably stabilized at 1.5 μs, although some fluctuate around the average value.

The possible cause of these fluctuations and deviations from the initial RMSD values is the introduction of new residues at the RBD interface with the hACE2 protein. Omicron subvariants contain 15 to 16 substitutions, 8 to 10 of each are located near the binding interface. So, when comparing to wt, it is clear that, for Omicron BA.1 and BA.2, the RBD is adjusting its position towards hACE2. As for Omicron BA.5, it appears to have stabilized before the other simulated systems. The reason may be the fewer substitutions on the RBD interface that impact the binding.

Taking into account that in the four complexes most replicates do not show large deviations from the X-ray structure after 1.5 μs, it can be stated that the RBD-hACE2 interfaces have overall stabilized at 1.5 μs, thus the trajectories before this point were considered equilibration and the remaining frames were used for further analysis. However, more simulation time would be required to clarify this matter.

3.1.2 Root-mean square fluctuation (RMSF)

The root-mean square fluctuation (RMSF) was calculated for alpha carbons (Cα) separately for the hACE2 and for the RBD for all replicates of each system using MDAnalysis. RMSF measures the fluctuations of atomic positions in the trajectory relative to its average position, after fitting to a reference frame, which in this case was the first one. The analysis is useful to reveal the most flexible regions of the system which corresponds to higher RMSF values.

In figure **3.4**, the average RMSF per residue of hACE2 and RBD over the last 500 ns of the simulations are represented, as well as the structure of the RBD-hACE2 complex colored by average RMSF values to represent the most stable and most flexible regions

After the system is equilibrated, overall average RMSF values of hACE2 and RBD show that there are no significant differences between the complexes.

Regarding hACE2 region, it is possible to observe higher flexibility between residues 115 and 125 (figure **3.4A**). Based on the hACE2 structure colored by average RMSF values, this region appears to be disordered, hence the most fluctuating region. Additionally, the regions with more flexibility in hACE2 corresponds to the movement of the C-terminal with respect to the N-terminal subdomain (figure **3.4C**). This correlates with the RMSD values of each subdomain, shown before, demonstrating that hACE2 presents different conformations throughout the simulations.

.

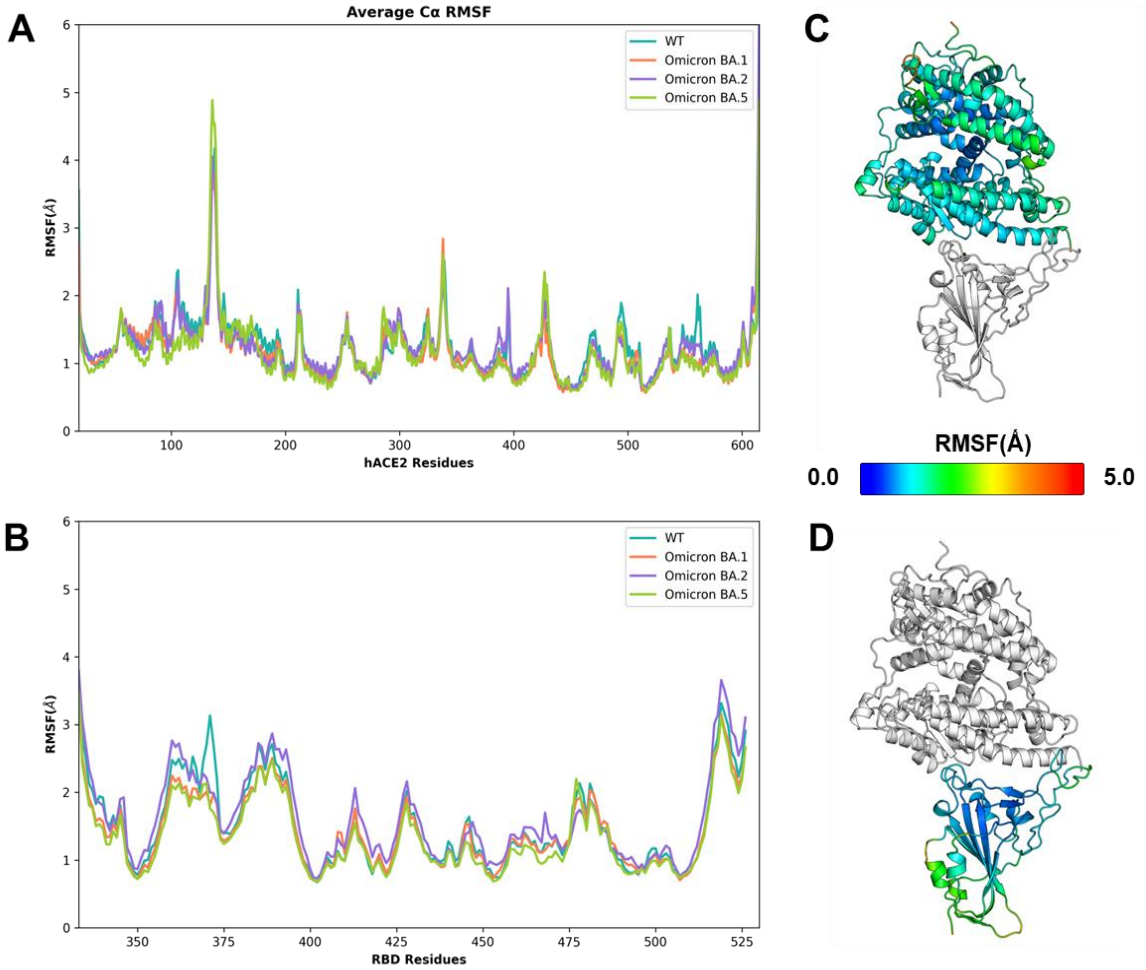


Figure 3.4 – Average alpha carbon (Cα) RMSF per residue computed over the last 500 ns of the simulations (equilibrated system). RMSF was computed separately for the **(A)** hACE2 and for the **(B)** RBD for each system and the average over the five replicates was calculated and plotted in the figure. Wt, Omicron BA.1, Omicron BA.2, and Omicron BA.5 are represented in blue, orange, purple, and green, respectively. **(C)** The average Cα RMSF of the hACE2 and **(D)** the RBD are represented in the wt RBD-hACE2 complex. The colors are distributed over the range of RMSF values, with blue representing the lowest (most stable) and red the highest (most flexible).

The values of average RMSF per RBD residue indicate that this protein bound to hACE2 presents a similar pattern among the different systems (figure **3.4B**). Observing figure **3.4D**, the most stable region corresponds to the RBD core, which is expect given its β-sheet structure. The C-terminal of the RBD and the helix with residues 370-375 are the most flexible regions on the RBD, followed by the RBM region (residues 450-500), which is explained by the quite dynamic RBM loop.

However, comparing the RBDs from each complex, it is possible to observe that Omicron BA.2 RBD bound to the hACE2 presents a slightly more flexible structure than the wt, as for the Omicron BA.5 RBD seems to be the most stable one of the variants (Figure **3.5**). These results correlate with the low RMSD values during the last 500 ns of the simulations for Omicron BA.5, and with oscillations in RMSD values for BA.2.

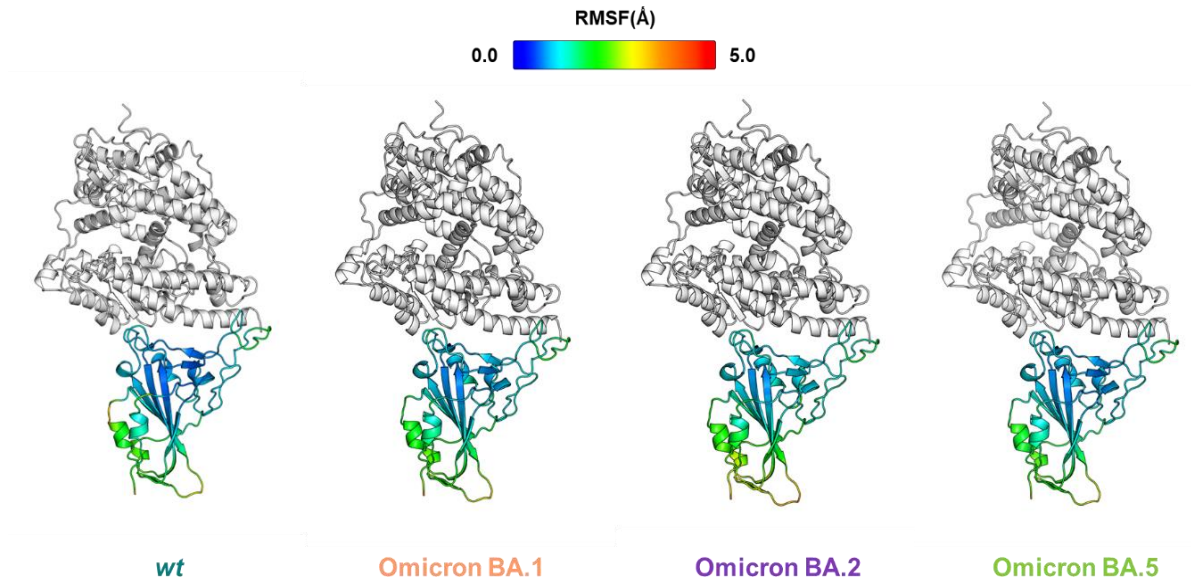


Figure 3.5 - Structures of the wt, Omicron BA.1, BA.2 and BA.5 RBD-hACE2 complexes colored by the average Cα RMSF. The colors are distributed over the range of RMSF values, with blue representing the lowest (most stable) and red the highest (most fluctuating).

3.1.3 RBD-hACE2 Interface: Secondary structure analysis

In order to evaluate the stability of the interface between the RBDs and hACE2, a secondary structure analysis of all simulations was performed, using the GROMACS tool *gmx do_dssp*. This tool was used to read the trajectory file and compute the secondary structure for each residue and frame. Afterward, the information obtained from *do_dssp* was used as an input to an in-house Python script to analyze whether the secondary structure of the hACE2 helix (residues 19-53) and the RBM (residues 438-506), highlighted in figure **3.6A**, that constitute the interface, were maintained in the last 500 ns of all simulations.

Observing the results, for all systems simulated, the RBD interface with hACE2 protein is constituted mostly by coil secondary structure (~76%), followed by beta-sheets (~20%). Additionally, there are traces of helices (~4%). There are no significant differences between the complexes (Figure **3.6B**).

Regarding the hACE2 interface, the plots show that the N-terminal helix maintains the secondary structure among all systems (~98%), although there are minor losses of the regular structure (~2% coil). For Omicron BA.2, there are more residues losing its secondary structure during the simulation, leading to higher mobility of this helix in relation to the RBD. This event can explain the fluctuations of the RMSD values of the Cα Omicron BA.2 RBD-hACE2 interface. In order for BA.2 RBD substitutions to adapt to new interactions with hACE2 there may be few movements of the interface (Figure **3.6C**).

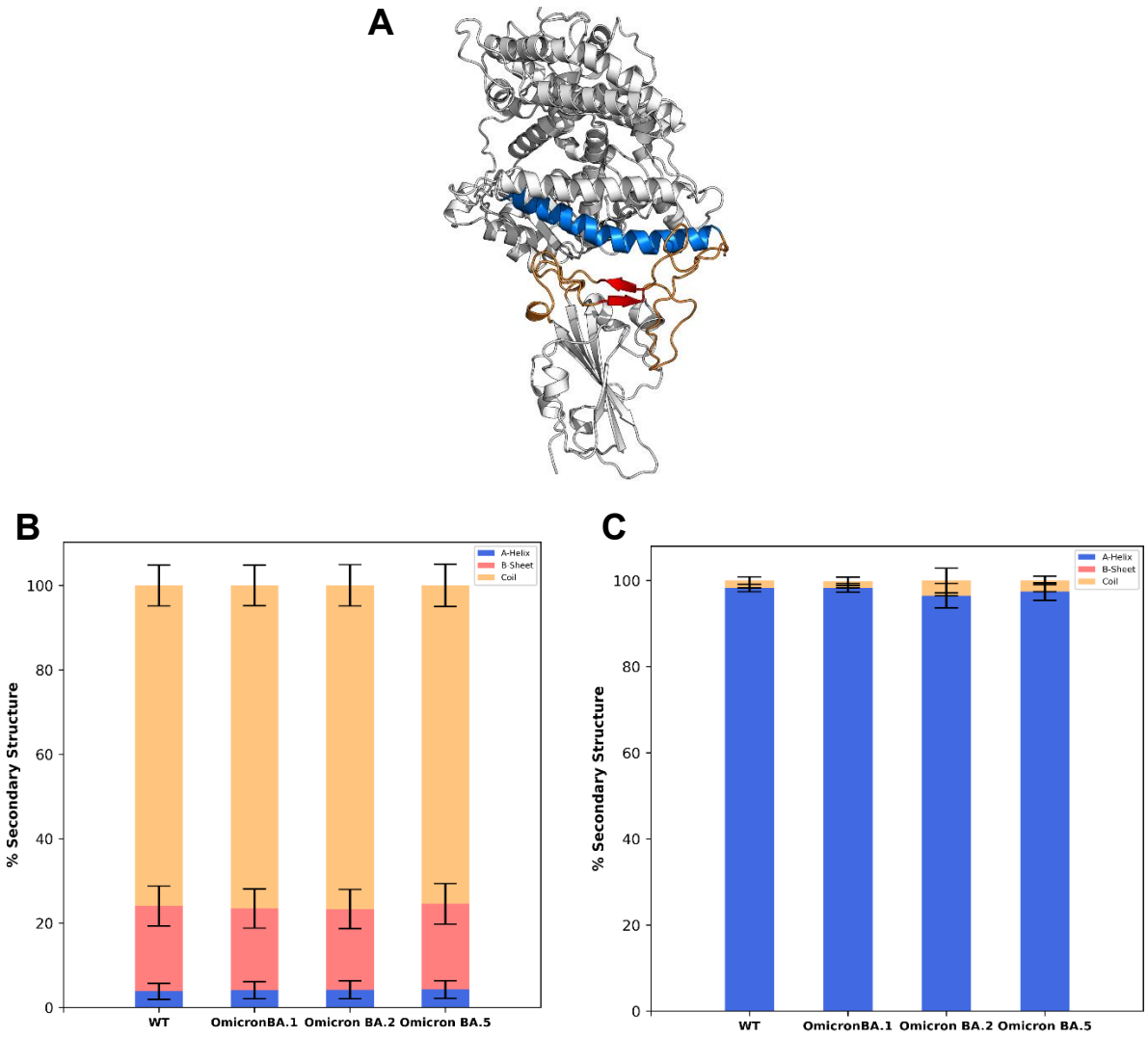


Figure 3.6 – (A) Representation of the RBD-hACE2 complex (PDB ID 6M0J) with the interface region highlighted by the secondary structure: alpha-helix (A-Helix) in blue, beta-sheet (B-Sheet) in red and coil in orange. Therefore, hACE2 helix that interacts with the RBM is colored blue, and the RBM loops and the beta-sheets are colored in orange and red, respectively. **(B)** Average percentage of secondary structure for the RBM region and **(C)** for the hACE2 helix during the last 500 ns of all replicates of each system. Each type of secondary structure is colored as the representation in **(A)**. The values correspond to the average obtained in the five simulations for each complex and errors are expressed as twice the standard deviation obtained by bootstrapping.

Overall, the secondary structure of the interface is not altered throughout the simulations and in comparison with all RBD-hACE2 complexes, therefore, it can be considered stable and there are no major alterations that can impact the binding of the variants.

3.1.4 RBD-hACE2 Interface Contact Analysis

For a detailed characterization of the binding interface between wt and Omicron subvariants and hACE2, RIN analysis was performed for the equilibrated trajectories of each simulation to obtain the intermolecular interactions established between the wt, BA.1, BA.2 and BA.5 RBDs and the hACE2 protein. From the identified interactions, the selected are those that were present in over 30% of the simulation frames (Supplementary figure **S6**). The remaining interactions are detailed in the supplementary tables **S1-S4**.

Based on the results, the RBD interface area can be primarily classified into three regions, each involving important interactions with hACE2. To an easier interpretation of each location, here it is adapted the naming convention for the RBD given its resemblance to human torso[115] (figure **3.7A-C**). The different regions of the RBD that interact with hACE2 are the neck, that comprises the two beta-sheets of the RBM, and the two shoulders corresponding to the loops situated in opposite sides, being the right one the RBM "ridge" (figure **3.7A**).

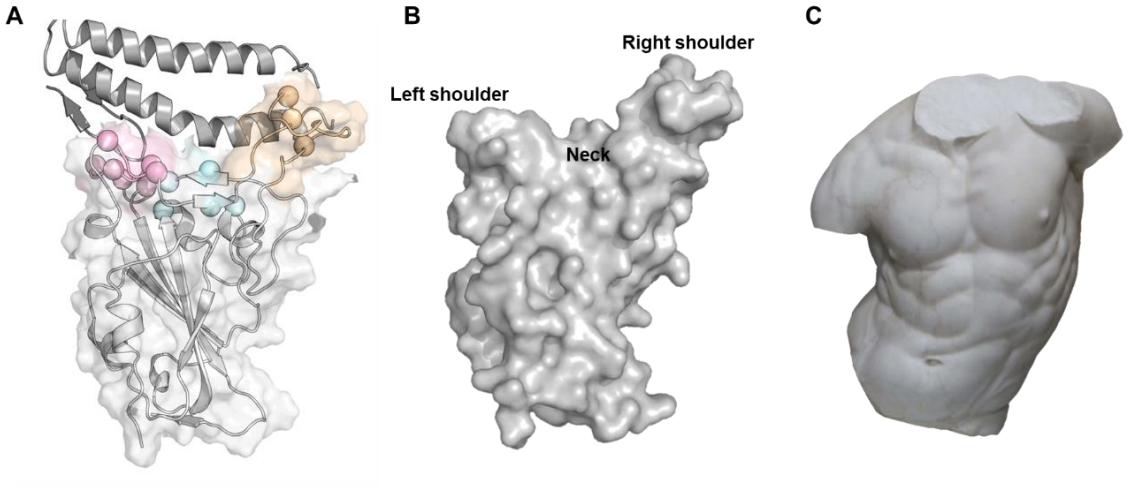


Figure 3.7 – RBD interface area definition based on RIP-MD results and RBD anatomy. **(A)** Representation of the surface and cartoon RBD with the three interface regions, comprising key residues (spheres) that interact with hACE2. The regions highlighted in pink, cyan and orange, correspond to the left shoulder, neck and right shoulder. **(B)** RBD surface labeled to correspond to the **(C)** human torso (Torso Gaddi, Wikipedia), used by analogy for a better definition of the RBD interface regions with hACE2.

Wt RBD-hACE2 interface

The identified interactions between wt RBD and hACE2 correspond to those that are described in the literature. Overall, the binding interface is mostly characterized by polar and charged residues, thus having more hydrogen bonds and salt bridges, although there are also hydrophobic interactions.

In the left shoulder of wt RBD, the residues T500, Y449, Y505, and G502 form hydrogen bonds with hACE2 residues Y41, D38, E37, and K353 (figures **3.8A, B**). Additionally, there are relevant interactions that are not shown in the RIN since they are slightly below the selection cutoff. The hydroxyl group of Y41 also establishes a transient hydrogen bond with N501 with ~27% prevalence, and D38 and K353 form other transient hydrogen bonds with Q498 with ~20% and ~23%, respectively.

On the RBD neck, the residues K417, R403, and E484 interact with D30, E37 and K31 of hACE2, respectively, forming salt bridges. Also, Q493 forms transient hydrogen bonds with hACE2 E35 and K31, although the prevalence of the Q493-K31 is below the selection cutoff (~22%).

On the right shoulder, F486 forms a stable pi-stacking with hACE2 Y83. Moreover, there is a hydrogen bond network, in which the hydroxyl group of Y83 and the side chain of Q24 interact with both side chain groups of the N487.

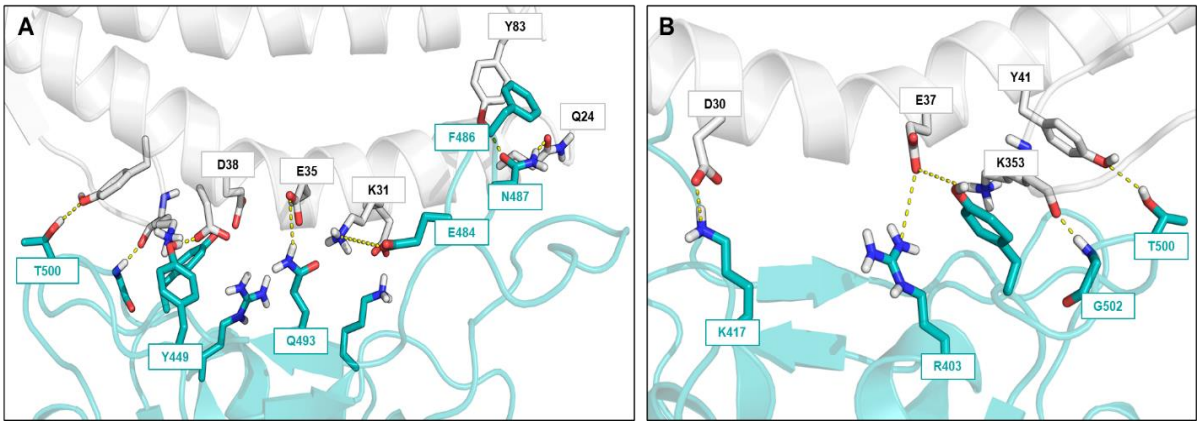


Figure 3.8 – Representation of the binding interface of wt RBD (cyan cartoon) with hACE2 (white cartoon). The identified pairwise interactions present in over 30% of the simulation frames are represented in yellow dashed lines and the participating residues from RBD and hACE2 are represented in sticks, colored by atom group, and labeled in cyan or in black, respectively. **(A)** Front-view of the overall binding interface, where interactions are established by the RBD's left shoulder (residues 449, 500, 502, and 505), neck (403, 417, 484 and 493) and right shoulder (486 and 487). **(B)** View of the biding interface from behind.

Omicron BA.1 RBD-hACE2 interface

Comparing with wt, the interface between Omicron BA.1 RBD and hACE2 presents major differences since there are fifteen substitutions located in the RBD, ten of which are present in the binding interface – K417N, Q493R, and E484A occurring on the RBD neck; G446S, G496S, Q498R, N501Y and Y505H located on the left shoulder; and S477N, T478K on the right one.

The RBD neck region is altered with the loss of two key interactions, the K417-D30 and K31-E484 salt bridges since the RBD residue 417 is altered to an asparagine (N417) and 484 is altered to an alanine (A484). Additionally, on the left shoulder, the altered H505 is responsible for the loss of Y505-E37 hydrogen bond. Nevertheless, new interactions are established between BA.1 RBD and hACE2 (figures **3.9A, B**).

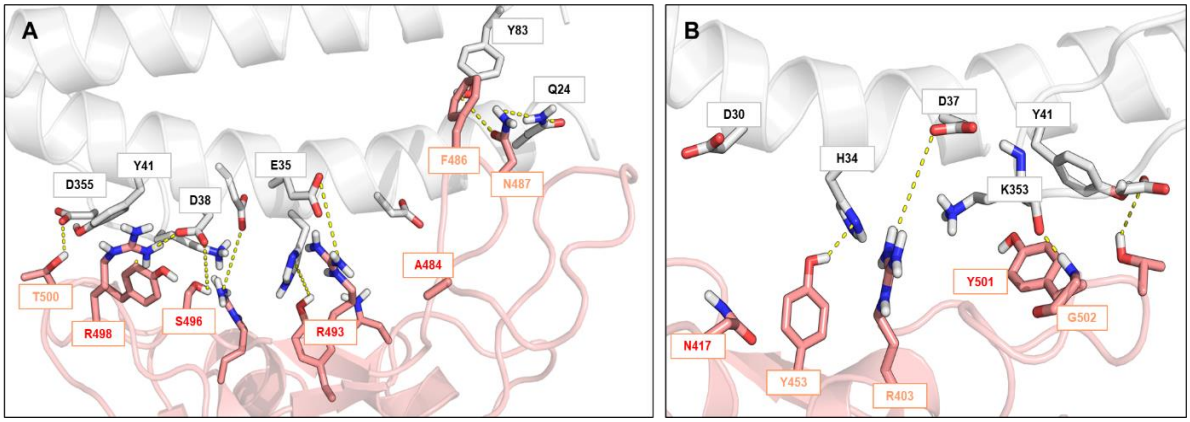


Figure 3.9 – Representation of the binding interface of Omicron BA.1 RBD (salmon cartoon) with hACE2 (white cartoon). The identified pairwise interactions present in over 30% of the simulation frames are represented in yellow dashed lines and the participating residues from RBD and hACE2 are represented in sticks, colored by atom group, and labeled in salmon or in black, respectively. The RBD substitutions are labeled in red. **(A)** Front-view of the overall binding interface, where interactions are established by the RBD's left shoulder (residues 496, 498, 500 and 501), neck (403, 453, and 493) and right shoulder (486 and 487). **(B)** View of the biding interface from behind.

On the left shoulder, the single hydrogen bond between the wt N501 and the hACE2 Y41 is replaced by a stable pi-stacking between the aromatic side chains of the Y41 and the substituted RBD Y501. This interaction is also observed in other VOCs that contained N501Y substitution and it is highly associated with stronger binding to hACE2 receptor given the interaction nature.

In addition, the hACE2 residues D38 and D355 form new hydrogen bons with other left shoulder RBD residues, the substituted S496 and T500, while the wt Q498-D38 hydrogen bond is replaced with a transient salt bridge between the substituted R498 and D38. The G502-K353 hydrogen bond is maintained, as well as Y449-D38 hydrogen bond, although is not shown (~23% simulation frames).

From wt to Omicron BA.1, the substitution Q493R has altered the RBD neck interface with hACE2. Instead of the hydrogen bonds formed between the side chain groups of wt Q493 with K31 and E35, R493 forms a salt bridge network with the hACE2 E35 and D38 (figure **3.10**). These interactions connect the neck and the left shoulder interface regions into one and may restrict the movement of the BA.1 RBD with respect to the hACE2 helix α1, indicating an increased binding affinity to the receptor.

Moreover, visual observation of the trajectories obtained from the simulation of BA.1 RBD-hACE2 revealed different conformations of the R493 side chain in order to interact with both E35 and D38. This dynamic may affect the interaction between the hACE2 H34 and the RBD Y453. Once the side chain of R493 rotates to interact with D38, H34 also moves backwards to maintain the hydrogen bond between the imidazole group and the Y453 hydroxyl group.

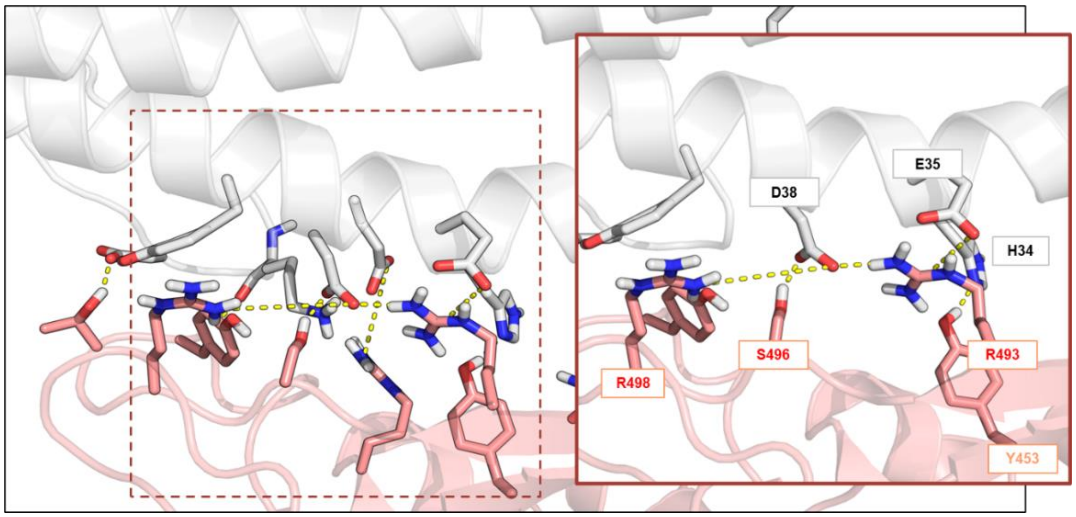


Figure 3.10 – Different conformation of the substituted R493 throughout the simulation. Snapshot from the simulation with Omicron BA.1 RBD complexed with hACE2, focused in the interactions between the substituted residues of Omicron BA.1 with hACE2 D38, represented in yellow dashed lines. The residues of RBD and hACE2 are labeled in salmon and black, respectively. The RBD substitutions are labeled in red.

The right shoulder of BA.1 RBD interface is not altered. The F486-Y83 pi-stacking and the Q24-N487-Y83 hydrogen bonds are maintained, which demonstrates that the substituted K478 may not interfere with the binding. Nevertheless, the substituted N477 can interact with the hACE2 S19 although it is not shown due to the prevalence being below the cutoff (~14%).

Omicron BA.2 RBD-hACE2 interface

The interface between Omicron BA.2 RBD and hACE2 is quite similar to the BA.1 complex, although some interactions were lost (figures **3.11A, B**). It is noteworthy that the substitutions G446S and G496S present in BA.1, do not occur in Omicron BA.2. Additionally, BA.2 presents two unique substitutions, T376A, D405N, and R408S. Only the residue 405 is located close to the interface region with hACE2, although no interaction was found between this residue and hACE2.

On the BA.2 RBD's left shoulder, the Y501-Y41 pi-stacking and the G502-K353 hydrogen bonds are maintained, however, the salt bridge between the substituted R498 and the hACE2 D38 is practically non-existing, given the selection cutoff (~11% simulation frames). In Omicron BA.1, the D38 from hACE2 formed a triple interaction network with the substituted R498, R493, and S496.

In the case of BA.2, the absence of the substituted S496 may have resulted in a competition between R493 and R498 for the hACE2 D38. Thus, the salt bridge between the substituted R493 and the D38 has become more stable and the salt bridge with R498 less prevalent. With this, in both subvariants, it can be explained the decreased or even loss of the Y449-D38 hydrogen bond.

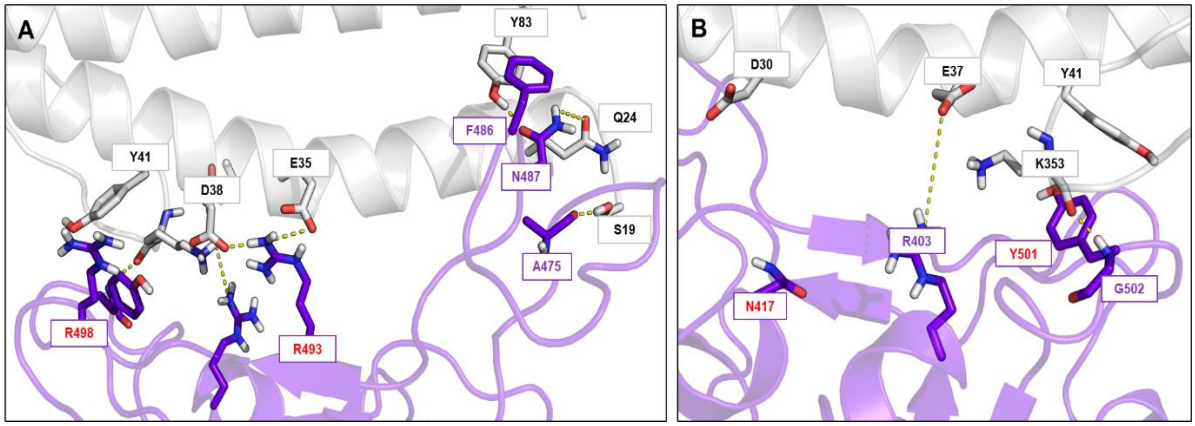


Figure 3.11 – Representation of the binding interface of Omicron BA.2 RBD (purple cartoon) with hACE2 (white cartoon). The identified pairwise interactions present in over 30% of the simulation frames are represented in yellow dashed lines and the participating residues from RBD and hACE2 are represented in sticks, colored by atom group, and labeled in purple and black, respectively. The RBD substitutions are labeled in red. **(A)** Front-view of the overall binding interface, where interactions are established by the RBD's left shoulder (residues 498, 501, and 502), neck (403 and 493) and right shoulder (475, 486 and 487). **(B)** View of the biding interface from behind.

The interactions between the BA.2 RBD neck and hACE2 are also maintained, such as the R403-E37 and the E35-R493 salt bridge. Because of the rotation of the arginine side chain to interact with D38, this last interaction could be lost. However, R493 adopts a side chain conformation that allows both interactions with E35 and D38, thus having a ~95% and ~99% prevalence, respectively.

Lastly, the interface region located on BA.2 RBD's right shoulder is similar to the one observed in wt and BA.1, where the F486-Y83 pi-staking and the Q24-N487-Y83 hydrogen bonds are kept. In addition to this a new hydrogen is formed between the backbone of RBD A475 and the side chain of hACE2 S19. This residue may also be involved in the interaction with the substituted N477, but with low prevalence (~19%).

Omicron BA.5 RBD-hACE2 interface

In contrast to the other Omicron subvariants, Omicron BA.5 does not contain the substitution Q493R, although it has two new substitutions, L452R and F486V. This last substitution is located on the RBD's right shoulder interface region (figures **3.12A, B**).

The left shoulder and the neck regions are not very different comparing with the other Omicron RBDs. The Y501-Y41 pi-stacking and the G502-K353 hydrogen bonds are still maintained, as for the T500-D355 hydrogen bond it is also a preserved interaction among wt and Omicron subvariants but in BA.1 and BA.5 showed a higher frequency (~88%) with respect to wt and BA.2. Since the residue 493 may not interact with D38 from hACE2, the substituted R498 and Y449 of RBD can establish a salt bridge and a stable hydrogen bond, respectively.

On the BA.5 RBD's neck, the R403-E37 salt bridge is kept, and the Q493 forms a hydrogen bond with E35 and with K31, similar to wt but less frequent (~35% and ~13%, respectively).

The major difference of the RBD interface region is located in the right shoulder, due to the substitution F486V. The wt F486 residue was found to be very important in the binding of the RBD to hACE2, due to the stable pi-stacking interaction formed between the aromatic rings between the side chains of F486 and hACE2 Y83. Other interactions involving F486 were shown in the literature, but they were not observed in the present work. In this study, no interaction was identified between BA.5 RBD V486 and hACE2 residues, which could lead to less stable interface due to the dynamic RBM loop. Although, the N487-Y83 hydrogen bond is maintained, the N487-Q24 hydrogen bond has become less prevalent (~17%). In addition, the interactions A475-S19-N477 are kept.

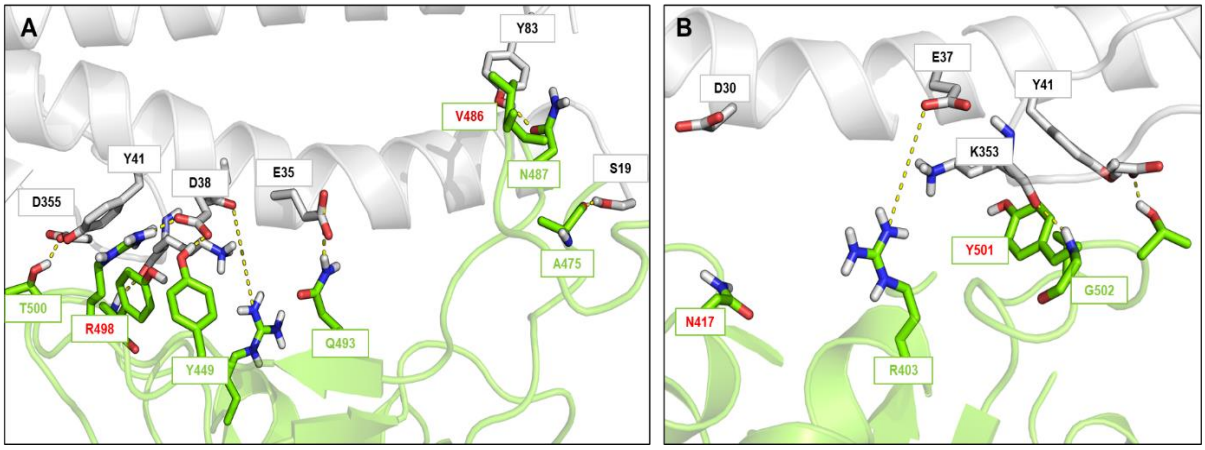


Figure 3.12 – Representation of the binding interface of Omicron BA.5 RBD (purple cartoon) with hACE2 (white cartoon). The identified pairwise interactions present in over 30% of the simulation frames are represented in yellow dashed lines and the participating residues from RBD and hACE2 are represented in sticks, colored by atom group, and labeled in purple and black, respectively. The RBD substitutions are labeled in red. **(A)** Front-view of the overall binding interface, where interactions are established by the RBD's left shoulder (residues 449, 498, 500, 501 and 502), neck (493 and 403) and right shoulder (475,486 and 487). **(B)** View of the biding interface from behind.

Notice that the substituted R452 does not interact with hACE2, which has no significant direct effect on the binding to the receptor, although it may have other impact, such as antibody escape, as mentioned above.

3.1.5 Contact surface analysis

The contact area between the RBD and hACE2 was calculated for each complex by computing the solvent-accessible surface area (SASA) of the RBD-hACE2 complex, the SASA of the two proteins alone for the last 500 ns of the simulations, using the GROMACS tool *gmx sasa*.

The sum of the areas obtained from the hACE2 and from the RBD correspond to the area of the complex plus two times the contact area between the two proteins. In order to obtain the value of the contact surface it is necessary to subtract the area of the complex from the sum of the areas of the two proteins, and afterward divide by two. In addition to the total contact area, we also calculated the hydrophobic and the hydrophilic components of this surface by specifying additional selections based on the charge.

The average values of the contact surfaces of each RBD-hACE2 complex are shown in figure **3.13** and in the supplementary table **S5**.

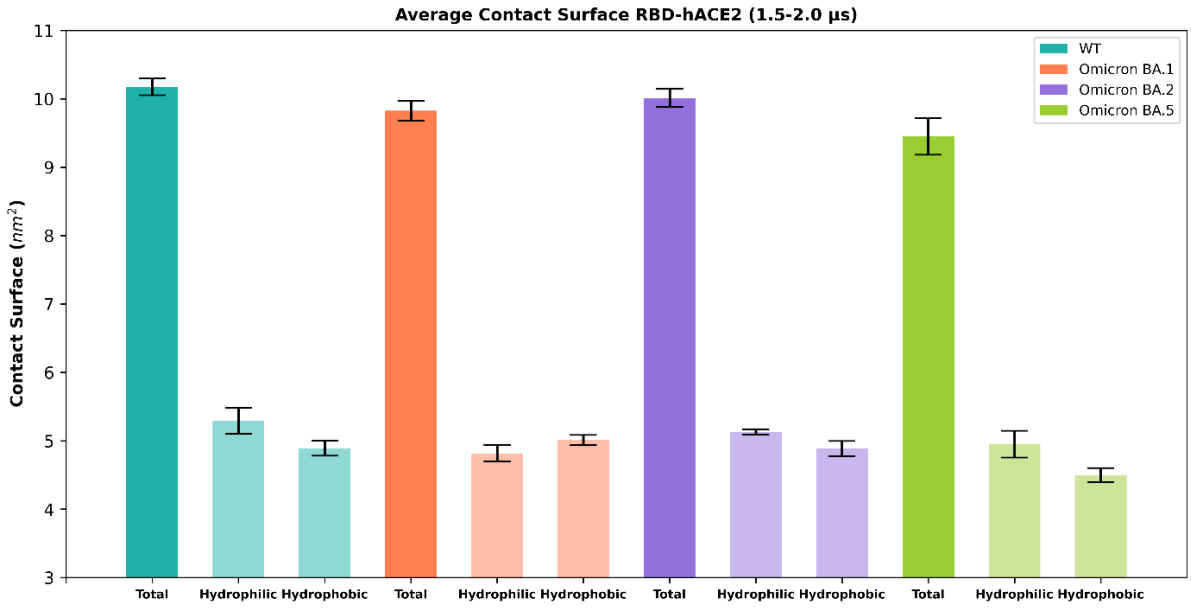


Figure 3.13 – Average values of the contact surfaces (nm^2) of each RBD-hACE2 complex for the last 500 ns: total, hydrophilic, and hydrophobic surface areas. The values correspond to the average of the averages obtained in the five simulations for each complex and errors are expressed as twice the standard deviation obtained by bootstrapping. The specific values of the contact values can be found in the supplementary table **S5**.

Based on this analysis, Omicron BA.5 RBD-hACE2 presents the lowest value of the total contact surface. Omicron BA.2 and wt RBDs have similar values of total contact surface and have a larger interface area with hACE2.

Wt RBD-hACE2 shows a higher hydrophilic contribution on the interface area, followed by Omicron BA.2 and Omicron BA.5 complexes. In contrast, the BA.1 complex shows a different pattern, in which the hydrophobic contribution appears to be higher compared to the other complexes.

Observing the wt RBD-hACE2 interface, the wt RBD not only presents more hydrophilic residues at the binding interface than Omicron RBDs but also contains more residues contributing to the interaction itself, thus demonstrating a higher contact surface with a greater hydrophilic contribution.

In Omicron BA.1, the occurring RBD substitutions – K417N, E484A, and Y505H - lead to a loss of key hydrophilic interactions with hACE2, accounting for two salt bridges and four hydrogen bonds. However, some of those interactions were replaced by three salt bridges and one stable hydrophobic contact due to Q493R, R49,8 and N501Y, respectively. Furthermore, other hydrogen bonds were formed. With this balance, it is expected for Omicron BA.1-hACE2 contact area to have an equal contribution from hydrophilic and hydrophobic interactions, as observed in the plot.

Although there are interactions that are lost compared to Omicron BA.1, the Omicron BA.2-hACE2 shows a contact surface area comparable to wt, having higher hydrophilic contribution due to the presence of more polar residues and more stable hydrophilic interactions at the binding interface.

Moreover, the substitution Q493R that occurs in BA.1 and BA.2 RBDs may have an impact on the higher values of contact surface with hACE2 due to the connection of the two main regions of the binding interface, as described above. Since this substitution doesn't occur in BA.5, this could be one of the reasons why Omicron BA.5-hACE2 presents a lower contact surface area with respect to the other Omicron subvariants. This subvariant shows the lowest value of interface area among the studied systems.

Another possible explanation of this result is the presence of the F486V substitution on the RBD right shoulder. As mentioned before, the loss of this key interaction with hACE2 residues could lead to a rearrangement of the flexible RBM loop, loss of other interactions, and a consequent decrease of the binding interface area in this region. Furthermore, the contact surface is mostly due to hydrophilic contacts, as the loss of the pi-stacking interaction F486-Y83 may be responsible for a decrease in the hydrophobic contact area.

With the results obtained from the interface analysis of the Omicron RBDs-hACE2 complexes comparing to the wt, Omicron subvariant substitutions located at the RBD's left shoulder, Q493R, G496S, Q498R, N501Y, Q498R could contribute significantly to the binding affinity with hACE2. The interactions between the Omicron residues 493, 496 498, and 501 and hACE2 appear to compensate for the loss of key interactions due to other substitutions in wt residues N417 and Y505, restoring the effectiveness for the binding, and consequently may lead to an increase in viral fitness. Additionally, S477N substitution, in contrast to T478K, may contribute to the binding by establishing and helping neighboring residues to form new hydrogen bonds with the hACE2 helix. On the other hand, F486V may have a negative impact on the hACE2 affinity.

Currently, new Omicron subvariants have emerged, being BQ.1 the most concerning one due to its increased neutralization resistance. This subvariant contains two additional substitutions, K444T and N460K, both located outside the binding interface with hACE2. Although residue 444 appears to be closer to the left shoulder of the RBD that interacts with hACE2, the substitution is not favorable given the smaller side chain of threonine compared to lysine.

Although these analyses were helpful to understand the binding mechanism of the SARS-CoV-2 Omicron RBDs to the hACE2 receptor and to predict the possible impact of the substitutions on the affinity to the receptor, the enthalpic and entropic contributions of each interaction are not known.

Thus, this study should be complemented with a method to estimate interaction free energies. In this way, it would be possible to compute the binding affinity compared with experimental results and calculate the energy contributions of the residues to the RBD-hACE2 interaction. The Molecular mechanics Poisson−Boltzmann surface area (MM-PBSA) approach has become to be one of the most widely used for these calculations and is often employed to study biomolecular complexes. This method can incorporate conformational fluctuations and entropic contributions to the binding energy, although with some limitations.[116] As future work this is going to be implemented to have insights of the impact of Omicron subvariants in the binding to hACE2 receptors.

Since hACE2 contains *N*-glycosylation sites that can interact with S protein RBD, another improvement to this work should be the consider the *N*-glycans in the simulation systems located closer to the binding interface to have a more realistic system.

Following the study of the direct effects of the occurring substitutions in Omicron subvariants on the hACE2 binding, it is important to understand the impact on the dynamic of the unbounded Omicron RBDs. This analysis can reveal other features that are related to the binding mode to the receptor and is presented below.

3.2 MD simulations of the unbounded Omicron RBDs

For the second part of this thesis, atomistic MD simulations with the unbounded RBDs of Omicron subvariants, BA.2 and BA.5, were performed in solution, obtaining in total two simulations with five replicates each. The wt RBD was previously simulated by the Protein Modeling group under the same conditions, although it was simulated for 15 μs over 5 replicates. The Omicron BA.1 subvariant was also included in the same study, although this system was simulated for only 7 μs[85]. The goal of this study is to compare the RBD conformational dynamics of the Omicron subvariants BA.2 and BA.5 to the wt RBD, giving specific emphasis to the effect of the substitutions.

3.2.1 RMSD

As mentioned earlier, the RBD can be divided into two subdomains with different dynamics, as mentioned earlier, in which the RBM is a particular irregular region that comprises the “ridge” (figure **3.14**), the region of interest.

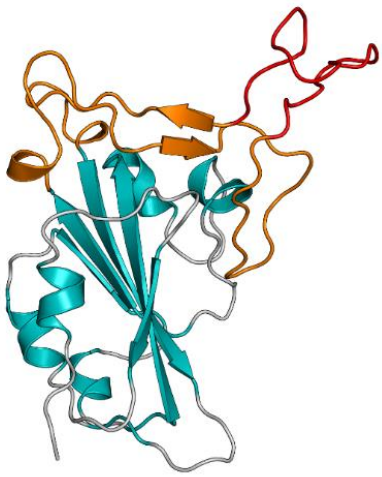


Figure 3.14 - Representation of hACE2-bound RBD (PDB ID 6M0J), highlighting the RBD core, colored cyan, RBM, colored orange, and the “ridge” region, colored red.

The RMSD was calculated for the simulated free Omicron RBDs, using MDAnalysis. In this case, the Cα atoms of the RBD core were fitted against the reference structure, the RBD X-ray structure (PDB ID 6M0J) and the RMSD was calculated for the Cα RBD, to observe the time evolution of the flexible region of RBD (figure **3.15**).

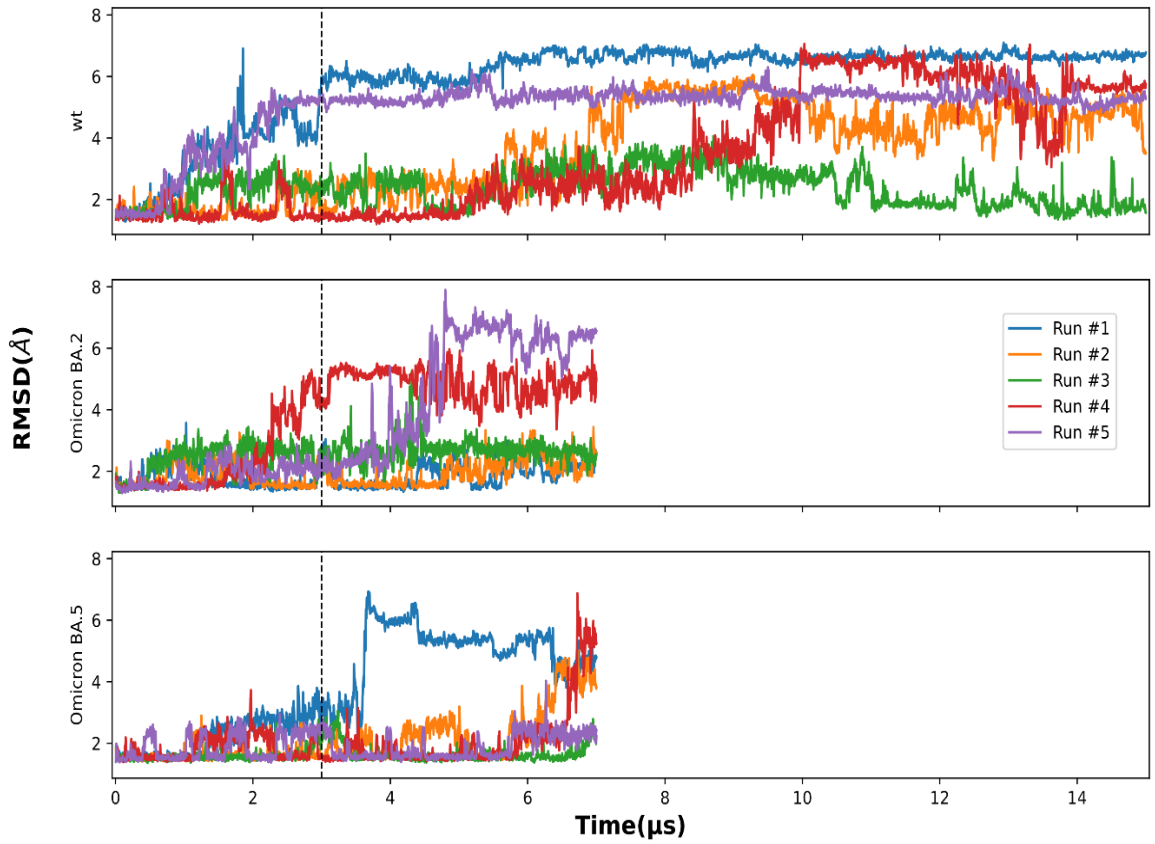


Figure 3.15 – Moving average of RBD Cα RMSD in solution. Data are shown for the five replicas for each variant tested. RBD Core Cα were fitted against the RBD X-ray structure from PDB ID: 6M0J. The moving average was calculated using the neighboring 50 frames. The first 3 μs of simulation were used for equilibration (black dashed line) and the remaining frames were used for further PCA and RIN analysis.

From the overall RMSD values of the wt replicates, it is possible to observe that there are major oscillations, between 3-7 Å, that correspond to the movement of the loop located at the RBM, comprising the “ridge” region. When RMSD values start to increase above 3-4 Å, by visualizing the trajectories it is possible to observe that this RMSD increase corresponds to the movement of the RBD towards the “closed” state.

In the case of the wt RBD, this system was already simulated twice the time of the studied systems in this study. By analyzing the RMSD plot for wt, the structure appears to stabilize only after 10 μs of simulation time for all replicates. This indicates that the RBM loop is equilibrated after this time where it explores both “open” and “closed” states, as mentioned before. It is also worth noting that in systems with this type of dynamic behavior, the equilibrium can only be inspected by looking at the global behavior of all replicates.

In the case of Omicron BA.2 RBD, the stabilization is achieved after 5 μs of simulation time for all replicates. Once again it is observed that different replicates have different behaviors and the analysis of the system in equilibrium can only be done by looking at the replicates combined. As for Omicron BA.5 RBD, only the replicate 1 (blue) stabilized after 4 μs, and the RMSD values for the remaining replicates started to oscillate after 6 μs.

In order to compare the time evolution of the systems for further analysis of the RBM conformational dynamics, it would be interesting to prolong the simulations, especially for the Omicron BA.5 system, and this will be done in the future. Nevertheless, taking in account all replicates of each system, to have a wider conformational space and reliable results, the first 3 μs were chosen as equilibration time, and the remaining frames were used for further analysis.

3.2.2 RMSF

In order to understand the magnitude of flexibility in the RBM region, the RMSF was calculated for Cα atoms for all replicates of each system using MDAnalysis, fitting the RBD Cα core.

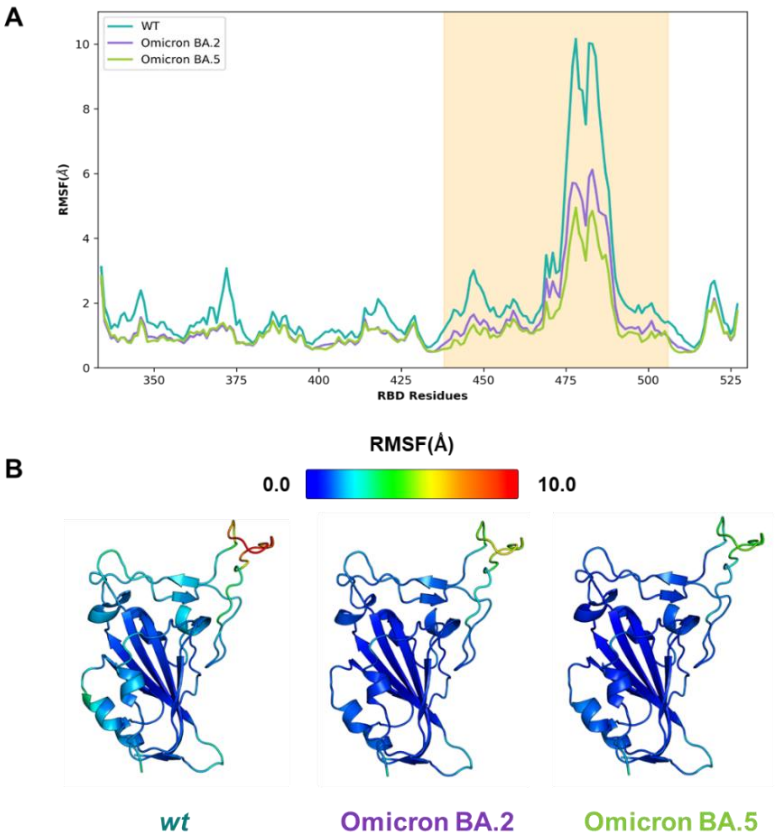


Figure 3.16 – (A) Average Cα RMSF per RBD residue computed over the last 12 μs of the simulations (equilibrated system) for all replicates of wt and over the last 4 μs of the simulations (equilibrated system) for all replicates of Omicron BA.2 and BA.5. The average over the five replicates was calculated and plotted in figure. Wt, Omicron BA.2 and BA.5 are represented in blue, orange, purple, and green, respectively. The RBM region is highlighted in light-orange. **(B)** The average Cα RMSF of RBD are represented in each RBD structure. The colors are distributed over the range of RMSF values, with blue representing the lowest (most stable) and red the highest (most flexible).

In figure **3.16**, the average RMSF per residue of the RBD over the last 4 μs of the simulations are represented, as well as the structure of the RBDs colored by average RMSF values to represent the most stable and most flexible regions.

From the RMSF plot, it is possible to observe that the fluctuation pattern differs from the RBD bound to the hACE2 protein, as shown in the last chapter. In this case, the unbounded RBD demonstrates a significant increase in the RMSF values in the RBM region, highlighted in light-orange (figure **3.16A**). This indicates a large movement of this region, especially in the zone of the “ridge” loop (residues 471-491), which is consistent with what was observed during the visualization of the trajectories and with the overall RMSD values.

Comparing the simulated RBDs with wt, it is possible to observe that not only the “ridge” region, but the overall RBM region presents more flexibility (figure **3.16B**). Looking at the Omicron RBDs, the “ridge” region in the unbounded BA.2 RBD is quite more flexible relative to the BA.5 RBD, which leads to a larger conformational dynamic of the RBM “ridge”.

3.2.3 Conformational Dynamics of the Omicron RBDs: PCA analysis

Visual observation of the trajectories obtained in the simulation of the SARS-CoV-2 Omicrons RBD in water indicated that the RBM region exhibits significant conformational alterations. To characterize the conformational dynamics of this region, PCA was performed on the coordinates recovered from RBD simulations of the previous work [85], reducing them to 2 PCs. Then, the 2D conformation space was converted into free-energy landscapes (figure **3.17 A, C,** and **E**). The information extracted from energy surface landscape analysis from 2D-PCA of RBD conformational dynamics can be found in the supplementary table **S6**, including the energy minima, percentage of frames, and loop conformation for each of the basins.

In the energy landscape of wt 2D-PCA (figure **3.17A**) it is possible to observe two deep basins which correspond to two sets of RBM conformations. Basin 1 corresponds to conformations similar to the crystallographic hACE2-bound RBD structure, or so-called “open” conformations. In contrast, basin 0 is constituted by different conformations, with the loop of the RBM twisted, covering the hACE2 binding surface - RBD “closed” conformations.

Like in wt, the 2D-PCA of Omicron BA.5 RBD showed two deep basin clusters, wherein basin 0 corresponds to the wt “open” conformations and basin 1 to the “closed” state (figure **3.17E**).

In contrast to wt and other Omicron subvariants, including the BA.1 that was already studied [85], Omicron BA.2 RBD presents a new conformational state in the 2D-PCA, in addition to the two other basins corresponding to the wt “open” (basin 0 and 1) and “closed” (basin 2) conformations. In this conformation (basin 3), the RBM folds into a state that differs from the previously describes “open” and “closed” conformations and is here designated “shrunken” (figure **3.17C**).

Afterward, the respective frames from each basin were extracted for further RIN analysis to characterize the RBD conformations and the intramolecular interactions that stabilize them, as well as understand the impact of the substitutions in the Omicron RBD’s conformation.

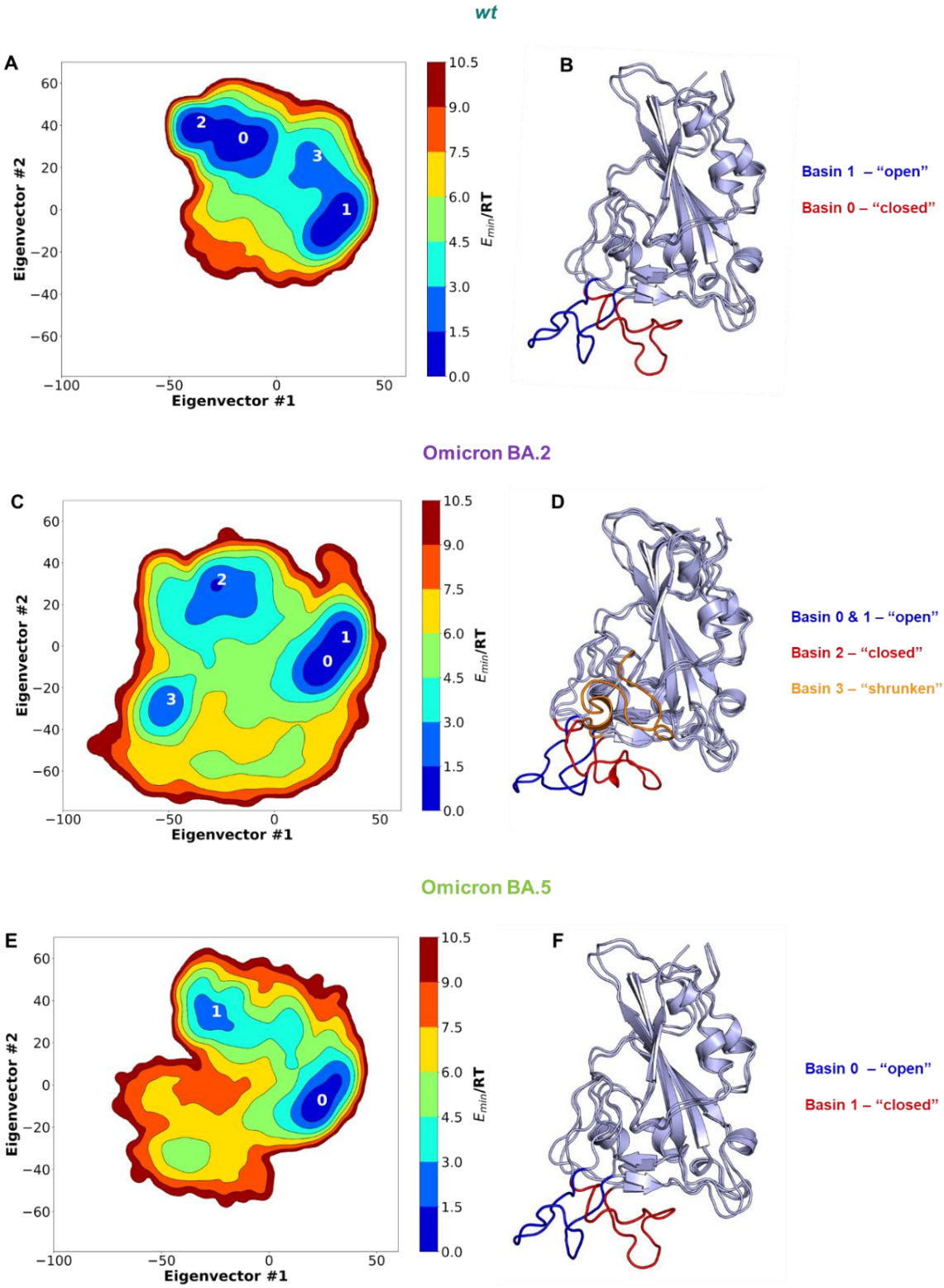


Figure 3.17 - Two-dimension principal component analysis (PCA) of SARS-CoV-2 RBD conformational dynamics in water. Representation of the first two principal components obtained from the Cα backbone of the wt RBD [85] **(A)**, Omicron BA.2 RBD **(C)** and Omicron BA.5 RBD **(E)**, wherein the basins with $k_BT < 3$ are numbered. **(B,D,F)** Representation of the lowest energy structures of each selected basin, in which the "ridge" region of the "open", "closed", and "shrunken" conformations are colored in blue, red, and orange, respectively.

Omicron BA.2 RBD presents a distinct conformation: “Shrunken”

As mentioned before, the Omicron BA.2 RBD presents a distinct conformational dynamic from the other Omicron subvariants and the other VOCs.

Like wt, MD simulations of the Omicron BA.2 RBD show prevalence of the wt “open” and “closed” conformations, shifting the “open”/”closed” equilibrium towards an “open” state relative to the wt (~44.5% occurrence[85]) and is in this state for most of the simulation time (~68,79% occurrence, Supplementary Table **S6**). This may have an impact on the RBD conformational dynamics promoting efficient binding to hACE2, similar to Omicron BA.1 (~96,6% “open” of the simulation frames) that emerged in November 2021. [85]

RIN analysis was conducted in order to identify the intramolecular interactions established in each conformation. We selected the interactions that were present in over 50% of the simulation frames (Supplementary figures **S9-S13**).

In BA.2 RBD, the interactions that stabilize the “open” conformation are common to those present in wt and Omicron BA.1 lineage, as well as in the earlier variants [85], namely the triple pi-stacking between residues Y489-F456-Y473 and the hydrogen bond between Y489 and Y473. Moreover, the Q493R mutation does not disrupt the hydrogen bond with Y453 which stabilizes the β6-strand (figure **3.18**).

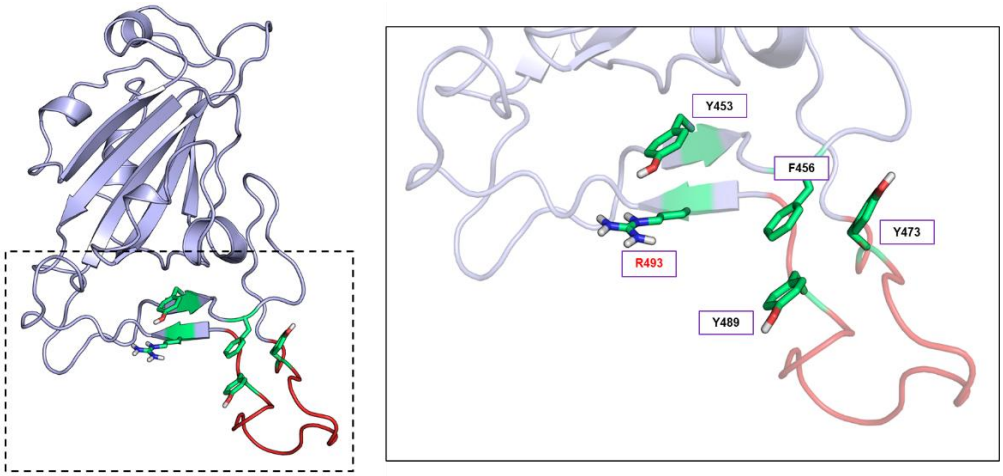


Figure 3.18 – Snapshot of a simulation of Omicron BA.2 RBD showing the intramolecular interactions that stabilize the “open” conformation. The “ridge” is colored in red and the residues of interest are represented in green sticks. The right-side panel corresponds to a zoom in on the interactions. Text labels indicate crucial residues and the substitutions are represented in red. The figure is rotated 180º relative to figure **3.17D.**

As opposed to wt RBD, the “closed” conformation is not stabilized by the salt bridges formed by K417, E484, and R403. As described ahead, the Omicron RBD presents the E484A substitution, preventing these interactions.

Similar to what is observed in the wt RBD “closed” conformation, the triple pi-stacking observed in the “open” state is lost in Omicron BA.2. Instead, in the “closed” state, new hydrogen bonds are formed between the backbone of V483 and S494, and between the backbone of the substituted A484 and N487. Additionally, the one of the cysteines located in the “ridge” region, C480 interacts with the S494 and L492 backbones. Nevertheless, R493-Y453 interaction is maintained, stabilizing the β-sheets of the RBM (figure **3.19**).

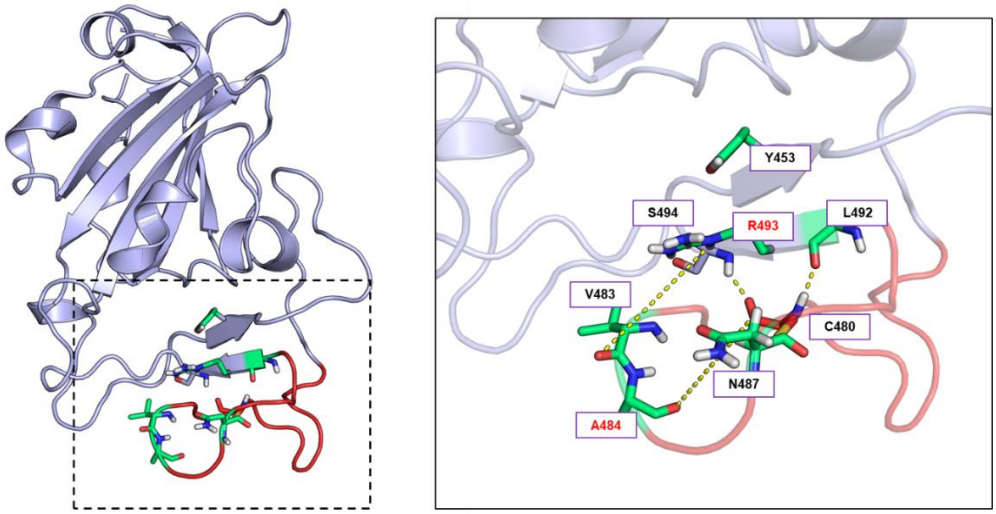


Figure 3.19 - Snapshot of a simulation of Omicron BA.2 RBD showing the intramolecular interactions that stabilize the “closed” conformation. The “ridge” is colored in red and the residues of interest are represented in green sticks. The right-side panel corresponds to a zoom in on the interactions. Text labels indicate crucial residues and the substitutions are represented in red. The figure is rotated 180º relative to figure **3.17D**.

Interestingly, in contrast to wt and other variants, BA.2 RBD exhibits an alternative RBD conformation, “shrunken” (~7.94% occurrence, Supplementary Table **S6**). The name is based on the attachment of the “ridge” with the loop located at the opposite side of RBM, recalling the “shrunken shoulders” of the RBD.

This conformation is primarily stabilized by the formation of hydrogen bonds between three asparagine residues. Instead of the hydrogen bond network N448-N450-K444-F497 that occurred in “open” and “closed” conformation, the side chain of the substituted N477 forms a hydrogen bond network with the backbone of N448 and N450. Additionally, the substituted K478 interacts with backbone I472 and may interact with the backbone of A475, although it does not appear in the RIN due to the cutoff of 50% (figure **3.20A**).

Similar to the “open” and “closed” conformations, the interaction between R493 and Y453 is maintained (figure **3.20B**). This interaction appears to be important for the stability of the three observed conformations of Omicron BA.2.

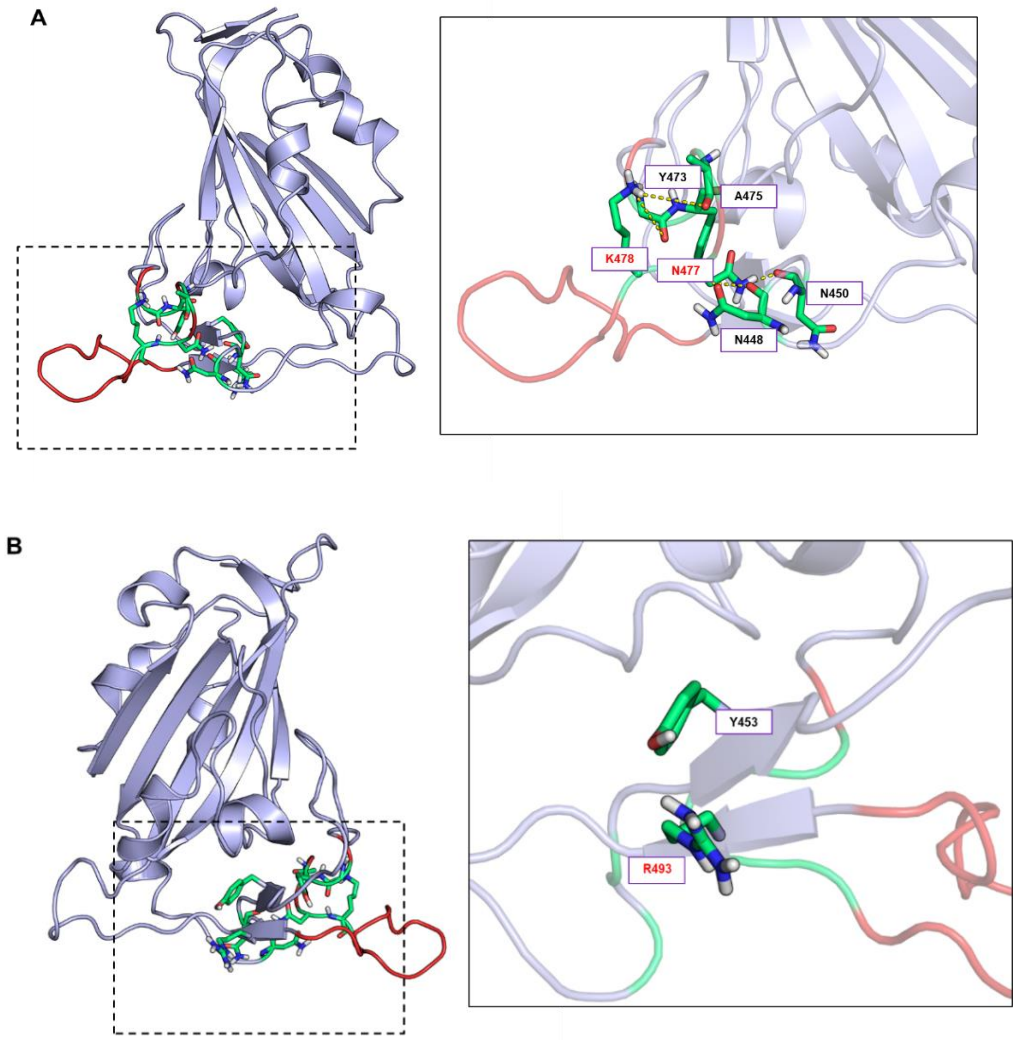


Figure 3.20 - Snapshot of a simulation of Omicron BA.2 RBD showing the intramolecular interactions that stabilize the “shrunken” conformation. The “ridge” is colored in red and the residues of interest are represented in green sticks. The right-side panel corresponds to a zoom in on the interactions. Text labels indicate crucial residues and the substitutions are represented in red. The figure **3.20B** is rotated 180º relative to figure **3.20A**.

This conformation is quite different from the ones that were already mentioned in the literature and it appears to be another alternative “open” conformation. Similar to the “reversed” conformation observed for Delta variant, the “shrunken” state seems to alter the conformation of the “ridge” which bends backwards towards the opposite RBM loop, maintaining part of the hACE2 binding surface open and accessible for binding and further infection, as demonstrated in figure **3.21**. The remaining binding region that is lost with the “ridge” conformational change is hypothesized to be compensated by increased capacity for antibody escape, providing a fitness advantage. Due to the several RBD-targeting antibodies that recognize the “ridge” region, the class II nAbs, this alternative conformation of the “ridge” may be not as easily recognized.

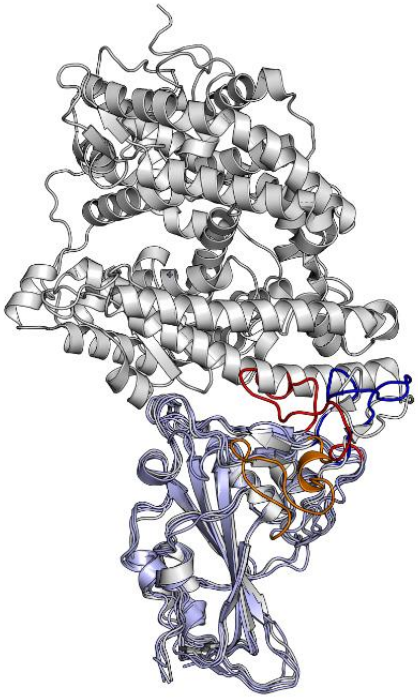


Figure 3.21 – Representation of Omicron BA.2 RBD conformational states in complex with hACE2. The RBD from X-ray structure (PDB ID 6M0J) was superimposed to the snapshots of the lowest energy structures of the basins, in which the “ridge” regions of the “open”, “closed”, and “shrunken” snapshots are colored in blue, red, and orange, respectively. The X-ray structure is represented in gray cartoon.

Omicron BA.5 RBD improves RBM accessibility compared to BA.2

As mentioned before, Omicron BA.5 RBD showed prevalence of two conformations that correspond to the wt “open” and “closed” states. Similar to BA.2, for BA.5 RBD the open conformation accounts for the majority of configurations sampled during the simulations, and the frequency of the “open” state is even larger than in Omicron BA.2 (~77.8% occurrence, Supplementary Table **S6**), which can improve hACE2 accessibility compared to BA.2.

For Omicron BA.5, the “open” conformation is stabilized by the same interactions observed in wt and the other VOCs [85]. The triple interaction network Y489-F456-Y473 and the two hydrogen bonds between the Y453 and the Q493 are maintained (figure **3.22**).

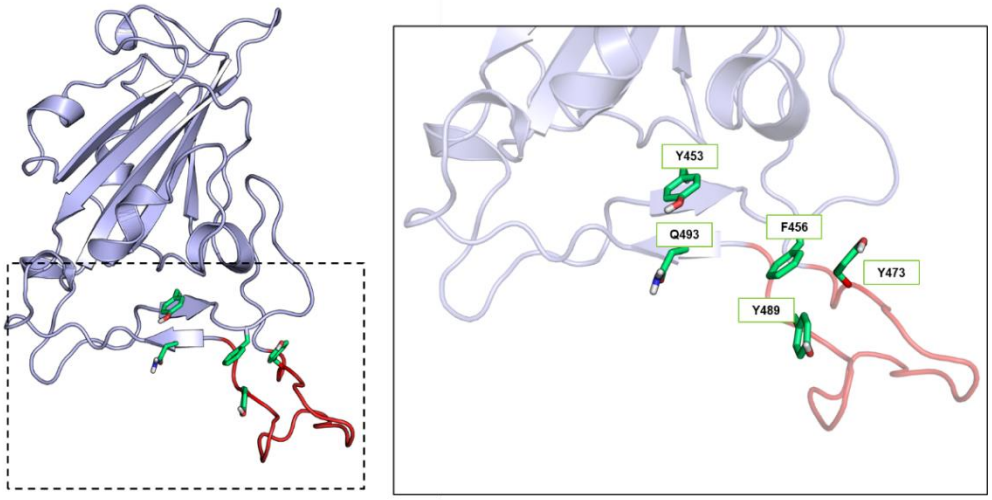


Figure 3.22 - Snapshot of a simulation of Omicron BA.5 RBD showing the intramolecular interactions that stabilize the “open” conformation. The “ridge” is colored in red and the residues of interest are represented in green sticks. The right-side panel corresponds to a zoom in on the interactions. Text labels indicate crucial residues and the substitutions are represented in red. The figure is rotated 180º relative to figure **3.17F**.

Due to the common substitutions that occur in Omicron subvariants, such as K417N and E484A, the salt bridges between these residues and R403 do not appear to stabilize the “close” conformation. Instead, Q493 interacts with the backbone of C480, and the backbone of C488 forms another interaction with F490 backbone, although it is not shown in RIN (cutoff 50 %). The hydrogen bond formed between the backbone N448 and the substituted A484 is replaced by a transient interaction with N481 (figure **3.23**).

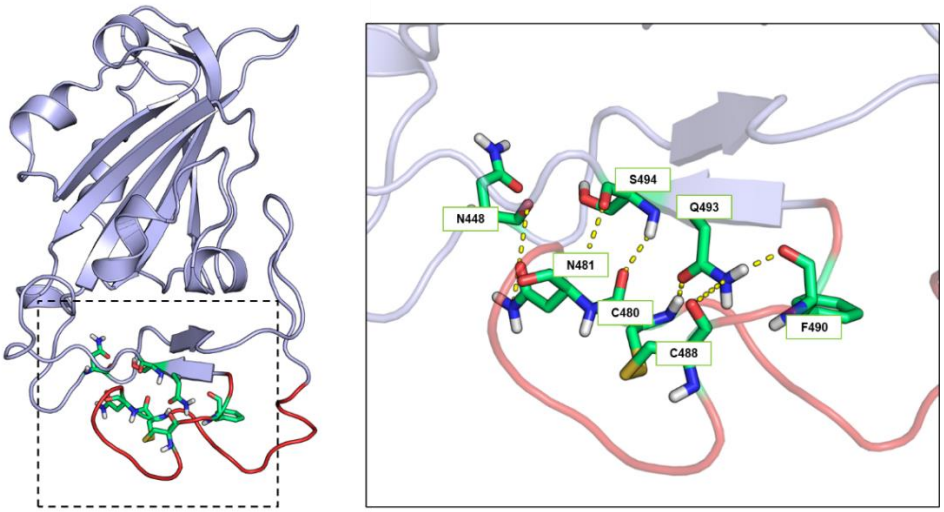


Figure 3.23 - Snapshot of a simulation of Omicron BA.5 RBD showing the intramolecular interactions that stabilize the “closed” conformation. The “ridge” is colored in red and the residues of interest are represented in green sticks. The right-side panel corresponds to a zoom in on the interactions. Text labels indicate crucial residues and the substitutions are represented in red. The figure is rotated 180º relative to figure **3.17F.**

Moreover, the presence of both Delta RBD substitutions, T478K and L452R does not have an impact on the conformational dynamics towards a “reversed” state, as we would expect. Nevertheless, given the RMSD values shown before the system did not reach stabilization within the 7 μs of simulation, therefore, more simulation time would be required to clarify this.

3.2.4 RBD conformational dynamics impacts its secondary structure

The conformational dynamics of the “ridge” region might have an impact on the secondary structure of the RBM region. To measure the alterations in the RBM, a secondary structure analysis was performed using GROMACS *do_dssp* tool to obtain the percentage of each type of secondary structure for all conformations of the Omicron subvariants, BA.2 and BA.5. The difference between the “closed” and the “open” conformations was calculated for both Omicron RBDs. Since Omicron BA.2 RBD presents an alternative conformation to “open” state, the difference between these two conformations was also calculated (Figure **3.24**).

From the obtained results, in Omicron BA.2, the “closed” conformation does not seem to significantly impact the secondary structure of the RBD. The major impact appears to occur in residues 473-475 and 485-490, which corresponds to the loss of the “ridge” beta-sheet structure during the closing. This effect is possible to observe for Omicron BA.5 as well and can be explained with the lack of the triple pi-stacking that was likely stabilizing this region in the “open” conformation. Interestingly, both Omicron subvariants seem to form a small alpha-helix in the “ridge” region (residues 473-477) during closing. Additionally, Omicron BA.2 shows some helicity in residues 484-488 which may correspond to the interactions between these residues with the residue 493 and 494 upon closing.

Regarding the secondary structure of the Omicron BA.2 “shrunken” conformation, there are substantial differences comparing with the “open” state. Similar to the “closed” state, the two beta-sheets of the “ridge” region are lost. Curiously, there is a significant alpha helical character between residues 445-455 that are formed probably due to the interactions with the “ridge” region to stabilize this conformation.

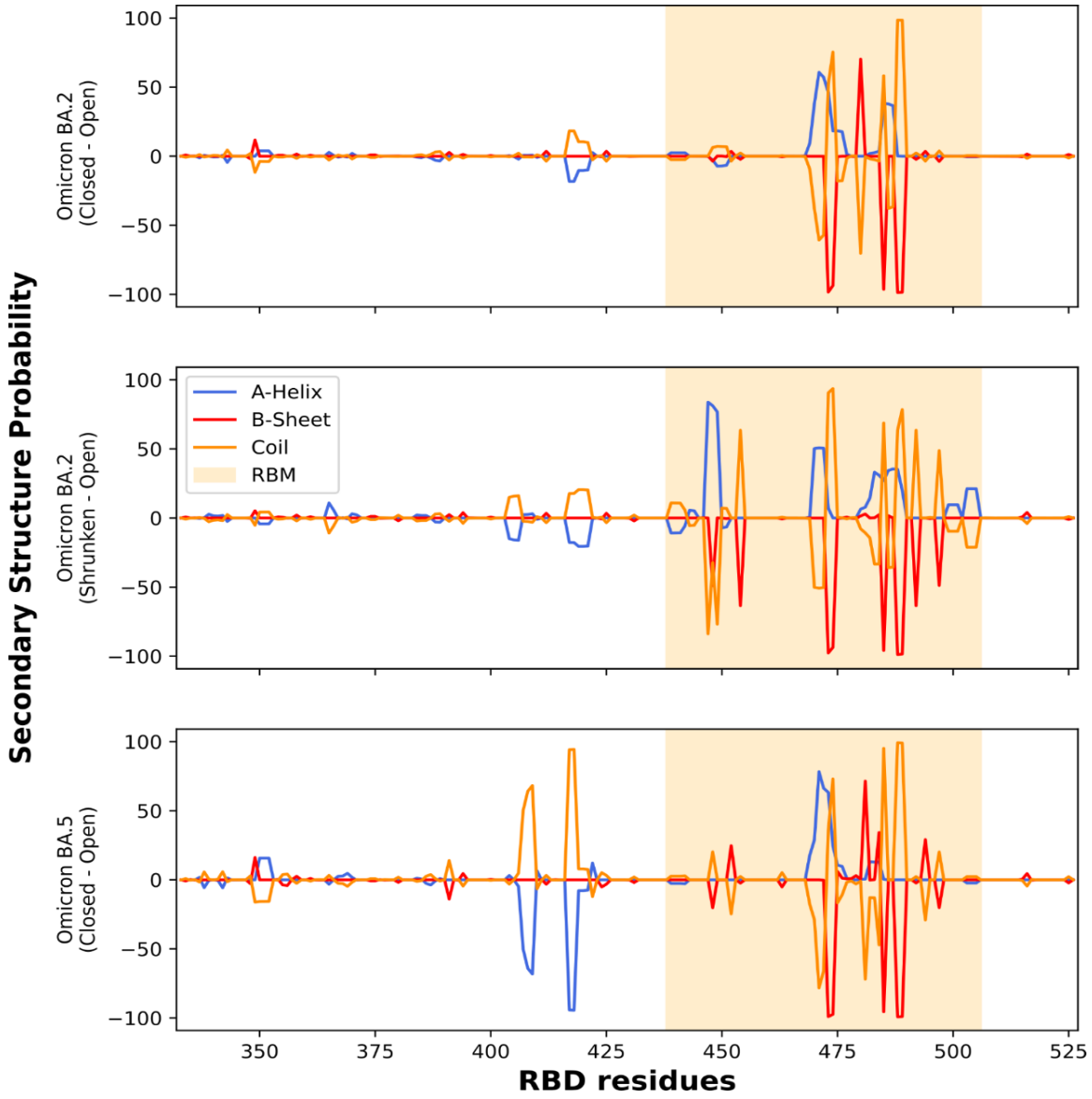


Figure 3.24 – Secondary structure difference between "closed" and "open" conformations of Omicron RBDs, and "closed" and "shrunken" conformations for Omicron BA.2. Probability of coil, alpha-helix and beta-sheet secondary structures was obtained using the GROMACS tool *do_dssp* for all conformations of the Omicron subvariants. RBM region is highlighted in the graphics in light-orange.

According to the study previously done on the impact of the variants in RBD conformational dynamics [85], the substitutions that were mostly associated with the dynamics of the "ridge" region are common to all Omicron subvariants, such as K417N, E484A, Q493R, and N501Y. Thus, the RBD dynamics of the Omicron sublineages was expected to be similar to BA.1. However, the results from PCA and RIN analysis of the intramolecular interactions in Omicron BA.2 and BA.5 RBDs demonstrated that there are differences, especially due to the new conformation found for BA.2 RBD, the "shrunken" conformation, which reveals a fitness advantage due to its features.

Moreover, the “open” conformation for Omicron RBDs is similar among all the variants, whereas the “closed” conformation is stabilized by different intramolecular interactions. Nevertheless, the wt Q493 in Omicron BA.5 and the substituted R493 present in Omicron BA.2 RBD have shown to stabilize the “closed” conformation. Additionally, the larger movements of the “ridge” region have a significant impact on the secondary structure of the RBM region.

Regarding the percentage of “open” conformations sampled during the simulations of the Omicron variant, for BA.1 subvariant it is substantially higher than for the wt, as described in the literature [85], followed by BA.5 and BA.2. Given that, similar to Omicron BA.1, Omicron BA.5 subvariant may have a significant impact on the RBD conformational dynamics promoting efficient binding to hACE2 receptor. Although BA.2 presents a lower percentage of “open” conformation compared to the other Omicron subvariants, the viral fitness may be reestablished by the “shrunken” conformation by maintaining the accessibility to hACE2 and increasing antibody escape.

Finally, the RBM “ridge” dynamics can have an impact on the hACE2 binding surface for receptor engagement and further infection. With this study, it was possible to complement the knowledge obtained from the RBD-hACE2 complexes and the effect of Omicron subvariants BA.2 and BA.5 on the binding to hACE2.

3.3 Omicron subvariants impact the binding affinity to hACE2

The overall results from the RBD-hACE2 interface analysis of the simulated systems and the impact of the RBD substitutions on the conformational dynamics can be compared to experimental binding assays from recent studies. [68,117,118] These results were all obtained by surface plasmon resonance (SPR) and encompass the binding kinetics data for wt and Omicron subvariants S protein or S RBD with hACE2 (Table **3.1**).

Table 3.1 – Binding affinities and kinetics of Omicron subvariants spike proteins to hACE2 measured by SPR from recent studies. [68,117,118]

S protein variant/RBD-hACE2	Binding kinetics data (SPR)	Wt	Omicron BA.1	Omicron BA.2	Omicron BA.5
Spike (Wang Q. *et al*) 68	K_D (nM)	5.20	2.10	2.21	1.66
	k_{on} (M^{-1} s^{-1})	7.45×10^4	1.73×10^5	1.29×10^5	9.52×10^4
	k_{off}(s^{-1})	3.87×10^{-4}	3.65×10^{-4}	2.85×10^{-4}	1.58×10^{-4}
RBD (Tuekprakhon A. *et al*) 118	K_D (nM)	7.3	7.8	4.0	2.4
	k_{on} (M^{-1} s^{-1})	7.00×10^5	8.20×10^5	1.10×10^6	8.10×10^5
	k_{off} (s^{-1})	5.10×10^{-3}	6.40×10^{-3}	4.40×10^{-3}	1.90×10^{-3}
RBD (Cao Y. et al.) 117	K_D (nM)	21.90	14.50	10.80	14.40
	k_{on} (M^{-1} s^{-1})	3.59×10^5	5.81×10^5	7.37×10^6	6.48×10^5
	k_{off}(s^{-1})	7.88×10^{-3}	8.45×10^{-3}	7.92×10^{-2}	9.32×10^{-3}

Since the experimental data was extracted under different conditions, the values from binding affinity of the RBDs to hACE2 are not fully consistent with each other. Nevertheless, regarding the equilibrium dissociation constant (K_D) values in both Wang *et al.* [68] and Cao Y. *et al.* [118] studies, it is possible to observe that there is a pattern. These studies show that all Omicron subvariants have an increased binding affinity to the wt, although in the Tuekprakhon A. *et al* study, BA.1 presents a similar binding affinity as wt. Additionally, BA.5 is shown to be the Omicron subvariant with higher affinity to hACE2, except in Cao Y. *et al*. [117]

Overall, these results can be correlated to the ones obtained from the analysis of the MD simulations performed with the Omicron RBD-hACE2 complexes and the unbounded RBDs.

From the overall analysis of the binding interface between Omicron RBDs and hACE2, the interactions between the substitutions and hACE2 may increase the binding affinity for the receptor. However, due to the reduced contact surface for BA.5 and the F486V substitution that may have a negative impact in the binding, the value of binding affinity was not expected. Nevertheless, this can be compensated by other interactions stronger interactions to reach a better binding, and there may also be compensated by entropic factors.

Regarding the RBD conformational dynamics, the Omicron RBDs stayed in the “open” conformation for most of the simulation time due to the presence of substitutions such as K417N and E484A that prevent key interactions that stabilize the “closed” conformation, which correlates to the higher values of binding affinity compared to the wt.

Once more, since BA.1 presents a larger percentage of “open” conformation it would be expected to have a better binding than BA.5. However, the effect of the BA.2 “shrunken” conformation is reflected on the K_D value compared to the other subvariants. The movement of the RBM “ridge” towards the opposite side causes the loss of interactions at an important region of the RBD binding interface with hACE2. Therefore, it is predicted a negative impact on the binding.

In Cao Y. *et al*. study, there is a different tendency, in which BA.2 RBD presents a better binding compared to the other Omicron subvariants.[117] This could be due to the different experimental conditions, including the immobilization of the RBD that may affect the “open”/”closed” equilibrium. If RBD is presented in its “open” conformation, due to the substitutions located at the binding interface, it is possible to bind more efficiently to the hACE2.

The association and dissociation rates were also analyzed, in which k_{on} reflects the efficiency of protein-protein collisions that lead to a bound state, and k_{off} is considered the lifetime of the protein-protein complex, i.e, the strength of the interaction.

There is an increase on k_{on} values for all Omicron subvariants, reflecting their higher accessibility to hACE2. This can be explained by the significant conformational changes in the RBM region described for all Omicron subvariants, where some substitutions lead to a prevalence of the “open” state and consequent binding. Out of the three studies, Omicron BA.2 presents higher RBD accessibility to hACE2, followed by Omicron BA.1 and Omicron BA.5. However, in Cao Y. *et al*. study, Omicron BA.2 and Omicron BA.5 RBD show higher accessibility for the receptor, which once more can be explained by the different experimental setups.

Regarding the k_{off} values, Omicron BA.1 and BA.2 present higher values compared to wt, whereas BA.5 subvariant has a lower value. These overall results hint that Omicron BA.5 complex is substantially more stable than the Omicron BA.1 and BA.2 complexes having stronger interactions with hACE2 compared to the other subvariants. This correlates with the K_D value for BA.5 presented a better binding. However, in Cao *et al*. study [117], the wt presented a lower k_{off} value, which demonstrates a more stable complex compared to the Omicron subvariants. The K_D value for wt, in this study, is explained by the lower value of k_{on}, i.e, the lower accessibility to hACE2.

It is noteworthy that these studies were performed with the complete Omicron S proteins or in presence of glycosylated hACE2 which are more complex systems than the RBD itself or complexed with hACE2 simulated in this thesis. Nevertheless, the binding affinities obtained experimentally can correlate to the formulated hypothesis of the direct and indirect impact of the Omicron subvariants on the binding to the host receptor, as all of them present a stronger binding to hACE2 compared to wt.

4. Conclusion and Future Perspectives

The aim of this thesis was the investigation of the impact of the Omicron subvariants of concern, BA.1, BA.2, and BA.5, on the binding to the hACE2 cellular receptor. To this end, for the first part of this work, long atomistic MD simulations were performed with the Omicron RBDs bound to hACE2 to assess the stability of the binding interface and compare the binding properties between Omicron subvariants and with wt.

With the results obtained from the simulations with RBDs-hACE2 complexes it is possible to conclude that Omicron subvariants share common substitutions in the RBD interface region that may have a favorable impact on the binding to hACE2 – S477N, Q493R, G496S, Q498R, and N501Y. Due to the other Omicron substitutions, K417N and Y505H, crucial interactions with hACE2 are lost. Nevertheless, substitutions in residues 477, 493, 496, 408, and 501, seem to have a compensatory effect by restoring the binding effectiveness, which consequently could give Omicron subvariants a fitness advantage compared to wt.

Among Omicron subvariants, by comparing the results of the RBD-hACE2 contact interface, BA.1 and BA.2 demonstrated larger contact areas with hACE2, although the contributions from hydrophilic and hydrophobic interactions are different in these complexes. In the case of BA.2 the hydrophilic contribution to the contact surface is higher, whereas the BA.1 contact surface presents equal hydrophilic and hydrophobic contributions. Additionally, the substitution Q493R that occurs in BA.1 and BA.2 RBDs may have an impact on the higher values of contact surface with hACE2 due to the connection of the two main regions of the binding interface.

In contrast to the other Omicron subvariants, BA.5 RBD-hACE2 presents the lowest value of the total contact surface. BA.5 RBD binding region presents a “reversion” to wt at residue 493 (R493Q) and the F486V substitution which may lead to a decrease in the contact area with hACE2 due to the smaller side chain of Q493 and loss of a stable pi-stacking interaction. Therefore, F486V may have a negative impact on the hACE2 affinity.

Moreover, the wt RBD, in equilibrium, exhibits “open” and “closed” conformations, in which either the RBD is free to engage with hACE2 or the RBM “ridge” blocks the hACE2 binding surface, leading to possible inefficient binding, respectively.

Concerning the effect of Omicron substitutions on the RBD conformational dynamics, a prior investigation on the effect of variations on this "open"/"closed" dynamics revealed that Omicron BA.1 presented a higher percentage of “open” conformations sampled during the simulations. Additionally, all Omicron subvariants present the substitutions K417N, E484A, Q493R, and N501Y that are associated with the RBM “ridge” dynamics, thus being possible to infer about the dynamics of BA.2 and BA.5 RBDs.

However, to understand the impact of the BA.2 and BA.5 subvariants, MD simulations of the unbound RBDs were performed. The analysis of these simulations demonstrated that there are differences, especially due to the new conformation found for BA.2 RBD, the new alternative “shrunken” conformation. This conformation is quite different from the “open”,

“closed” and the Delta “reversed” conformation, mentioned in the literature. The “shrunken” state seems to alter the conformation of the “ridge” which bends backwards towards the opposite RBM loop, maintaining part of the hACE2 binding surface open and accessible for binding and further infection. The loss of the remaining binding region is hypothesized to be compensated by increased capacity for antibody escape by class II-epitope occluding, providing a fitness advantage.

Through RIN analysis, it was also possible to characterize the intramolecular interactions that stabilize the different Omicron conformations. For the “open” conformation is similar among all the variants, whereas the “closed” conformation is stabilized by different interactions.

Regarding the percentage of “open” conformations sampled during the simulations of the Omicron variant, similarly to the Omicron BA.1, the Omicron BA.5 subvariant may have a significant impact on the RBD conformational dynamics promoting efficient binding to hACE2 receptor. Although, BA.2 presents a lower percentage of “open” conformation compared to the other subvariants, the viral fitness may be reestablished by the “shrunken” conformation by maintaining the accessibility to hACE2 and increasing antibody escape.

Finally, the results obtained from MD simulations with the free Omicron RBDs and bound to hACE2 were compared with recent experimental binding assays to validate formulated hypothesis about the binding behavior of Omicron subvariants. Overall binding affinities and kinetic values demonstrated an increased binding affinity for all Omicron subvariants, which is possible to relate to our findings.

This thesis has limitations by considering a reductionist approach in simulating these systems. The RBD was simulated without considering the rest of the S protein, and the RBD and hACE2 glycans were not considered. To overcome this, there is an ongoing investigation to understand the impact of the *N*-glycosylation sites not only at the binding interface but also in the RBD conformational dynamics.

Regarding the impact of Omicron substitutions on the direct binding to hACE2, to better understand the binding mechanism, the MM-PBSA method is going to be implemented to estimate interaction free energies. Thus, it would be possible to compute the binding affinity comparing with the experimental data available, and calculate the energy contributions of the residues in the RBD-hACE2 interface.

.

Bibliography

(1) Lan, J.; Ge, J.; Yu, J.; Shan, S.; Zhou, H.; Fan, S.; Zhang, Q.; Shi, X.; Wang, Q.; Zhang, L.; Wang, X. Structure of the SARS-CoV-2 Spike Receptor-Binding Domain Bound to the ACE2 Receptor. *Nature* **2020**, *581* (7807), 216. https://doi.org/10.1038/s41586-020-2180-5.

(2) Wang, C.; Horby, P. W.; Hayden, F. G.; Gao, G. F. A Novel Coronavirus Outbreak of Global Health Concern. *The Lancet* **2020**, *395* (10223), 470–473. https://doi.org/10.1016/S0140-6736(20)30185-9.

(3) Zhou, P.; Yang, X. lou; Wang, X. G.; Hu, B.; Zhang, L.; Zhang, W.; Si, H. R.; Zhu, Y.; Li, B.; Huang, C. L.; Chen, H. D.; Chen, J.; Luo, Y.; Guo, H.; Jiang, R. di; Liu, M. Q.; Chen, Y.; Shen, X. R.; Wang, X.; Zheng, X. S.; Zhao, K.; Chen, Q. J.; Deng, F.; Liu, L. L.; Yan, B.; Zhan, F. X.; Wang, Y. Y.; Xiao, G. F.; Shi, Z. L. A Pneumonia Outbreak Associated with a New Coronavirus of Probable Bat Origin. *Nature 2020 579:7798* **2020**, *579* (7798), 270–273. https://doi.org/10.1038/s41586-020-2012-7.

(4) Guan, W.; Ni, Z.; Hu, Y.; Liang, W.; Ou, C.; He, J.; Liu, L.; Shan, H.; Lei, C.; Hui, D. S. C.; Du, B.; Li, L.; Zeng, G.; Yuen, K.-Y.; Chen, R.; Tang, C.; Wang, T.; Chen, P.; Xiang, J.; Li, S.; Wang, J.; Liang, Z.; Peng, Y.; Wei, L.; Liu, Y.; Hu, Y.; Peng, P.; Wang, J.; Liu, J.; Chen, Z.; Li, G.; Zheng, Z.; Qiu, S.; Luo, J.; Ye, C.; Zhu, S.; Zhong, N. Clinical Characteristics of Coronavirus Disease 2019 in China. *New England Journal of Medicine* **2020**, *382* (18), 1708–1720. https://doi.org/10.1056/NEJMoa2002032.

(5) *WHO Coronavirus (COVID-19) Dashboard | WHO Coronavirus (COVID-19) Dashboard With Vaccination Data*. https://covid19.who.int/ (accessed 2022-10-16).

(6) Fontanet, A.; Autran, B.; Lina, B.; Kieny, M. P.; Karim, S. S. A.; Sridhar, D. SARS-CoV-2 Variants and Ending the COVID-19 Pandemic. *Lancet* **2021**, *397* (10278), 952–954. https://doi.org/10.1016/S0140-6736(21)00370-6.

(7) Edwards, A. M.; Baric, R. S.; Saphire, E. O.; Ulmer, J. B. Stopping Pandemics before They Start: Lessons Learned from SARS-CoV-2. **2022**, *375* (6585), 1133–1139. https://doi.org/10.1126/science.abn1900.

(8) Mathieu, E.; Ritchie, H.; Rodés-Guirao, L.; Appel, C.; Giattino, C.; Hasell, J.; Macdonald, B.; Dattani, S.; Beltekian, D.; Ortiz-Ospina, E.; Roser, M. *Coronavirus Pandemic (COVID-19)*. Our World in Data. https://ourworldindata.org/coronavirus (accessed 2022-10-17).

(9) Forchette, L.; Sebastian, W.; Liu, T. A Comprehensive Review of COVID-19 Virology, Vaccines, Variants, and Therapeutics. *Current Medical Science 2021 41:6* **2021**, *41* (6), 1037–1051. https://doi.org/10.1007/S11596-021-2395-1.

(10) Wang, M. Y.; Zhao, R.; Gao, L. J.; Gao, X. F.; Wang, D. P.; Cao, J. M. SARS-CoV-2: Structure, Biology, and Structure-Based Therapeutics Development. *Frontiers in*

Cellular and Infection Microbiology. Frontiers Media S.A. November 25, 2020, p 724. https://doi.org/10.3389/fcimb.2020.587269.

(11) Naqvi, A. A. T.; Fatima, K.; Mohammad, T.; Fatima, U.; Singh, I. K.; Singh, A.; Atif, S. M.; Hariprasad, G.; Hasan, G. M.; Hassan, M. I. Insights into SARS-CoV-2 Genome, Structure, Evolution, Pathogenesis and Therapies: Structural Genomics Approach. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* **2020**, *1866* (10), 165878. https://doi.org/10.1016/j.bbadis.2020.165878.

(12) Alanagreh, L.; Alzoughool, F.; Atoum, M. The Human Coronavirus Disease COVID-19: Its Origin, Characteristics, and Insights into Potential Drugs and Its Mechanisms. *Pathogens 2020, Vol. 9, Page 331* **2020**, *9* (5), 331. https://doi.org/https://doi.org/10.3390/pathogens9050331.

(13) Chan, J. F. W.; Lau, S. K. P.; To, K. K. W.; Cheng, V. C. C.; Woo, P. C. Y.; Yue, K. Y. Middle East Respiratory Syndrome Coronavirus: Another Zoonotic Betacoronavirus Causing SARS-like Disease. *Clin Microbiol Rev* **2015**, *28* (2), 465–522. https://doi.org/10.1128/CMR.00102-14.

(14) Guan, W.; Ni, Z.; Hu, Y.; Liang, W.; Ou, C.; He, J.; Liu, L.; Shan, H.; Lei, C.; Hui, D. S. C.; Du, B.; Li, L.; Zeng, G.; Yuen, K.-Y.; Chen, R.; Tang, C.; Wang, T.; Chen, P.; Xiang, J.; Li, S.; Wang, J.; Liang, Z.; Peng, Y.; Wei, L.; Liu, Y.; Hu, Y.; Peng, P.; Wang, J.; Liu, J.; Chen, Z.; Li, G.; Zheng, Z.; Qiu, S.; Luo, J.; Ye, C.; Zhu, S.; Zhong, N. Clinical Characteristics of Coronavirus Disease 2019 in China. *New England Journal of Medicine* **2020**, *382* (18), 1708–1720. https://doi.org/10.1056/NEJMoa2002032.

(15) Lamers, M. M.; Haagmans, B. L. SARS-CoV-2 Pathogenesis. *Nature Reviews Microbiology 2022 20:5* **2022**, *20* (5), 270–284. https://doi.org/10.1038/s41579-022-00713-0.

(16) Lu, R.; Zhao, X.; Li, J.; Niu, P.; Yang, B.; Wu, H.; Wang, W.; Song, H.; Huang, B.; Zhu, N.; Bi, Y.; Ma, X.; Zhan, F.; Wang, L.; Hu, T.; Zhou, H.; Hu, Z.; Zhou, W.; Zhao, L.; Chen, J.; Meng, Y.; Wang, J.; Lin, Y.; Yuan, J.; Xie, Z.; Ma, J.; Liu, W. J.; Wang, D.; Xu, W.; Holmes, E. C.; Gao, G. F.; Wu, G.; Chen, W.; Shi, W.; Tan, W. Genomic Characterisation and Epidemiology of 2019 Novel Coronavirus: Implications for Virus Origins and Receptor Binding. *The Lancet* **2020**, *395* (10224), 565–574. https://doi.org/10.1016/S0140-6736(20)30251-8.

(17) Cui, J.; Li, F.; Shi, Z. L. Origin and Evolution of Pathogenic Coronaviruses. *Nature Reviews Microbiology 2018 17:3* **2018**, *17* (3), 181–192. https://doi.org/10.1038/s41579-018-0118-9.

(18) Zhu, Z.; Lian, X.; Su, X.; Wu, W.; Marraro, G. A.; Zeng, Y. From SARS and MERS to COVID-19: A Brief Summary and Comparison of Severe Acute Respiratory Infections Caused by Three Highly Pathogenic Human Coronaviruses. *Respir Res* **2020**, *21* (1), 1–14. https://doi.org/https://doi.org/10.1186/s12931-020-01479-w.

(19) Hu, T.; Liu, Y.; Zhao, M.; Zhuang, Q.; Xu, L.; He, Q. A Comparison of COVID-19, SARS and MERS. *PeerJ* **2020**, *8*. https://doi.org/https://doi.org/10.7717/peerj.9725.

(20) Du, L.; Yang, Y.; Zhou, Y.; Lu, L.; Li, F.; Jiang, S. MERS-CoV Spike Protein: A Key Target for Antivirals. *Expert Opin Ther Targets* **2017**, *21* (2), 131–143. https://doi.org/10.1080/14728222.2017.1271415.

(21) Walls, A. C.; Park, Y. J.; Tortorici, M. A.; Wall, A.; McGuire, A. T.; Veesler, D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **2020**, *181* (2), 281–292. https://doi.org/10.1016/j.cell.2020.02.058.

(22) Hoffmann, M.; Kleine-Weber, H.; Schroeder, S.; Krüger, N.; Herrler, T.; Erichsen, S.; Schiergens, T. S.; Herrler, G.; Wu, N. H.; Nitsche, A.; Müller, M. A.; Drosten, C.; Pöhlmann, S. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* **2020**, *181* (2), 271–280. https://doi.org/10.1016/j.cell.2020.02.052.

(23) Jackson, C. B.; Farzan, M.; Chen, B.; Choe, H. Mechanisms of SARS-CoV-2 Entry into Cells. *Nature Reviews Molecular Cell Biology 2021 23:1* **2021**, *23* (1), 3–20. https://doi.org/10.1038/s41580-021-00418-x.

(24) Ou, T.; Mou, H.; Zhang, L.; Ojha, A.; Choe, H.; Farzan, M. Hydroxychloroquine-Mediated Inhibition of SARS-CoV-2 Entry Is Attenuated by TMPRSS2. *PLoS Pathog* **2021**, *17* (1). https://doi.org/10.1371/journal.ppat.1009212.

(25) Lebeau, G.; Vagner, D.; Frumence, É.; Ah-Pine, F.; Guillot, X.; Nobécourt, E.; Raffray, L.; Gasque, P. Deciphering SARS-CoV-2 Virologic and Immunologic Features. *International Journal of Molecular Sciences 2020, Vol. 21, Page 5932* **2020**, *21* (16), 5932. https://doi.org/10.3390/ijms21165932.

(26) Guadarrama-Ortiz, P.; Choreño-Parra, J. A.; Sánchez-Martínez, C. M.; Pacheco-Sánchez, F. J.; Rodríguez-Nava, A. I.; García-Quintero, G. Neurological Aspects of SARS-CoV-2 Infection: Mechanisms and Manifestations. *Front Neurol* **2020**, *11*, 1039. https://doi.org/10.3389/fneur.2020.01039.

(27) Huang, Y.; Yang, C.; Xu, X. feng; Xu, W.; Liu, S. wen. Structural and Functional Properties of SARS-CoV-2 Spike Protein: Potential Antivirus Drug Development for COVID-19. *Acta Pharmacologica Sinica 2020 41:9* **2020**, *41* (9), 1141–1149. https://doi.org/10.1038/s41401-020-0485-4.

(28) Duan, L.; Zheng, Q.; Zhang, H.; Niu, Y.; Lou, Y.; Wang, H. The SARS-CoV-2 Spike Glycoprotein Biosynthesis, Structure, Function, and Antigenicity: Implications for the Design of Spike-Based Vaccine Immunogens. *Front Immunol* **2020**, *11*, 2593. https://doi.org/10.3389/fimmu.2020.576622.

(29) Peng, R.; Wu, L. A.; Wang, Q.; Qi, J.; Gao, G. F. Cell Entry by SARS-CoV-2. *Trends Biochem Sci* **2021**, *46* (10), 848–860. https://doi.org/10.1016/j.tibs.2021.06.001.

(30) Delano, W. L. The PyMOL Molecular Graphics System. Schrödinger, LLC. 2015.

(31) Watanabe, Y.; Allen, J. D.; Wrapp, D.; McLellan, J. S.; Crispin, M. Site-Specific Glycan Analysis of the SARS-CoV-2 Spike. *Science (1979)* **2020**, *369* (6501), 330–333. https://doi.org/10.1126/science.abb9983.

(32) Gong, Y.; Qin, S.; Dai, L.; Tian, Z. The Glycosylation in SARS-CoV-2 and Its Receptor ACE2. *Signal Transduction and Targeted Therapy 2021 6:1* **2021**, *6* (1), 1–24. https://doi.org/10.1038/s41392-021-00809-8.

(33) Henderson, R.; Edwards, R. J.; Mansouri, K.; Janowska, K.; Stalls, V.; Kopp, M.; Haynes, B. F.; Acharya, P. Glycans on the SARS-CoV-2 Spike Control the Receptor Binding Domain Conformation. *bioRxiv* **2020**. https://doi.org/10.1101/2020.06.26.173765.

(34) Sztain, T.; Ahn, S. H.; Bogetti, A. T.; Casalino, L.; Goldsmith, J. A.; Seitz, E.; McCool, R. S.; Kearns, F. L.; Acosta-Reyes, F.; Maji, S.; Mashayekhi, G.; McCammon, J. A.; Ourmazd, A.; Frank, J.; McLellan, J. S.; Chong, L. T.; Amaro, R. E. A Glycan Gate Controls Opening of the SARS-CoV-2 Spike Protein. *Nat Chem* **2021**, *13* (10), 963–968. https://doi.org/10.1038/s41557-021-00758-3.

(35) Rajpoot, S.; Ohishi, T.; Kumar, A.; Pan, Q.; Banerjee, S.; Zhang, K. Y. J.; Baig, M. S. A Novel Therapeutic Peptide Blocks SARS-CoV-2 Spike Protein Binding with Host Cell ACE2 Receptor. *Drugs in R and D* **2021**, *21* (3), 273–283. https://doi.org/10.1007/s40268-021-00357-0.

(36) Gross, L. Z. F.; Sacerdoti, M.; Piiper, A.; Zeuzem, S.; Leroux, A. E.; Biondi, R. M. ACE2, the Receptor That Enables Infection by SARS-CoV-2: Biochemistry, Structure, Allostery and Evaluation of the Potential Development of ACE2 Modulators. *ChemMedChem* **2020**, *15* (18), 1682–1690. https://doi.org/10.1002/cmdc.202000368.

(37) Shang, J.; Ye, G.; Shi, K.; Wan, Y.; Luo, C.; Aihara, H.; Geng, Q.; Auerbach, A.; Li, F. Structural Basis of Receptor Recognition by SARS-CoV-2. *Nature* **2020**, *581* (7807), 221–224. https://doi.org/10.1038/s41586-020-2179-y.

(38) Yan, R.; Zhang, Y.; Li, Y.; Xia, L.; Guo, Y.; Zhou, Q. Structural Basis for the Recognition of SARS-CoV-2 by Full-Length Human ACE2. *Science (1979)* **2020**, *367* (6485), 1444–1448. https://doi.org/10.1126/science.abb2762.

(39) Shajahan, A.; Archer-Hartmann, S.; Supekar, N. T.; Gleinich, A. S.; Heiss, C.; Azadi, P. Comprehensive Characterization of N- and O- Glycosylation of SARS-CoV-2 Human Receptor Angiotensin Converting Enzyme 2. *Glycobiology* **2021**, *31* (4), 410–424. https://doi.org/10.1093/glycob/cwaa101.

(40) Mehdipour, A. R.; Hummer, G. Dual Nature of Human ACE2 Glycosylation in Binding to SARS-CoV-2 Spike. *Proc Natl Acad Sci U S A* **2021**, *118* (19). https://doi.org/10.1073/pnas.2100425118.

(41) Cao, L.; Goreshnik, I.; Coventry, B.; Case, J. B.; Miller, L.; Kozodoy, L.; Chen, R. E.; Carter, L.; Walls, A. C.; Park, Y. J.; Strauch, E. M.; Stewart, L.; Diamond, M. S.; Veesler,

D.; Baker, D. De Novo Design of Picomolar SARS-CoV-2 Miniprotein Inhibitors. *Science* **2020**, *370* (6515). https://doi.org/10.1126/SCIENCE.ABD9909.

(42) Muhseen, Z. T.; Hameed, A. R.; Al-Hasani, H. M. H.; Tahir ul Qamar, M.; Li, G. Promising Terpenes as SARS-CoV-2 Spike Receptor-Binding Domain (RBD) Attachment Inhibitors to the Human ACE2 Receptor: Integrated Computational Approach. *J Mol Liq* **2020**, *320*. https://doi.org/10.1016/j.molliq.2020.114493.

(43) Kumar, V.; Liu, H.; Wu, C. Drug Repurposing against SARS-CoV-2 Receptor Binding Domain Using Ensemble-Based Virtual Screening and Molecular Dynamics Simulations. *Comput Biol Med* **2021**, *135*. https://doi.org/10.1016/j.compbiomed.2021.104634.

(44) Dejnirattisai, W.; Zhou, D.; Ginn, H. M.; Duyvesteyn, H. M. E.; Supasa, P.; Case, J. B.; Zhao, Y.; Walter, T. S.; Mentzer, A. J.; Liu, C.; Wang, B.; Paesen, G. C.; Slon-Campos, J.; López-Camacho, C.; Kafai, N. M.; Bailey, A. L.; Chen, R. E.; Ying, B.; Thompson, C.; Bolton, J.; Fyfe, A.; Gupta, S.; Tan, T. K.; Gilbert-Jaramillo, J.; James, W.; Knight, M.; Carroll, M. W.; Skelly, D.; Dold, C.; Peng, Y.; Levin, R.; Dong, T.; Pollard, A. J.; Knight, J. C.; Klenerman, P.; Temperton, N.; Hall, D. R.; Williams, M. A.; Paterson, N. G.; Bertram, F. K. R.; Siebert, C. A.; Clare, D. K.; Howe, A.; Radecke, J.; Song, Y.; Townsend, A. R.; Huang, K.-Y. A.; Fry, E. E.; Mongkolsapaya, J.; Diamond, M. S.; Ren, J.; Stuart, D. I.; Screaton, G. R. The Antigenic Anatomy of SARS-CoV-2 Receptor Binding Domain. *Cell* **2021**, *184* (8), 2183–2200. https://doi.org/10.1016/j.cell.2021.02.032.

(45) Barnes, C. O.; Jette, C. A.; Abernathy, M. E.; Dam, K.-M. A.; Esswein, S. R.; Gristick, H. B.; Malyutin, A. G.; Sharaf, N. G.; Huey-Tubman, K. E.; Lee, Y. E.; Robbiani, D. F.; Nussenzweig, M. C.; West, A. P.; Bjorkman, P. J. SARS-CoV-2 Neutralizing Antibody Structures Inform Therapeutic Strategies. *Nature* **2020**, *588* (7839), 682–687. https://doi.org/10.1038/s41586-020-2852-1.

(46) Greaney, A. J.; Starr, T. N.; Barnes, C. O.; Weisblum, Y.; Schmidt, F.; Caskey, M.; Gaebler, C.; Cho, A.; Agudelo, M.; Finkin, S.; Wang, Z.; Poston, D.; Muecksch, F.; Hatziioannou, T.; Bieniasz, P. D.; Robbiani, D. F.; Nussenzweig, M. C.; Bjorkman, P. J.; Bloom, J. D. Mapping Mutations to the SARS-CoV-2 RBD That Escape Binding by Different Classes of Antibodies. *Nat Commun* **2021**, *12* (1), 4196. https://doi.org/10.1038/s41467-021-24435-8.

(47) Chen, Y.; Zhao, X.; Zhou, H.; Zhu, H.; Jiang, S.; Wang, P. Broadly Neutralizing Antibodies to SARS-CoV-2 and Other Human Coronaviruses. *Nat Rev Immunol* **2022**. https://doi.org/10.1038/s41577-022-00784-3.

(48) Shrestha, L. B.; Tedla, N.; Bull, R. A. Broadly-Neutralizing Antibodies Against Emerging SARS-CoV-2 Variants. *Front Immunol* **2021**, *12*. https://doi.org/10.3389/fimmu.2021.752003.

(49) Yuan, M.; Wu, N. C.; Zhu, X.; Lee, C.-C. D.; So, R. T. Y.; Lv, H.; Mok, C. K. P.; Wilson, I. A. A Highly Conserved Cryptic Epitope in the Receptor Binding Domains of SARS-

CoV-2 and SARS-CoV. *Science (1979)* **2020**, *368* (6491), 630–633. https://doi.org/10.1126/science.abb7269.

(50) *Tracking SARS-CoV-2 variants*. https://www.who.int/activities/tracking-SARS-CoV-2-variants (accessed 2022-10-31).

(51) *SARS-CoV-2 variants of concern as of 27 October 2022*. https://www.ecdc.europa.eu/en/covid-19/variants-concern (accessed 2022-10-31).

(52) Celik, I.; Yadav, R.; Duzgun, Z.; Albogami, S.; El-Shehawi, A. M.; Fatimawali; Idroes, R.; Tallei, T. E.; Emran, T. bin. Interactions of the Receptor Binding Domain of SARS-CoV-2 Variants with HACE2: Insights from Molecular Docking Analysis and Molecular Dynamic Simulation. *Biology 2021, Vol. 10, Page 880* **2021**, *10* (9), 880. https://doi.org/10.3390/biology10090880.

(53) Lou, F.; Li, M.; Pang, Z.; Jiang, L.; Guan, L.; Tian, L.; Hu, J.; Fan, J.; Fan, H. Understanding the Secret of SARS-CoV-2 Variants of Concern/Interest and Immune Escape. *Front Immunol* **2021**, *12*, 4326. https://doi.org/10.3389/fimmu.2021.744242.

(54) Koley, T.; Goswami, A.; Kumar, M.; Upadhyay, N.; Hariprasad, G. Comparative Structural Analysis of Human ACE2 Receptor with Spike Protein of SARS-CoV-2 Variants: Implications to Understand Infectivity of the Virus. *Advances and Applications in Bioinformatics and Chemistry* **2022**, *15*, 21–27. https://doi.org/10.2147/aabc.s360787.

(55) Saito, A.; Irie, T.; Suzuki, R.; Maemura, T.; Nasser, H.; Uriu, K.; Kosugi, Y.; Shirakawa, K.; Sadamasu, K.; Kimura, I.; Ito, J.; Wu, J.; Iwatsuki-Horimoto, K.; Ito, M.; Yamayoshi, S.; Loeber, S.; Tsuda, M.; Wang, L.; Ozono, S.; Butlertanaka, E. P.; Tanaka, Y. L.; Shimizu, R.; Shimizu, K.; Yoshimatsu, K.; Kawabata, R.; Sakaguchi, T.; Tokunaga, K.; Yoshida, I.; Asakura, H.; Nagashima, M.; Kazuma, Y.; Nomura, R.; Horisawa, Y.; Yoshimura, K.; Takaori-Kondo, A.; Imai, M.; Chiba, M.; Furihata, H.; Hasebe, H.; Kitazato, K.; Kubo, H.; Misawa, N.; Morizako, N.; Noda, K.; Oide, A.; Suganami, M.; Takahashi, M.; Tsushima, K.; Yokoyama, M.; Yuan, Y.; Tanaka, S.; Nakagawa, S.; Ikeda, T.; Fukuhara, T.; Kawaoka, Y.; Sato, K. Enhanced Fusogenicity and Pathogenicity of SARS-CoV-2 Delta P681R Mutation. *Nature* **2021**, *602* (7896), 300–306. https://doi.org/10.1038/s41586-021-04266-9.

(56) Michael Rajah, M.; Hubert, M.; Bishop, E.; Saunders, N.; Robinot, R.; Grzelak, L.; Planas, D.; er emy Dufloo, J.; Gellenoncourt, S.; Bongers, A.; Zivaljic, M.; Planchais, C.; Guivel-Benhassine, F.; Porrot, F.; Mouquet, H.; Chakrabarti, L. A.; Buchrieser, J.; Schwartz, O. SARS-CoV-2 Alpha, Beta, and Delta Variants Display Enhanced Spike-Mediated Syncytia Formation. *EMBO J* **2021**, *40* (24), e108944. https://doi.org/10.15252/embj.2021108944.

(57) Saunders, N.; Planas, D.; Bolland, W. H.; Rodriguez, C.; Fourati, S.; Buchrieser, J.; Planchais, C.; Prot, M.; Staropoli, I.; Guivel-Benhassine, F.; Porrot, F.; Veyer, D.; Péré, H.; Robillard, N.; Saliba, M.; Baidaliuk, A.; Seve, A.; Hocqueloux, L.; Prazuck, T.; Rey, F. A.; Mouquet, H.; Simon-Lorière, E.; Bruel, T.; Pawlotsky, J. M.; Schwartz, O.

Fusogenicity and Neutralization Sensitivity of the SARS-CoV-2 Delta Sublineage AY.4.2. *EBioMedicine* **2022**, *77*, 103934. https://doi.org/10.1016/j.ebiom.2022.103934.

(58) Chen, C.; Nadeau, S.; Yared, M.; Voinov, P.; Xie, N.; Roemer, C.; Stadler, T. CoV-Spectrum: Analysis of Globally Shared SARS-CoV-2 Data to Identify and Characterize New Variants. *Bioinformatics* **2022**, *38* (6), 1735–1737. https://doi.org/10.1093/bioinformatics/btab856.

(59) Chakraborty, C.; Bhattacharya, M.; Sharma, A. R. Present Variants of Concern and Variants of Interest of Severe Acute Respiratory Syndrome Coronavirus *2*: Their Significant Mutations in S-Glycoprotein, Infectivity, Re-Infectivity, Immune Escape and Vaccines Activity. *Rev Med Virol* **2022**, *32* (2), e2270. https://doi.org/10.1002/rmv.2270.

(60) Graham, C.; Seow, J.; Huettner, I.; Khan, H.; Kouphou, N.; Acors, S.; Winstone, H.; Pickering, S.; Galao, R. P.; Dupont, L.; Lista, M. J.; Jimenez-Guardeño, J. M.; Laing, A. G.; Wu, Y.; Joseph, M.; Muir, L.; van Gils, M. J.; Ng, W. M.; Duyvesteyn, H. M. E.; Zhao, Y.; Bowden, T. A.; Shankar-Hari, M.; Rosa, A.; Cherepanov, P.; McCoy, L. E.; Hayday, A. C.; Neil, S. J. D.; Malim, M. H.; Doores, K. J. Neutralization Potency of Monoclonal Antibodies Recognizing Dominant and Subdominant Epitopes on SARS-CoV-2 Spike Is Impacted by the B.1.1.7 Variant. *Immunity* **2021**, *54* (6), 1276-1289.e6. https://doi.org/10.1016/j.immuni.2021.03.023.

(61) Ali, F.; Kasry, A.; Amin, M. The New SARS-CoV-*2* Strain Shows a Stronger Binding Affinity to ACE2 Due to N501Y Mutant. *Med Drug Discov* **2021**, *10*. https://doi.org/10.1016/j.medidd.2021.100086.

(62) Laffeber, C.; de Koning, K.; Kanaar, R.; Lebbink, J. H. G. Experimental Evidence for Enhanced Receptor Binding by Rapidly Spreading SARS-CoV-2 Variants. *J Mol Biol* **2021**, *433* (15), 167058. https://doi.org/10.1016/j.jmb.2021.167058.

(63) Tian, F.; Tong, B.; Sun, L.; Shi, S.; Zheng, B.; Wang, Z.; Dong, X.; Zheng, P. Mutation N501Y in RBD of Spike Protein Strengthens the Interaction between COVID-19 and Its Receptor ACE2. *Elife* **2021**. https://doi.org/10.7554/eLife.69091.

(64) Alenquer, M.; Ferreira, F.; Lousa, D.; Valério, M.; Medina-Lopes, M.; Bergman, M. L.; Gonçalves, J.; Demengeot, J.; Leite, R. B.; Lilue, J.; Ning, Z.; Penha-Gonçalves, C.; Soares, H.; Soares, C. M.; Amorim, M. J. Signatures in SARS-CoV-2 Spike Protein Conferring Escape to Neutralizing Antibodies. *PLoS Pathog* **2021**, *17* (8), e1009772. https://doi.org/10.1371/journal.ppat.1009772.

(65) Deng, X.; Garcia-Knight, M. A.; Khalid, M. M.; Servellita, V.; Wang, C.; Morris, M. K.; Sotomayor-González, A.; Glasner, D. R.; Reyes, K. R.; Gliwa, A. S.; Reddy, N. P.; Sanchez San Martin, C.; Federman, S.; Cheng, J.; Balcerek, J.; Taylor, J.; Streithorst, J. A.; Miller, S.; Sreekumar, B.; Chen, P. Y.; Schulze-Gahmen, U.; Taha, T. Y.; Hayashi, J. M.; Simoneau, C. R.; Kumar, G. R.; McMahon, S.; Lidsky, P. v.; Xiao, Y.; Hemarajata, P.; Green, N. M.; Espinosa, A.; Kath, C.; Haw, M.; Bell, J.; Hacker, J. K.; Hanson, C.; Wadford, D. A.; Anaya, C.; Ferguson, D.; Frankino, P. A.; Shivram, H.; Lareau, L. F.; Wyman, S. K.; Ott, M.; Andino, R.; Chiu, C. Y. Transmission, Infectivity, and

Neutralization of a Spike L452R SARS-CoV-2 Variant. *Cell* **2021**, *184* (13), 3426–3437. https://doi.org/10.1016/j.cell.2021.04.025.

(66) Kim, S.; Liu, Y.; Lei, Z.; Dicker, J.; Cao, Y.; Zhang, X. F.; Im, W. Differential Interactions Between Human ACE2 and Spike RBD of SARS-CoV-2 Variants of Concern. *Jornal of Chemical Theory and Computation* **2021**, *17*, 7972–7979. https://doi.org/10.1021/acs.jctc.1c00965.

(67) Motozono, C.; Toyoda, M.; Zahradnik, J.; Saito, A.; Nasser, H.; Tan, T. S.; Ngare, I.; Kimura, I.; Uriu, K.; Kosugi, Y.; Yue, Y.; Shimizu, R.; Ito, J.; Torii, S.; Yonekawa, A.; Shimono, N.; Nagasaki, Y.; Minami, R.; Toya, T.; Sekiya, N.; Fukuhara, T.; Matsuura, Y.; Schreiber, G.; Ikeda, T.; Nakagawa, S.; Ueno, T.; Sato, K. SARS-CoV-2 Spike L452R Variant Evades Cellular Immunity and Increases Infectivity. *Cell Host Microbe* **2021**, *29* (7), 1124–1136. https://doi.org/10.1016/j.chom.2021.06.006.

(68) Wang, Q.; Guo, Y.; Iketani, S.; Nair, M. S.; Li, Z.; Mohri, H.; Wang, M.; Yu, J.; Bowen, A. D.; Chang, J. Y.; Shah, J. G.; Nguyen, N.; Chen, Z.; Meyers, K.; Yin, M. T.; Sobieszczyk, M. E.; Sheng, Z.; Huang, Y.; Liu, L.; Ho, D. D. Antibody Evasion by SARS-CoV-2 Omicron Subvariants BA.2.12.1, BA.4 and BA.5. *Nature 2022 608:7923* **2022**, *608* (7923), 603–608. https://doi.org/10.1038/s41586-022-05053-w.

(69) Rao, S.; MD; MA; DPhill; Singh, M.; FACS. *View of The Newly Detected B.1.1.529 (Omicron) Variant of SARS-CoV-2 With Multiple Mutations*; 2021. https://doi.org/https://doi.org/10.47488/dhrp.v1iS5.35.

(70) Han, P.; Li, L.; Liu, S.; Wang, Q.; Zhang, D.; Xu, Z.; Han, P.; Li, X.; Peng, Q.; Su, C.; Huang, B.; Li, D.; Zhang, R.; Tian, M.; Fu, L.; Gao, Y.; Zhao, X.; Liu, K.; Qi, J.; Gao, G. F.; Wang, P. Receptor Binding and Complex Structures of Human ACE2 to Spike RBD from Omicron and Delta SARS-CoV-2. *Cell* **2022**, *185* (4), 630–640. https://doi.org/10.1016/j.cell.2022.01.001.

(71) Mannar, D.; Saville, J. W.; Zhu, X.; Srivastava, S. S.; Berezuk, A. M.; Tuttle, K. S.; Marquez, C.; Sekirov, I.; Subramaniam, S. SARS-CoV-2 Omicron Variant: ACE2 Binding, Cryo-EM Structure of Spike Protein-ACE2 Complex and Antibody Evasion. *bioRxiv* **2021**. https://doi.org/10.1101/2021.12.19.473380.

(72) Desingu, P. A.; Nagarajan, K.; Dhama, K. Emergence of Omicron Third Lineage BA.3 and Its Importance. *J Med Virol* **2022**, *94* (5), 1808–1810. https://doi.org/10.1002/jmv.27601.

(73) Kumar, S.; Karuppanan, K.; Subramaniam, G. Omicron (BA.1) and Sub-Variants (BA.1, BA.2 and BA.3) of SARS-CoV-2 Spike Infectivity and Pathogenicity: A Comparative Sequence and Structural-Based Computational Assessment. *bioRxiv* **2022**. https://doi.org/10.1101/2022.02.11.480029.

(74) Schubert, M.; Bertoglio, F.; Steinke, S.; Heine, P. A.; Ynga-Durand, M. A.; Maass, H.; Sammartino, J. C.; Cassaniti, I.; Zuo, F.; Du, L.; Korn, J.; Milošević, M.; Wenzel, E. V.; Krstanović, F.; Polten, S.; Pribanić-Matešić, M.; Brizić, I.; Baldanti, F.; Hammarström,

L.; Dübel, S.; Šustić, A.; Marcotte, H.; Strengert, M.; Protić, A.; Piralla, A.; Pan-Hammarström, Q.; Čičin-Šain, L.; Hust, M. Human Serum from SARS-CoV-2-Vaccinated and COVID-19 Patients Shows Reduced Binding to the RBD of SARS-CoV-2 Omicron Variant. *BMC Med* **2022**, *20* (1). https://doi.org/10.1186/S12916-022-02312-5.

(75) Zhao, H.; Lu, L.; Peng, Z.; Chen, L. L.; Meng, X.; Zhang, C.; Ip, J. D.; Chan, W. M.; Chu, A. W. H.; Chan, K. H.; Jin, D. Y.; Chen, H.; Yuen, K. Y.; To, K. K. W. SARS-CoV-2 Omicron Variant Shows Less Efficient Replication and Fusion Activity When Compared with Delta Variant in TMPRSS2-Expressed Cells. *Emerg Microbes Infect* **2022**, *11* (1), 277–283. https://doi.org/10.1080/22221751.2021.2023329.

(76) Meng, B.; Abdullahi, A.; Ferreira, I. A. T. M.; Goonawardane, N. *et al*. Altered TMPRSS2 Usage by SARS-CoV-2 Omicron Impacts Infectivity and Fusogenicity. *Nature* **2022**, *603* (7902), 706–714. https://doi.org/10.1038/s41586-022-04474-x.

(77) Peacock, T. P.; Brown, J. C.; Zhou, J.; Thakur, N.; Sukhova, K.; Newman, J.; Kugathasan, R.; Yan, A. W. C.; Furnon, W.; Lorenzo, G. de; Cowton, V. M.; Reuss, D.; Moshe, M.; Quantrill, J. L.; Platt, O. K.; Kaforou, M.; Patel, A. H.; Palmarini, M.; Bailey, D.; Barclay, W. S. The Altered Entry Pathway and Antigenic Distance of the SARS-CoV-2 Omicron Variant Map to Separate Domains of Spike Protein. *bioRxiv* **2022**. https://doi.org/10.1101/2021.12.31.474653.

(78) Focosi, D.; McConnell, S.; Casadevall, A. The Omicron Variant of Concern: Diversification and Convergent Evolution in Spike Protein, and Escape from Anti-Spike Monoclonal Antibodies. *Drug Resistance Updates* **2022**, *65*. https://doi.org/10.1016/j.drup.2022.100882.

(79) Liu, Z.; VanBlargan, L. A.; Bloyet, L. M.; Rothlauf, P. W.; Chen, R. E.; Stumpf, S.; Zhao, H.; Errico, J. M.; Theel, E. S.; Liebeskind, M. J.; Alford, B.; Buchser, W. J.; Ellebedy, A. H.; Fremont, D. H.; Diamond, M. S.; Whelan, S. P. J. Identification of SARS-CoV-2 Spike Mutations That Attenuate Monoclonal and Serum Antibody Neutralization. *Cell Host Microbe* **2021**, *29* (3), 477–488. https://doi.org/10.1016/j.chom.2021.01.014.

(80) Lan, J.; He, X.; Ren, Y.; Wang, Z.; Zhou, H.; Fan, S.; Zhu, C.; Liu, D.; Shao, B.; Liu, T.-Y.; Wang, Q.; Zhang, L.; Ge, J.; Wang, T.; Wang, X. Structural and Computational Insights into the SARS-CoV-2 Omicron RBD-ACE2 Interaction. *bioRxiv* **2022**, 2022.01.03.474855. https://doi.org/10.1101/2022.01.03.474855.

(81) Zhou, H.; Dcosta, B. M.; Landau, N. R.; Tada, T. Resistance of SARS-CoV-2 Omicron BA.1 and BA.2 Variants to Vaccine-Elicited Sera and Therapeutic Monoclonal Antibodies. *Viruses 2022, Vol. 14, Page 1334* **2022**, *14* (6), 1334. https://doi.org/10.3390/V14061334.

(82) Iketani, S.; Liu, L.; Guo, Y.; Liu, L.; Chan, J. F. W.; Huang, Y.; Wang, M.; Luo, Y.; Yu, J.; Chu, H.; Chik, K. K. H.; Yuen, T. T. T.; Yin, M. T.; Sobieszczyk, M. E.; Huang, Y.; Yuen, K. Y.; Wang, H. H.; Sheng, Z.; Ho, D. D. Antibody Evasion Properties of SARS-

CoV-2 Omicron Sublineages. *Nature 2022 604:7906* **2022**, *604* (7906), 553–556. https://doi.org/10.1038/s41586-022-04594-4.

(83) Qu, P.; Evans, J. P.; Faraone, J.; Zheng, Y.-M.; Carlin, C.; Anghelina, M.; Stevens, P.; Fernandez, S.; Jones, D.; Lozanski, G.; Panchal, A.; Saif, L. J.; Oltz, E. M.; Xu, K.; Gumina, R. J.; Liu, S.-L.; Davis, D. M. Distinct Neutralizing Antibody Escape of SARS-CoV-2 Omicron Subvariants BQ.1, BQ.1.1, BA.4.6, BF.7 and BA.2.75.2. *bioRxiv* **2022**. https://doi.org/10.1101/2022.10.19.512891.

(84) Lan, J.; Ge, J.; Yu, J.; Shan, S.; Zhou, H.; Fan, S.; Zhang, Q.; Shi, X.; Wang, Q.; Zhang, L.; Wang, X. Structure of the SARS-CoV-2 Spike Receptor-Binding Domain Bound to the ACE2 Receptor. *Nature 2020 581:7807* **2020**, *581* (7807), 215–220. https://doi.org/10.1038/s41586-020-2180-5.

(85) Valério, M.; Borges-Araújo, L.; Melo, M. N.; Lousa, D.; Soares, C. M. SARS-CoV-2 Variants Impact RBD Conformational Dynamics and ACE2 Accessibility. *Front Med Technol* **2022**, *0*, 72. https://doi.org/10.3389/fmedt.2022.1009451.

(86) Schlick, T. *Molecular Modeling and Simulation: An Interdiscisplinary Guide*, 2nd ed.; Antam, S. S., Marsden, J. E., Sirovich, L., Eds.; 2010; Vol. 21.

(87) Leach, A. R. *Molecular Modelling: Principles and Applications.*, 2nd ed.; 2001.

(88) van Gunsteren, W. F.; Bakowies, D.; Baron, R.; Chandrasekhar, I.; Christen, M.; Daura, X.; Gee, P.; Geerke, D. P.; Glättli, A.; Hünenberger, P. H.; Kastenholz, M. A.; Oostenbrink, C.; Schenk, M.; Trzesniak, D.; van der Vegt, N. F. A.; Yu, H. B. Biomolecular Modeling: Goals, Problems, Perspectives. *Angewandte Chemie International Edition* **2006**, *45* (25), 4064–4092. https://doi.org/10.1002/anie.200502655.

(89) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins* **2006**, *65* (3), 712–725. https://doi.org/10.1002/prot.21123.

(90) Doltsinis, N. L. *Molecular Dynamics Beyond the Born-Oppenheimer Approximation: Mixed Quantum-Classical Approaches*; Grotendorst, J., Blügel, S., Marx, D., Eds.; NIC Series, 2006; Vol. 31.

(91) González, M. A. Force Fields and Molecular Dynamics Simulations. *École thématique de la Société Française de la Neutronique* **2011**, *12*, 169–200. https://doi.org/10.1051/sfn/201112009.

(92) Aduri, R.; Psciuk, B. T.; Saro, P.; Taniga, H.; Schlegel, H. B.; SantaLucia, J. AMBER Force Field Parameters for the Naturally Occurring Modified Nucleosides in RNA. *J Chem Theory Comput* **2007**, *3* (4), 1464–1475. https://doi.org/10.1021/ct600329w.

(93) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber Biomolecular Simulation

Programs. *J Comput Chem* **2005**, *26* (16), 1668–1688. https://doi.org/10.1002/JCC.20290.

(94) Bayly, C. I.; Merz, K. M.; Ferguson, D. M.; Cornell, W. D.; Fox, T.; Caldwell, J. W.; Kollman, P. A.; Cieplak, P.; Gould, I. R.; Spellmeyer, D. C. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J Am Chem Soc* **1995**, *117* (19), 5179–5197. https://doi.org/10.1021/ja00124a002.

(95) Case Ross C Walker UC San Diego, D. A.; Thomas Cheatham, A. E. Amber 2022 Reference Manual Principal Contributors to the Current Codes. pp 272–280. https://ambermd.org/contributors.html (accessed 2022-11-05).

(96) Becker, O. M.; MacKerell Jr., A.; Roux, B.; Watanabe, M. *Computational Biochemistry and Biophysics*; CRC Press, 2001. https://doi.org/10.1201/9780203903827.

(97) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Dinola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J Chem Phys* **1998**, *81* (8), 3684. https://doi.org/10.1063/1.448118.

(98) Wu, W.; Owino, J.; Al-Ostaz, A.; Cai, L. Applying Periodic Boundary Conditions in Finite Element Analysis. 2014. https://imechanica.org/files/pbc.pdf (accessed 2022-11-05).

(99) Sharma, S.; Kumar, P.; Chandra, R. Introduction to Molecular Dynamics. *Molecular Dynamics Simulation of Nanocomposites using BIOVIA Materials Studio, Lammps and Gromacs* **2019**, 1–38. https://doi.org/10.1016/B978-0-12-816954-4.00001-2.

(100) Hinchliffe, A. *Molecular Modelling for Beginners*, 2nd ed.; 2008.

(101) Lindahl; Abraham; Hess; Spoel, van der. GROMACS 2020.3 Source Code. Zenodo 2020.

(102) Mark, P.; Nilsson, L. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *Journal of Physical Chemistry A* **2001**, *105* (43), 9954–9960. https://doi.org/10.1021/JP003020W.

(103) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J Chem Phys* **2007**, *126* (1), 014101. https://doi.org/10.1063/1.2408420.

(104) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J Appl Phys* **1998**, *52* (12), 7182. https://doi.org/10.1063/1.328693.

(105) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N·log(N) Method for Ewald Sums in Large Systems. *J Chem Phys* **1998**, *98* (12), 10089. https://doi.org/10.1063/1.464397.

(106) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J Mol Graph* **1996**, *14* (1), 33–38. https://doi.org/https://doi.org/10.1016/0263-7855(96)00018-5.

(107) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. *J Comput Chem* **2011**, *32* (10), 2319–2327. https://doi.org/10.1002/JCC.21787.

(108) Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Comput Sci Eng* **2007**, *9* (3), 90–95. https://doi.org/10.1109/MCSE.2007.55.

(109) Jolliffe, I. T. *Principal Component Analysis*; Springer Series in Statistics; Springer-Verlag: New York, 2002. https://doi.org/10.1007/b98835.

(110) Campos, S. R.; Baptista, A. M. Conformational Analysis in a Multidimensional Energy Landscape: Study of an Arginylglutamate Repeat. *J Phys Chem B* **2009**, *113* (49), 15989–16001.

(111) Sittel, F.; Jain, A.; Stock, G. Principal Component Analysis of Molecular Dynamics: On the Use of Cartesian vs. Internal Coordinates. *J Chem Phys* **2014**, *141* (1), 014111. https://doi.org/10.1063/1.4885338.

(112) Campos, S. R.; Baptista, A. M. *Molecular simulation lab — in-house software*. https://www.itqb.unl.pt/labs/molecular-simulation/in-house-software (accessed 2022-11-09).

(113) Contreras-Riquelme, S.; Garate, J. A.; Perez-Acle, T.; Martin, A. J. M. RIP-MD: A Tool to Study Residue Interaction Networks in Protein Molecular Dynamics. *PeerJ* **2018**, *6* (12). https://doi.org/10.7717/peerj.5998.

(114) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res* **2003**, *13* (11), 2498–2504. https://doi.org/10.1101/gr.1239303.

(115) Dejnirattisai, W.; Zhou, D.; Ginn, H. M.; Duyvesteyn, H. M. E.; Supasa, P.; Case, J. B.; Zhao, Y.; Walter, T. S.; Mentzer, A. J.; Liu, C.; Wang, B.; Paesen, G. C.; Slon-Campos, J.; López-Camacho, C.; Kafai, N. M.; Bailey, A. L.; Chen, R. E.; Ying, B.; Thompson, C.; Bolton, J.; Fyfe, A.; Gupta, S.; Tan, T. K.; Gilbert-Jaramillo, J.; James, W.; Knight, M.; Carroll, M. W.; Skelly, D.; Dold, C.; Peng, Y.; Levin, R.; Dong, T.; Pollard, A. J.; Knight, J. C.; Klenerman, P.; Temperton, N.; Hall, D. R.; Williams, M. A.; Paterson, N. G.; Bertram, F. K. R.; Siebert, C. A.; Clare, D. K.; Howe, A.; Radecke, J.; Song, Y.; Townsend, A. R.; Huang, K. Y. A.; Fry, E. E.; Mongkolsapaya, J.; Diamond, M. S.; Ren, J.; Stuart, D. I.; Screaton, G. R. The Antigenic Anatomy of SARS-CoV-2 Receptor Binding Domain. *Cell* **2021**, *184* (8), 2183-2200.e22. https://doi.org/10.1016/J.CELL.2021.02.032.

(116) Homeyer, N.; Gohlke, H. Free Energy Calculations by the Molecular Mechanics Poisson-Boltzmann Surface Area Method. *Mol Inform* **2012**, *31* (2), 114–122. https://doi.org/10.1002/MINF.201100135.

(117) Cao, Y.; Yisimayi, A.; Jian, F.; Song, W.; Xiao, T.; Wang, L.; Du, S.; Wang, J.; Li, Q.; Chen, X.; Yu, Y.; Wang, P.; Zhang, Z.; Liu, P.; An, R.; Hao, X.; Wang, Y.; Wang, J.; Feng, R.; Sun, H.; Zhao, L.; Zhang, W.; Zhao, D.; Zheng, J.; Yu, L.; Li, C.; Zhang, N.; Wang, R.; Niu, X.; Yang, S.; Song, X.; Chai, Y.; Hu, Y.; Shi, Y.; Zheng, L.; Li, Z.; Gu, Q.; Shao, F.; Huang, W.; Jin, R.; Shen, Z.; Wang, Y.; Wang, X.; Xiao, J.; Xie, X. S.

BA.2.12.1, BA.4 and BA.5 Escape Antibodies Elicited by Omicron Infection. *Nature 2022 608:7923* **2022**, *608* (7923), 593–602. https://doi.org/10.1038/s41586-022-04980-y.

(118) Tuekprakhon, A.; Nutalai, R.; Dijokaite-Guraliuc, A.; Zhou, D.; Ginn, H. M.; Selvaraj, M.; Liu, C.; Mentzer, A. J.; Supasa, P.; Duyvesteyn, H. M. E.; Das, R.; Skelly, D.; Ritter, T. G.; Amini, A.; Bibi, S.; Adele, S.; Johnson, S. A.; Constantinides, B.; Webster, H.; Temperton, N.; Klenerman, P.; Barnes, E.; Dunachie, S. J.; Crook, D.; Pollard, A. J.; Lambe, T.; Goulder, P.; Paterson, N. G.; Williams, M. A.; Hall, D. R.; Conlon, C.; Deeks, A.; Frater, J.; Frending, L.; Gardiner, S.; Jämsén, A.; Jeffery, K.; Malone, T.; Phillips, E.; Rothwell, L.; Stafford, L.; Fry, E. E.; Huo, J.; Mongkolsapaya, J.; Ren, J.; Stuart, D. I.; Screaton, G. R. Antibody Escape of SARS-CoV-2 Omicron BA.4 and BA.5 from Vaccine and BA.1 Serum. *Cell* **2022**, *185* (14), 2422-2433.e13. https://doi.org/10.1016/j.cell.2022.06.005.

Supplementary Information

1. Distance between hACE2 NTD and CTD helices when bound to wt RBD

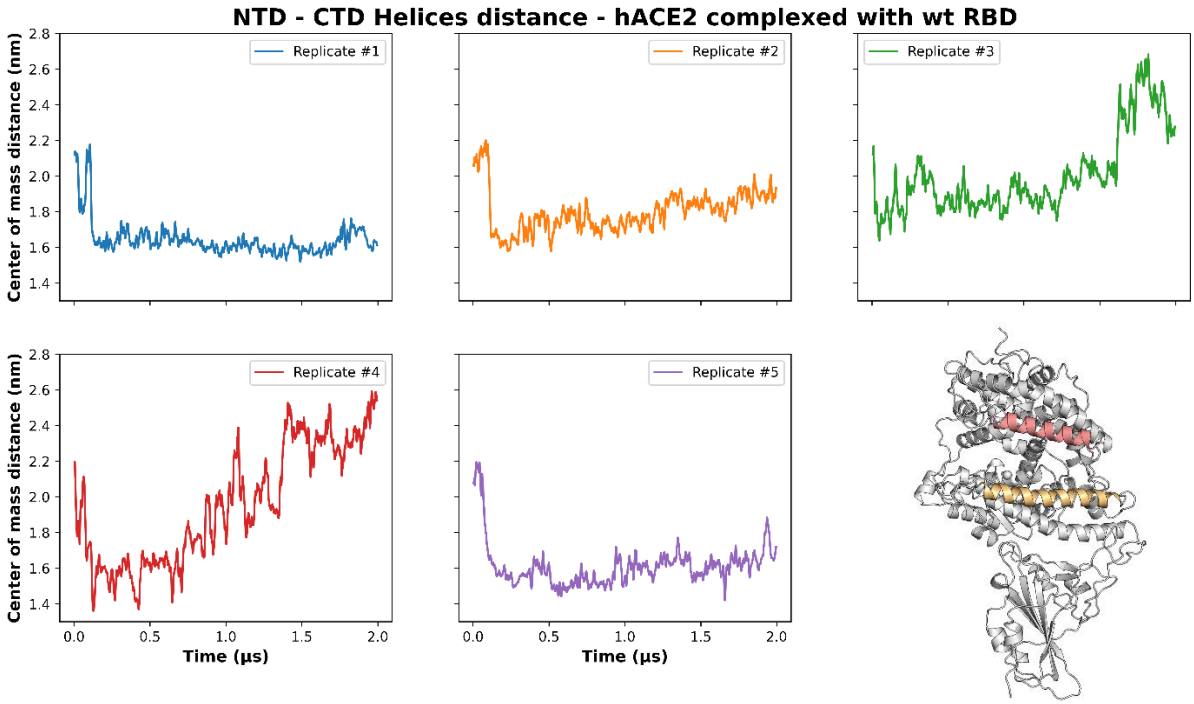


Figure S1 – Time evolution of the distance between the centers of mass of the N-terminal and C-terminal helices of hACE2 bound to wt RBD (residues 38-64 and residues 92-111, respectively). The central moving averages (considering 101 points = 10 ns) of the RMSD are represented to filter high-frequency oscillations. In the lower right corner, the N-terminal and C-terminal subdomain's helices are highlighted in the wt RBD-hACE2 structure.

2. Temporal Evolution of the RMSD: RBD-hACE2 complexes

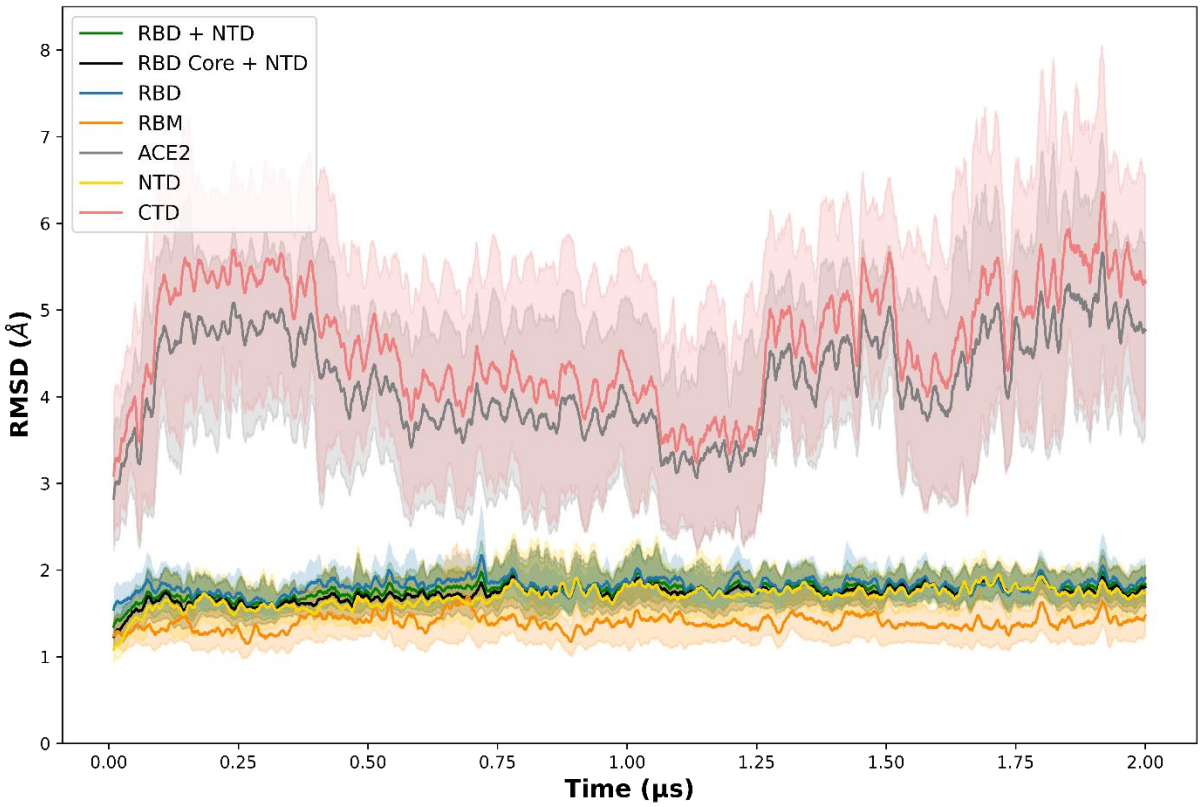


Figure S2 - Average (over the five replicates) time evolution of the Cα RMSD of selected zones of the wt RBD-hACE2 complex (only regular secondary structure zones). RBD Core and NTD Cα were fitted against the X-ray structure from PDB ID: 6M0J.The averages are represented for the RBD, RBM, hACE2, NTD and CTD subdomains, are represented with lines colored blue, orange, gray, yellow-gold, and coral, respectively. The wt RBD-NTD hACE2 complex RMSD is represented in green used for further analysis. The upper and lower bound error estimations are expressed as twice the standard deviation obtained using bootstrapping. The central moving averages (considering 101 points = 10 ns) of the RMSD are represented to filter high-frequency oscillations.

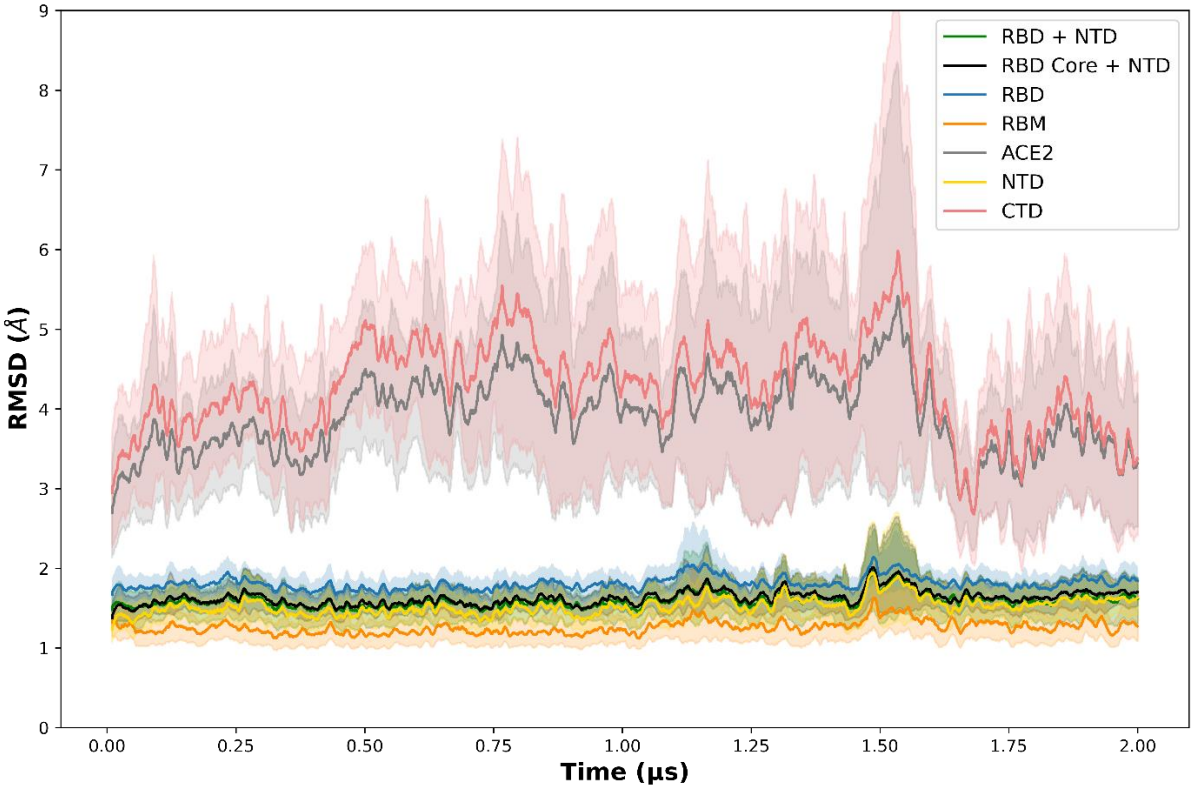


Figure S3 - Average (over the five replicates) time evolution of the Cα RMSD of selected zones of the BA.1 RBD-hACE2 complex (only regular secondary structure zones). RBD Core and NTD Cα were fitted against the X-ray structure from PDB ID: 6M0J.The averages for the selected domains are represented similarly to the figure **S2**. The BA.1 RBD-NTD hACE2 complex RMSD is represented in green used for further analysis. The upper and lower bound error estimations are expressed as twice the standard deviation obtained using bootstrapping. The central moving averages (considering 101 points = 10 ns) of the RMSD are represented to filter high-frequency oscillations.

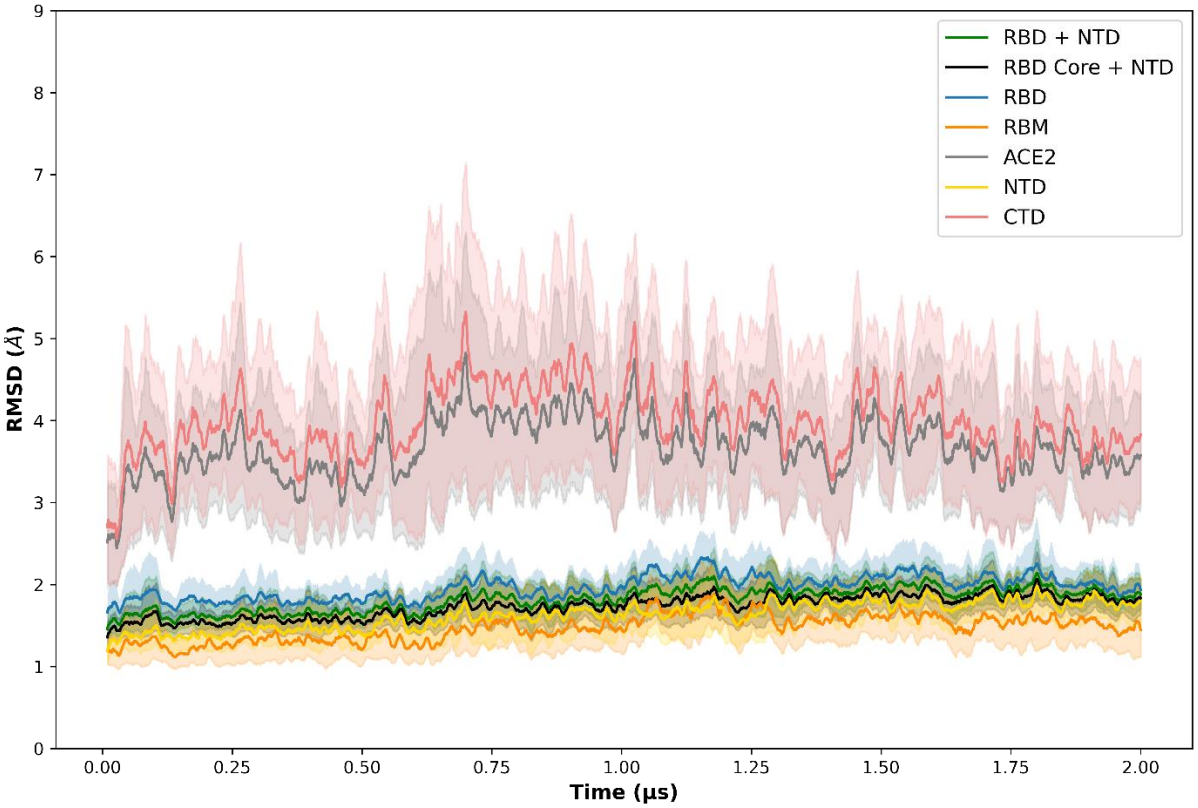


Figure S4 - Average (over the five replicates) time evolution of the Cα RMSD of selected zones of the BA.2 RBD-hACE2 complex (only regular secondary structure zones). RBD Core and NTD Cα were fitted against the X-ray structure from PDB ID: 6M0J.The averages for the selected domains are represented similarly to the figure **S2**. The BA.2 RBD-NTD hACE2 complex RMSD is represented in green used for further analysis. The upper and lower bound error estimations are expressed as twice the standard deviation obtained using bootstrapping. The central moving averages (considering 101 points = 10 ns) of the RMSD are represented to filter high-frequency oscillations.

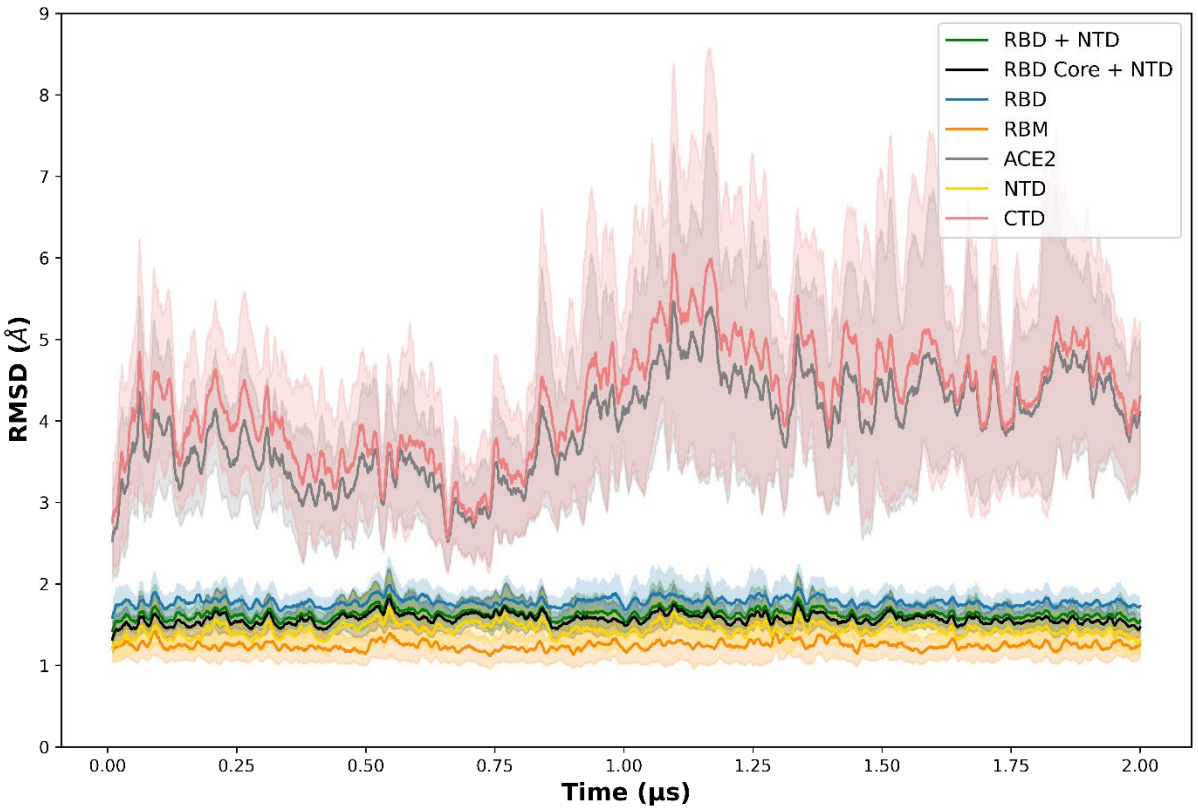


Figure S5 - Average (over the five replicates) time evolution of the Cα RMSD of selected zones of the BA.5 RBD-hACE2 complex (only regular secondary structure zones). RBD Core and NTD Cα were fitted against the X-ray structure from PDB ID: 6M0J.The averages for the selected domains are represented similarly to the figure **S2**. The BA.2 RBD-NTD hACE2 complex RMSD is represented in green used for further analysis. The upper and lower bound error estimations are expressed as twice the standard deviation obtained using bootstrapping. The central moving averages (considering 101 points = 10 ns) of the RMSD are represented to filter high-frequency oscillations.

3. Residue Interaction Networks (RIN): RBD-hACE2 complexes

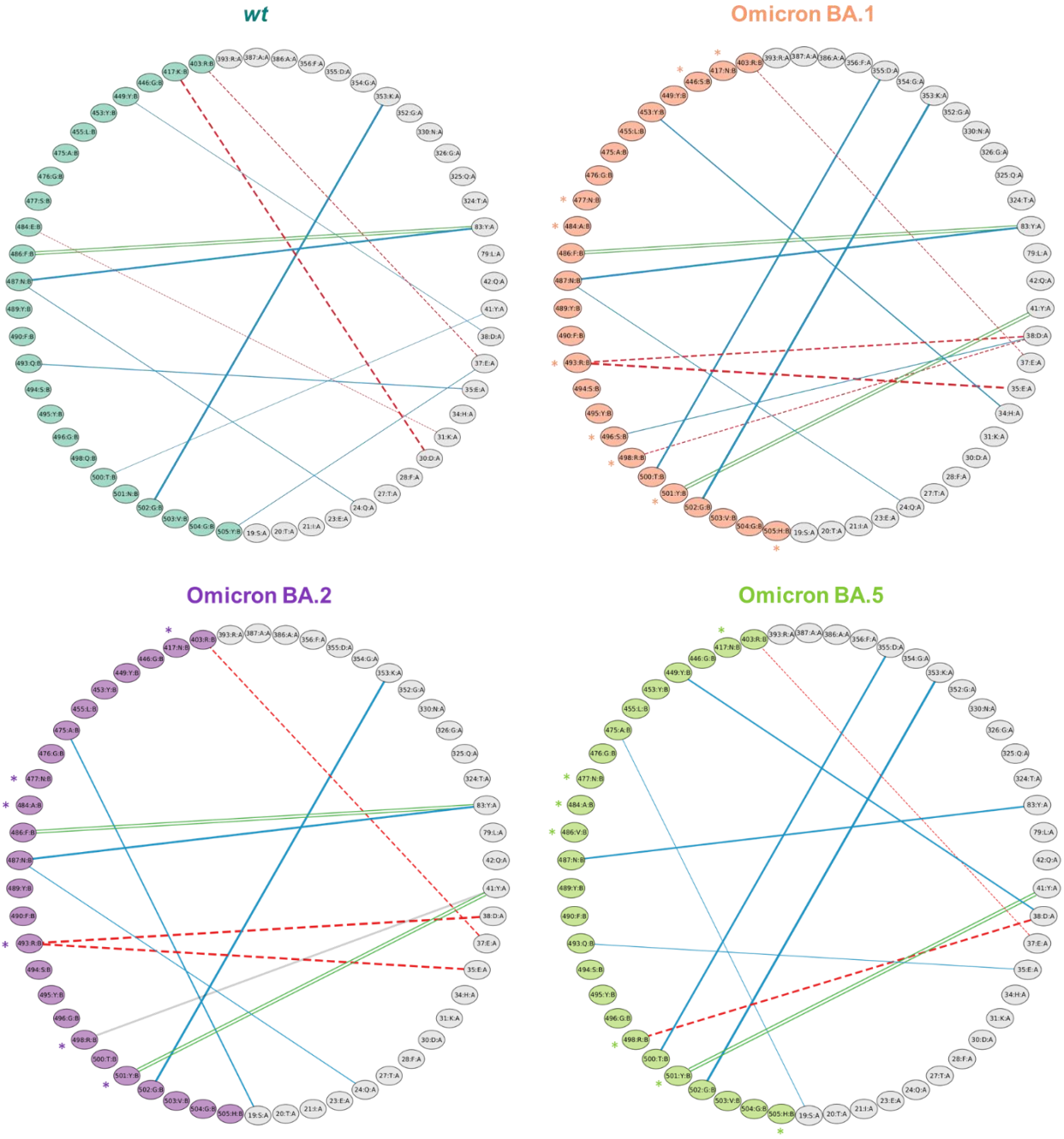


Figure S6–Residue interaction networks (RINs) for the RBD-hACE2 complexes in the last 500 ns of each simulation over 5 replicates. The intermolecular interactions between *wt*, BA.1, BA.2, and BA.5 RBDs, and hACE2 are represented: hydrogen bonds, salt bridges, pi-pi interactions and cation-pi, in blue, red, green and black, respectively. The residues of each RBD are colored by variant, in which wt RBD residues are represented in blue, BA.1 RBD residues in coral, BA.2 RBD residues in purple and BA.5 residues in green. The hACE2 residues are colored in gray. The RBD substitutions are represented with *****. The selected interactions are those that were present in over 30% of the simulation frames, and the width of each edge represents the frequency of interactions.

3.1 Detailed pairwise residue Interactions – wt RBD-hACE2 complex

Table S1– Description of the pairwise residue interactions between hACE2 and wt RBD of the last 500 ns of the simulations over 5 replicates, obtained using RIP-MD software. The type of interaction, the interacting residues, and the respective percentage of simulation frames are represented.

Interaction type	Pairwise residues		% Simulation frames
	hACE2 residue (Atom group)	wt RBD residue (Atom group)	
Hydrogen Bonds	K353(HZ2)	G496(O)	10.383
	S19(HG)	A475(O)	11.615
	A386(O)	Y505(HH)	12.934
	H34(NE2)	Y453(HH)	13.638
	H34(O)	Q493(2HE2)	15.31
	Q42(1HE2)	G446(O)	15.574
	Q24(2HE2)	A475(O)	17.422
	K353(HZ1)	Y495(O)	18.082
	Q42(2HE2)	Y449(OH)	18.522
	D38(OD2)	Q498(1HE2)	19.886
	D355(OD1)	T500(HG1)	21.117
	K31(HZ3)	Q493(OE1)	21.865
	K353(HZ1)	Q498(OE1)	23.449
	Y41(OH)	N501(2HD2)	26.617
	Y41(OH)	T500(HG1)	31.544
	D38(OD2)	Y449(HH)	34.404
	E37(OE1)	Y505(HH)	39.287
	Q24(OE1)	N487(1HD2)	42.367
	E35(OE1)	Q493(1HE2)	49.186
	Y83(HH)	N487(OD1)	92.213
	K353(O)	G502(H)	99.912
Salt Bridges	K31(NZ)	E484(OE2)	31.852
	E37(OE1)	R403(NH1)	48.658
	D30(OD1)	K417(NZ)	83.986
Pi-Pi	Y83(ring)	F486(ring)	88.869

3.2 Detailed pairwise residue Interactions – BA.1 RBD-hACE2 complex

Table S2 - Description of the pairwise residue interactions between hACE2 and Omicron BA.1 RBD of the last 500 ns of the simulations over 5 replicates, obtained using RIP-MD software. The type of interaction, the interacting residues, and the respective percentage of simulation frames are represented. The BA.1 RBD substitutions are represented with *.

Interaction type	Pairwise residues		% Simulation frames
	hACE2 residue (Atom group)	BA.1 RBD residue (Atom group)	
Hydrogen Bonds	Q42(OE1)	*R498(1HH2)	10.075
	H34(HE2)	S494(O)	11.791
	Q24(2HE2)	A475(O)	14.386
	S19(O)	*N477(2HD2)	14.386
	D38(OD2)	Y449(HH)	22.657
	Y41(OH)	T500(HG1)	27.013
	S19(HG)	A475(O)	28.817
	Q24(OE1)	N487(2HD2)	42.543
	D38(OD2)	*S496(HG)	44.391
	H34(ND1)	Y453(HH)	63.0
	D355(OD2)	T500(HG1)	87.637
	Y83(HH)	N487(OD1)	90.673
	K353(O)	G502(H)	100.0
Salt Bridges	E37(OE2)	*R493(NH1)	11.263
	E37(OE1)	R403(NH1)	40.607
	D38(OD1)	*R498(NH1)	48.394
	D38(OD1)	*R493(NH2)	71.139
	E35(OE1)	*R493(NH1)	92.477
Pi-Pi	Y83 (ring)	F486(ring)	84.778
	Y41(ring)	*Y501(ring)	94.941

3.3 Detailed pairwise residue Interactions – BA.2 RBD-hACE2 complex

Table S3- Description of the pairwise residue interactions between hACE2 and Omicron BA.2 RBD of the last 500 ns of the simulations over 5 replicates, obtained using RIP-MD software. The type of interaction, the interacting residues, and the respective percentage of simulation frames are represented. The BA.2 RBD substitutions are represented with *.

Interaction type	Pairwise residues		% Simulation frames
	hACE2 residue (Atom group)	BA.2 RBD residue (Atom group)	
Hydrogen Bonds	A386(O)	*H505(HE2)	13.682
	Q42(OE1)	*R498(2HH2)	14.078
	S19(H2)	*N477(OD1)	19.358
	D355(OD1)	T500(HG1)	22.305
	H34(ND1)	Y453(HH)	23.141
	Y41(OH)	T500(HG1)	23.273
	Q24(OE1)	N487(2HD2)	45.623
	S19(HG)	A475(O)	68.016
	Y83(HH)	N487(OD1)	91.509
	K353(O)	G502(H)	100.0
Salt Bridges	D38(OD2)	*R498(NH1)	10.691
	E37(OE2)	*R493(NH1)	19.182
	E37(OE1)	R403(NH1)	65.2
	D38(OD1)	*R493(NH2)	99.824
	E35(OE1)	*R493(NH1)	95.425
Pi-Pi	F28(ring)	F486(ring)	10.515
	Y83(ring)	F486(ring)	88.605
	Y41(ring)	*Y501(ring)	93.093

3.4 Detailed pairwise residue Interactions – BA.5 *RBD*-hACE2 complex

Table S4- Description of the pairwise residue interactions between hACE2 and Omicron BA.5 RBD of the last 500 ns of the simulations over 5 replicates, obtained using RIP-MD software. The type of interaction, the interacting residues, and the respective percentage of simulation frames are represented. The BA.5 RBD substitutions are represented with *.

Interaction type	Pairwise residues		% Simulation frames
	hACE2 residue (Atom group)	BA.5 RBD residue (Atom group)	
Hydrogen Bonds	S19(H2)	*N477(OD1)	10.647
	H34(ND1)	Q493(2HE2)	11.439
	K31(HZ3)	Q493(OE1)	12.934
	Q24(OE1)	N487(2HD2)	17.202
	Q24(2HE2)	A475(O)	18.082
	Y41(OH)	T500(HG1)	22.437
	H34(HE2)	S494(O)	24.241
	Q42(NE2)	*R498(2HH2)	24.593
	S19(HG)	A475(O)	35.416
	E35(OE1)	Q493(1HE2)	38.143
	D38(OD2)	Y449(HH)	71.271
	Y83(HH)	N487(OD1)	77.079
	D355(OD2)	T500(HG1)	88.297
	K353(O)	G502(H)	99.956
Salt Bridges	D38(OD1)	*R498(NH1)	36.912
	E37(OE1)	R403(NH1)	86.626
Pi-Pi interactions	Y41(ring)	*Y501(ring)	95.996

4. Average values of contact surfaces – RBD-hACE2 complexes

Table S5 – Average values of the total, hydrophilic and hydrophobic surfaces (nm^2) of each RBD-hACE2 complex for the last 500 ns represented in figure 3.13.

RBD-hACE2 Complex	Contact Surface (nm^2)	Hydrophilic Surface Area (nm^2)	Hydrophobic Surface Area (nm^2)
wt	10.18±0.12	5.29±0.19	4.89±0.11
Omicron BA.1	9.82±0.14	4.81±0.12	5.01±0.07
Omicron BA.2	10.01±0.13	5.13±0.04	4.88±0.11
Omicron BA.5	9.45±0.27	4.95±0.20	4.50±0.10

5. PCA Basins of Omicron BA.2 and BA.5 RBDs

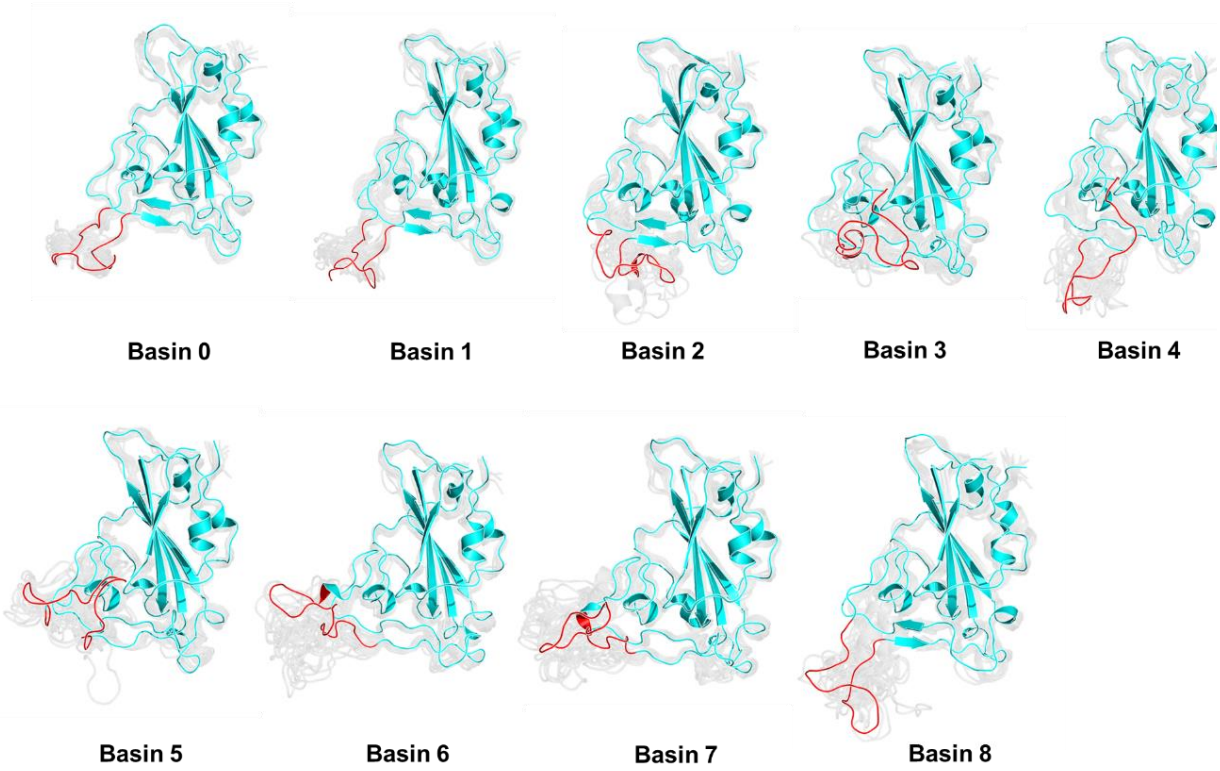


Figure S7 – Representation of all BA.2 RBD PCA basins. The structures corresponding to the free energy minima of all conformational basins are represented in blue, with the "ridge" region highlighted in red. The structures sampled from the same basin are in represented in the background as gray cartoon.

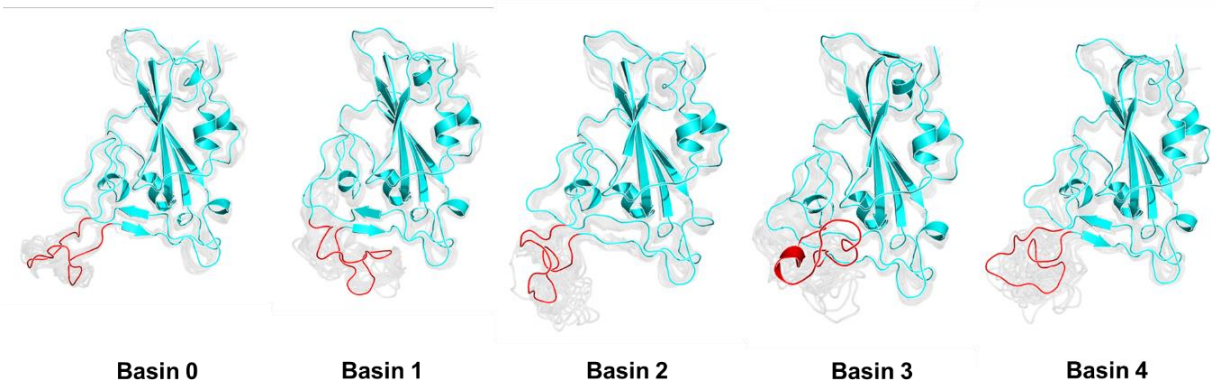


Figure S8 – Representation of all BA.5 RBD PCA basins. The structures corresponding to the free energy minima of all conformational basins are represented in blue, with the "ridge" region highlighted in red. The structures sampled from the same basin are in represented in the background as gray cartoon.

Table S6 - Energy surface landscape analysis from 2D PCA of RBD conformational dynamics in water. Energy surface landscape analysis and defined basins for each of the tested RBD variants. Energy minima, frame percentage and loop conformation for each of the basins is also given.

System	Basin	Free Energy	<E>/k_BT	S/R	Emin/k_BT	Frame Percentage %	Loop Conformation	"Open" state (%)	"Closed" state (%)
Omicron BA.2	0	-5.18	0.76	5.94	-0.00	36.25	"Open"	68.79	23.31
	1	-5.04	0.50	5.55	0.03	31.51	"Open"		
	2	-4.72	2.72	7.44	1.35	22.88	"Closed"		
	3	-3.65	2.45	6.11	1.66	7.94	"Shrunken"		
	4	-0.51	4.78	5.29	4.37	0.34	"Closed"		
	5	0.23	5.79	5.55	5.57	0.33	"Open"		
	6	0.85	5.78	4.93	5.61	0.16	"Open"		
	7	0.92	5.80	4.88	5.61	0.13	"Open"		
	8	1.01	4.66	3.65	4.55	0.07	"Open"		
					Total	99,61			
Omicron BA.5	0	-5.13	01.01	6.14	-0.00	77.84	"Open"	81.78	17.59
	1	-3.59	03.09	6.68	2.05	16.92	"Closed"		
	2	-1.88	4.10	5.98	3.42	3.14	"Open"		
	3	-0.37	5.37	5.74	4.87	1.26	"Closed"		
	4	-0.31	4.68	4.99	4.32	0.64	"Open"		
					Total	99,80			

6.1 Residue Interaction Networks (RIN): Omicron BA.2 RBD

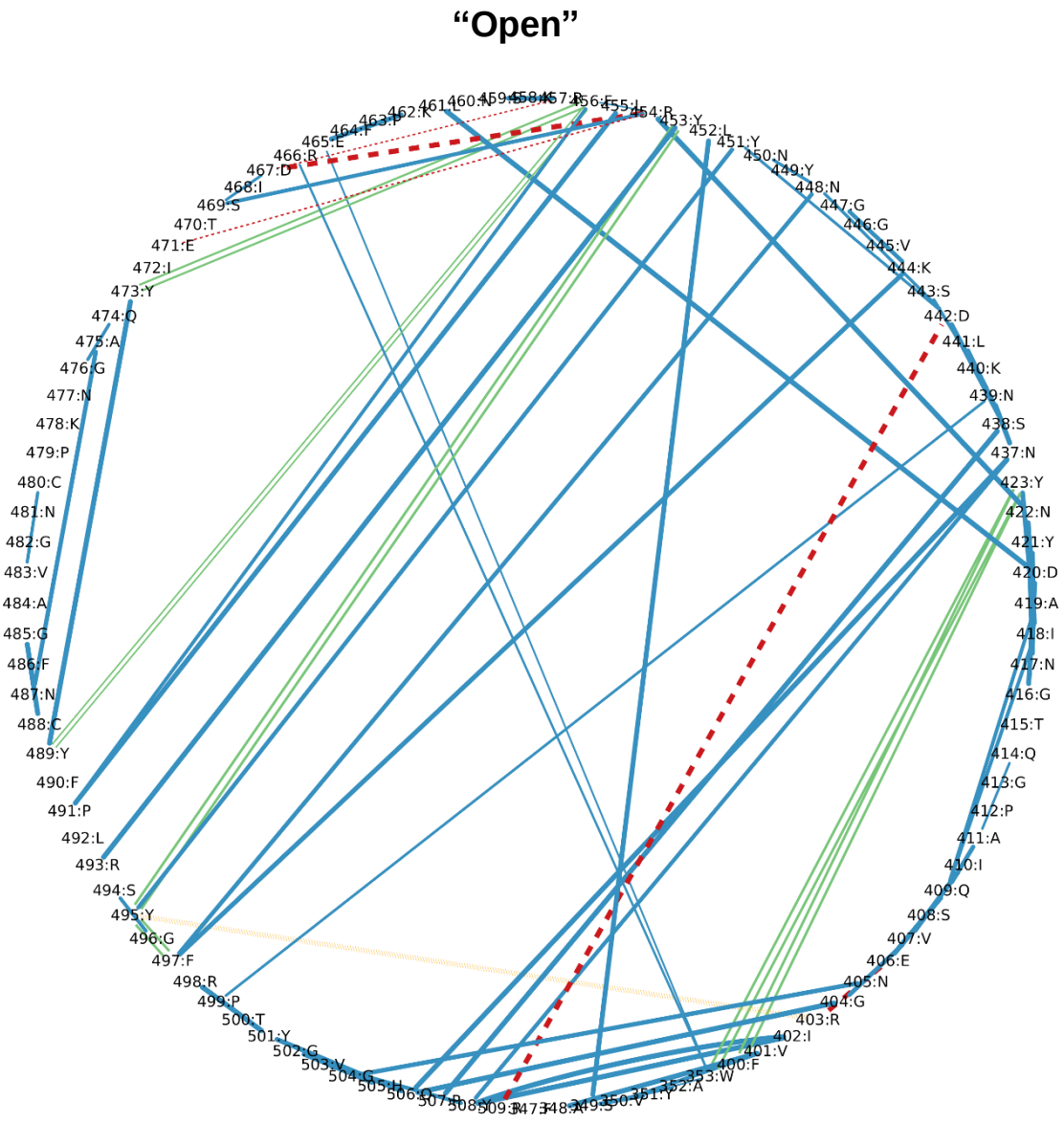


Figure S9 - Residue interaction networks (RINs) for the “shrunken” conformations of Omicron BA.2 RBD. RINs determined using RIP-MD for the 5000 lowest energy conformations obtained for basins 0 and 1 of 2D-PCA (Figure **3.17C**). Hydrogen bonds, salt bridges, and pi-pi interactions are shown in blue, red, and green, respectively.

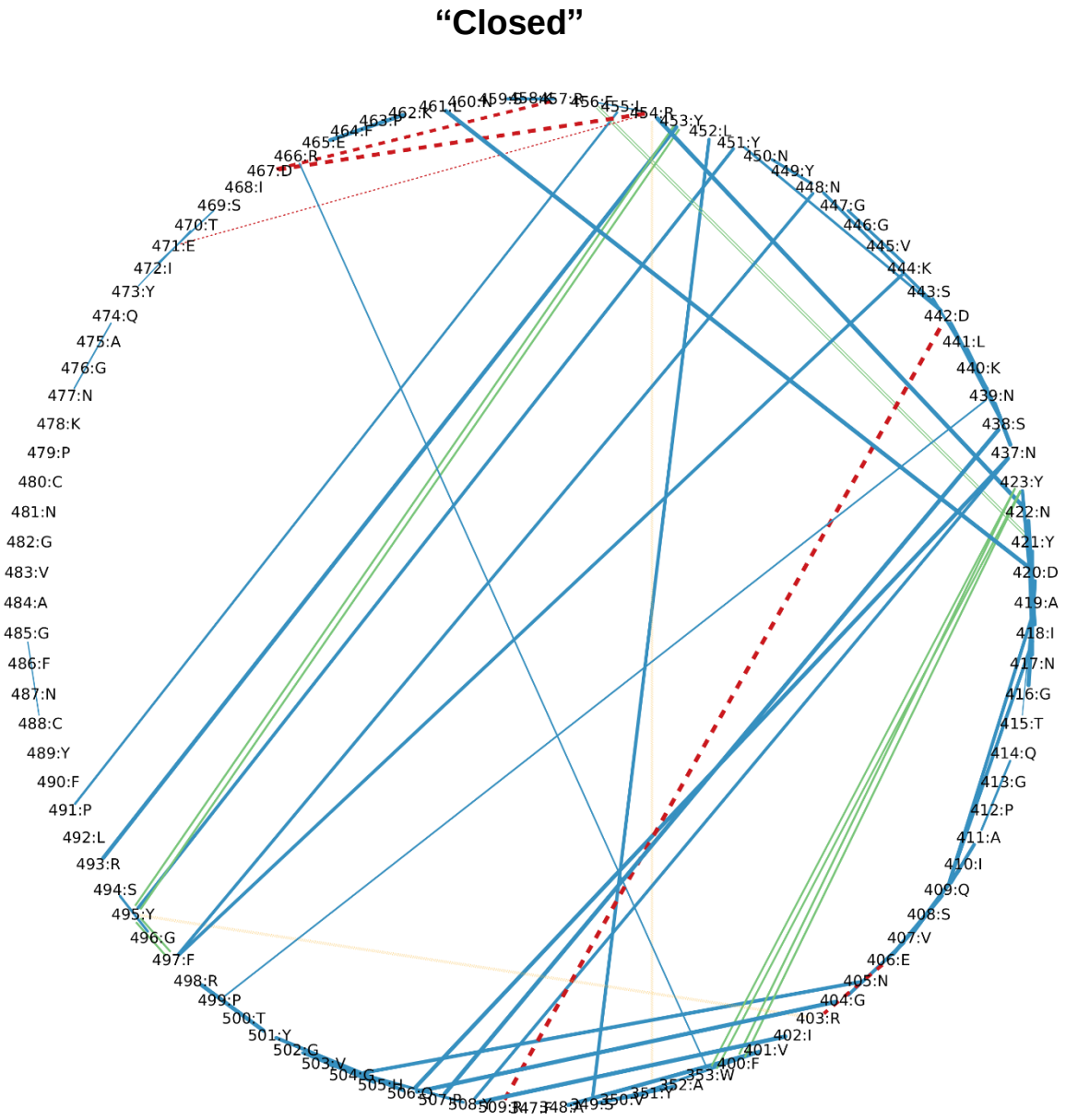


Figure S10 - Residue interaction networks (RINs) for the “shrunken” conformations of Omicron BA.2 RBD. RINs determined using RIP-MD for the 5000 lowest energy conformations obtained for basin 2 of 2D-PCA (Figure **3.17C**). Hydrogen bonds, salt bridges, and pi-pi interactions are shown in blue, red, and green, respectively.

“Shrunken”

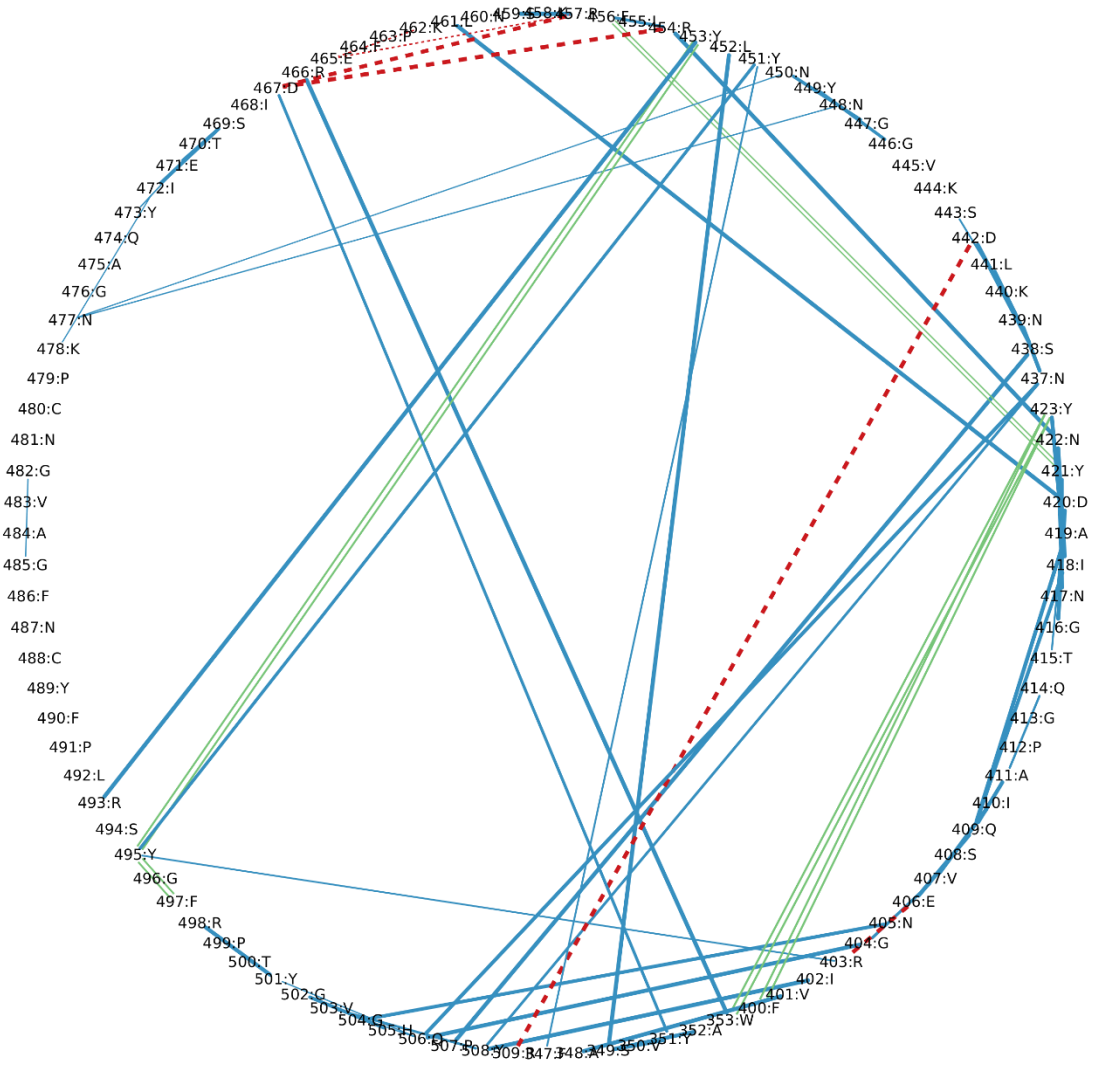


Figure S11 - Residue interaction networks (RINs) for the “shrunken” conformations of Omicron BA.2 RBD. RINs determined using RIP-MD for the 5000 lowest energy conformations obtained for basin 3 of 2D-PCA (Figure **3.17C**). Hydrogen bonds, salt bridges, and pi-pi interactions are shown in blue, red, and green, respectively.

6.2 Residue Interaction Networks (RIN): Omicron BA.5 RBD

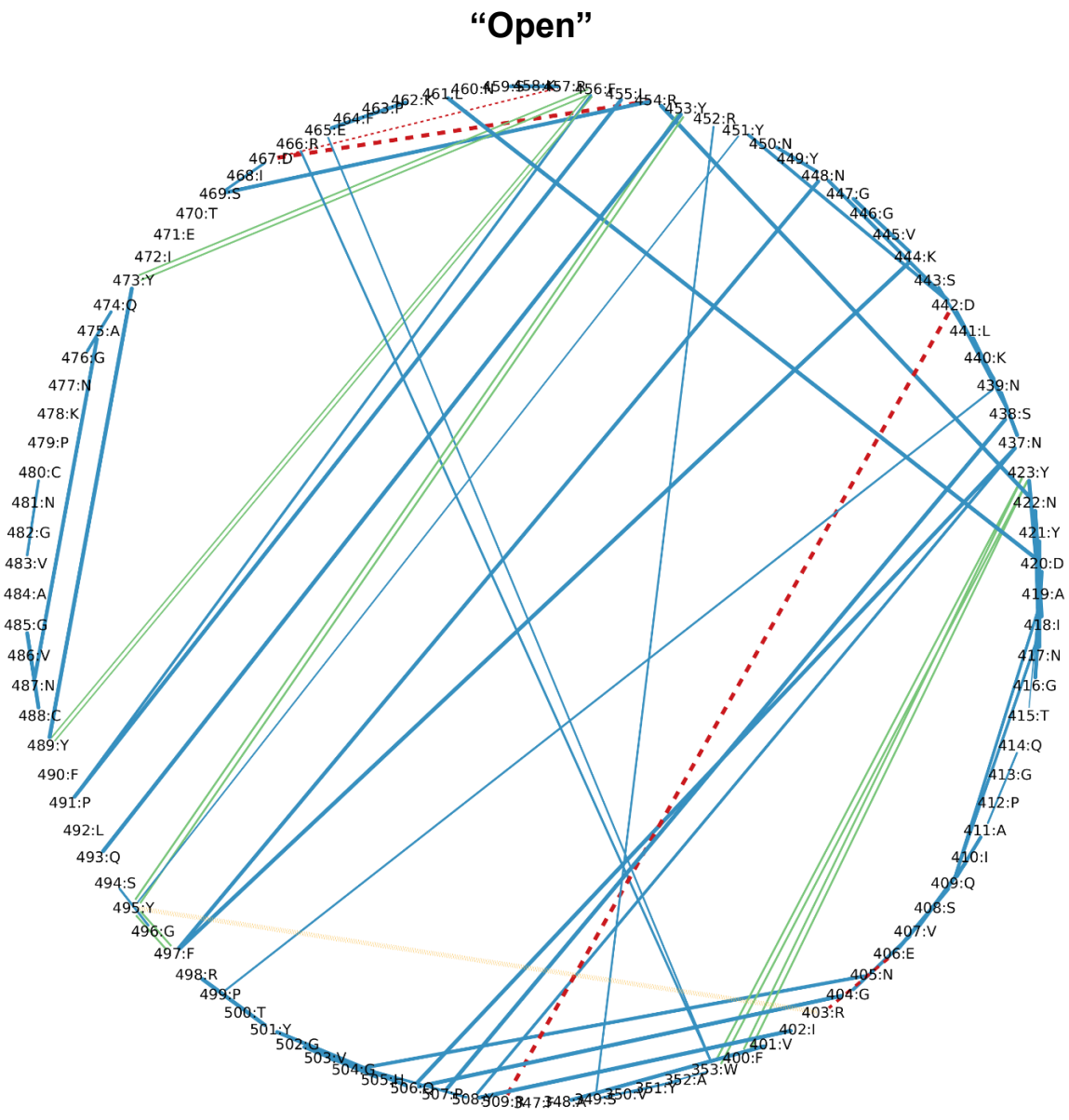


Figure S12 - Residue interaction networks (RINs) for the “open” and “closed” conformations of Omicron BA.5 RBD. RINs determined using RIP-MD for the 5000 lowest energy conformations obtained for the most populated basins. Hydrogen bonds, salt bridges, and pi-pi interactions are shown in blue, red, and green, respectively.

“Closed”

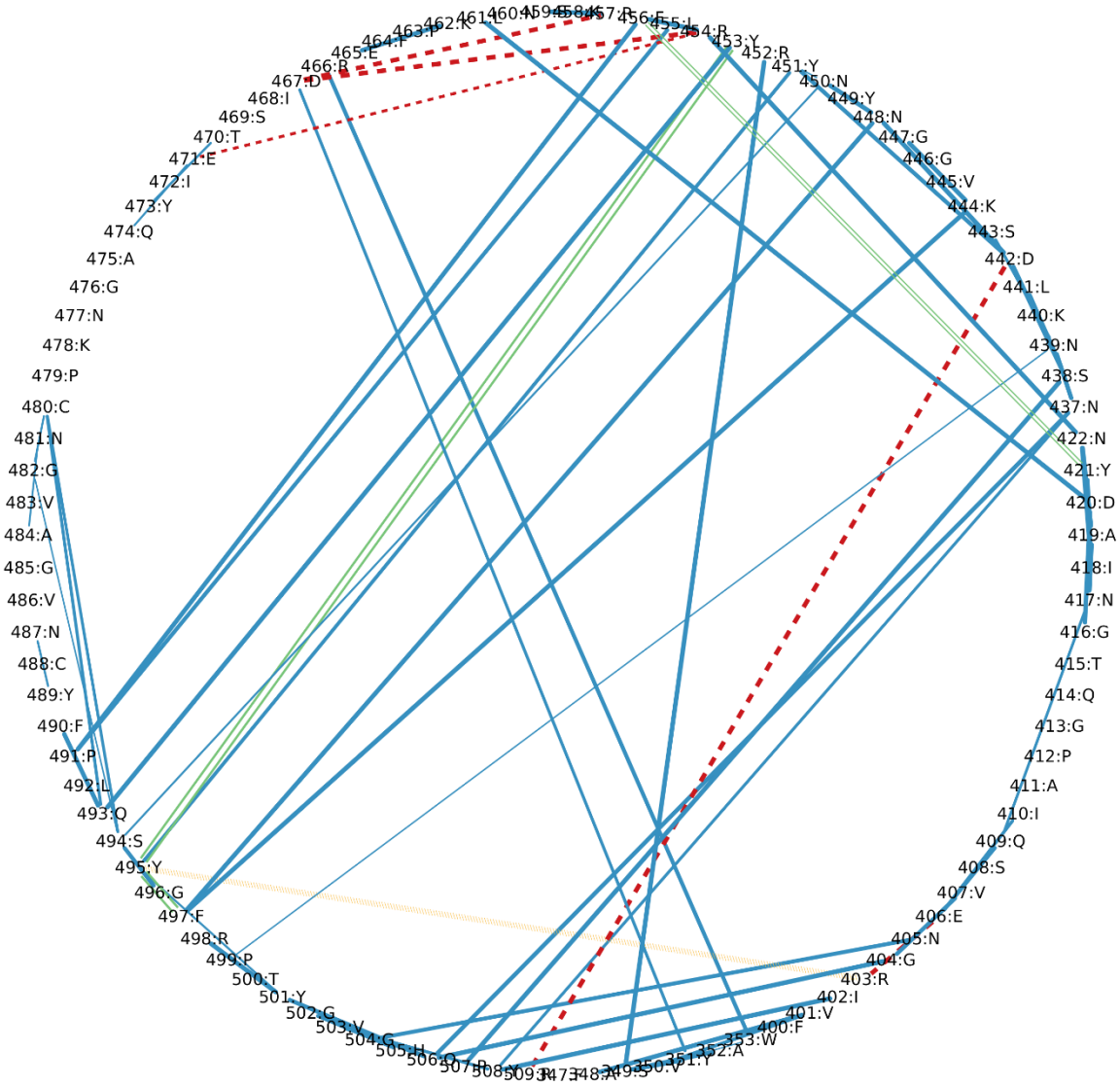


Figure S13 - Residue interaction networks (RINs) for the “open” and “closed” conformations of Omicron BA.5 RBD. RINs determined using RIP-MD for the 5000 lowest energy conformations obtained for the most populated basins. Hydrogen bonds, salt bridges, and pi-pi interactions are shown in blue, red, and green, respectively.