

**Impact of mutations in non-structural proteins
on SARS-CoV-2 replication**

Eugenia Afi Datsomor

Thesis submitted to the faculty of the Virginia Polytechnic Institute and State University in
partial fulfillment of the requirements for the degree of

Master of Science

Biological Sciences

Daniel Capelluto, Chair

Carla Finkielstein

William Huckle

Florian Schubot

May 7th, 2024
Blacksburg, Virginia

Abstract

The late 2019 marked the onset of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) that led to the unprecedented COVID-19 pandemic, with profound global health and socioeconomic impacts. This thesis offers a thorough examination of the molecular biology, evolution, and disease-causing mechanisms of SARS-CoV-2, as well as recent advancements in understanding the structural and functional implications of mutations in viral proteins.

The prevailing belief is that SARS-CoV-2 originated from a zoonotic transmission involving bats as the natural reservoir hosts, with an unknown intermediate host facilitating transmission to humans. Genomic sequencing and phylogenetic analysis have identified similarities between SARS-CoV-2 and bat coronaviruses, particularly RaTG13, indicating a potential bat origin. However, the exact circumstances and intermediate hosts of the spillover event remain under investigation.

In its structure, SARS-CoV-2 is an enveloped virus with a positive-sense single-stranded RNA genome. This genome encodes both structural and non-structural proteins crucial for viral replication and the development of the disease. The spike (S) protein facilitates viral entry by binding to the angiotensin-converting enzyme 2 (ACE2) receptor. Meanwhile, non-structural proteins are involved in viral RNA synthesis, immune evasion, and the assembly of virions. Alterations in the genetic makeup of the SARS-CoV-2 genome, notably within the spike protein, can impact transmission efficiency, viral load, and immune evasion. Notable mutations such as D614G, N501Y, and E484K have been associated with increased transmissibility and reduced neutralization by antibodies. Understanding the effects of these mutations on viral fitness and pathogenicity is crucial for informing public health interventions and vaccine development efforts. The impacts of Non-structural proteins (NSPs) on viral replication and transmission are however understudied.

In this study, we focused on mutations in the several NSPs including NSP1, 2, 3, 13,14, and 15 of the early Omicron (BA.1) and XBB 1.5 variants and investigated their impact on structure and the functional implications using bioinformatics tools and protein structure prediction methods. Our analysis focused on potential alterations in NSP1's structure and hence its ability to suppress host gene expression and modulate immune responses, shedding light on the mechanisms by which SARS-CoV-2 evolves to evade host defenses.

Overall, this thesis gives insights into the emergence, structure, replication cycle, evolution, and pathogenesis of SARS-CoV-2, highlighting the importance of ongoing research efforts in understanding and combatting this global health threat and provides a detailed structural analysis of mutations in NSPs.

Keywords: SARS-CoV-2, Non-structural proteins, Replication, Mutation, Homology Modeling

General Audience Abstract

The COVID-19 pandemic, instigated by the virus referred to as SARS-CoV-2, is a novel coronavirus believed to have originated in bats and possibly transmitted to humans via an intermediate host. Its genetic structure and protein interactions play crucial roles in how it spreads and causes illness. We need to understand where the virus came from, how it's built, its life cycle and how it's changing over time.

While the virus has undergone a lot of mutations over time, scientists are actively studying these changes, with a lot of focus on the structural ones, to understand their implications for public health measures and vaccine development. In our study, we focus on the non-structural proteins and aim to investigate the effect of selected mutations on the protein structure and function using bioinformatics. Understanding the virus is essential for effectively combating future pandemics and safeguarding public health.

Table of Contents

Acknowledgements.....	1
Abbreviations.....	2
CHAPTER ONE.....	3
Introduction.....	3
1.1 Background.....	3
1.1.1 SARS-CoV-2 emergence and origins.....	3
1.1.2 Classification and structure.....	3
1.1.3 Public health significance.....	4
1.2 Molecular Biology and Characterization of SARS-CoV-2.....	5
1.2.1 Genomic functional domains.....	5
1.2.2 Cell entry mechanisms.....	6
1.2.3 Intracellular replication machinery.....	7
1.2.4 Virion assembly and release.....	8
1.3 Viral components supporting replication cycle.....	10
1.3.1 Structural proteins.....	10
1.3.2 Non-structural proteins (NSPs).....	10
1.3.3 Functions of key NSPs.....	10
1.4 SARS-CoV-2 Evolution and Pathogenesis.....	13
1.4.1 Mutation rates and types.....	13
1.4.2 Immune evasion mechanisms.....	15
1.4.3 Determinants of clinical severity.....	16
1.4.4 Implications for diagnostics, interventions, therapeutics, vaccines.....	17
CHAPTER TWO.....	19
Research Question and Hypothesis.....	19
CHAPTER THREE.....	21
Research Methodology.....	21
3.1 Sequence Analysis.....	21
3.2 Comparative Analysis.....	22
3.3 Structure Prediction and Visualization.....	23
3.4 Look for changes in enzymatic activity, substrate binding, or other relevant functional aspects.....	27
3.5 Experimental Procedure.....	28
3.5.1 Sequence Alignment of XBB and early omicron variants.....	29
3.5.2 Structure Modeling of SARS-CoV-2 NSP1 and I-TASSER Accuracy Assessment.....	30
Results and Discussion.....	32

4.1 Homology Modeling NSP1.....	32
4.1.1 Mutation Impact: Glycine to Serine (G49S):.....	33
4.1.2 Mutation Impact: Lysine to Arginine (K47R):.....	35
4.2 Homology Modeling of NSP2	36
4.3 Homology Modeling of NSP13	39
4.3.1 Mutation Impact: Serine to Proline Mutation (S36P).....	42
4.4 Homology Modeling of NSP14	42
4.5 Homology Modeling of NSP15	46
Conclusions and Recommendations	52
References	56

Acknowledgements

I am immensely grateful to my advisor, Dr. Daniel Capelluto, whose unwavering support has been a source of strength throughout my challenging journey in graduate education. His patience, encouragement, and compassion have been inspiration for me. I could not have imagined having a better advisor and mentor.

I would also like to extend my heartfelt gratitude to Dr. Daniella Cimini for her consistent support, both within and outside the university. Her unwavering presence and assistance have been invaluable to me, especially during times of need.

I am deeply grateful to the other members of my committee, Dr. Carla Finkelstein, Dr. William Huckle, and Dr. Florian Schubot, for their invaluable insights and suggestions on my research project. Their patience and guidance have been instrumental throughout this journey. I particularly want to thank Dr. Finkelstein for her exceptional advice, guidance, and for entrusting me with her project.

I would like to say a big thank you to Rebecca Zimmerman and Brent Bowden. I cannot count the number of times they have been there for me and the number of things they have done for me. Thank you so much for everything!

Abbreviations

ACE2	Angiotensin-converting enzyme 2
CTD	C-terminal domain
COVID-19	Coronavirus disease 2019
DMVs	Double-membrane vesicles
ERGIC	ER-Golgi intermediate compartment
EndoU	Endoribonuclease
ExoN	Exoribonuclease
NTD	N-terminal domain
NSP	Non-structural protein
ORF	Open Reading Frame
PLpro	Papain-like protease
RdRp	RNA-dependent RNA polymerase
RNA	Ribonucleic acid
RMSD	Root-mean-square deviation
RTC	Replication-transcription complex
SARS	Severe acute respiratory syndrome
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
SPs	Structural proteins
UTRs	Untranslated regions
VOCs	Variants of concern

CHAPTER ONE

Introduction

1.1 Background

1.1.1 SARS-CoV-2 emergence and origins

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the etiological agent of the COVID-19 pandemic, originated in Wuhan, China in late 2019. The initial reported cases were associated with the Huanan Seafood Wholesale Market, indicating a potential zoonotic origin (Zhu et al., 2020). Genomic sequencing and phylogenetic analysis revealed that SARS-CoV-2 shares a significant sequence similarity with bat coronaviruses, particularly RaTG13, indicating that bats may serve as the natural reservoir host (Zhou et al., 2020). However, the intermediate animal host that facilitated the transmission to humans remains unknown. Pangolins, minks, and other mammals have been proposed as potential intermediate hosts due to the presence of related coronaviruses (Lam et al., 2020; Oude Munnink et al., 2021). The precise details of the spillover event and the particular ecological interactions that led to the emergence of SARS-CoV-2 are still being investigated.

1.1.2 Classification and structure

SARS-CoV-2 is classified within the genus Beta coronavirus, which is part of the Coronaviridae family. SARS-CoV-2 is characterized as an enveloped virus with a positive-sense single-stranded RNA genome. Its genome size is approximately 30 kilobases (Wu et al., 2020). The viral particle is spherical and measures about 100-160 nm in diameter. The genome of SARS-CoV-2 contains at least 10 open reading frames (ORFs) that encode both structural (SPs) and non-structural proteins (NSPs). The structural proteins include the spike (S) glycoprotein, membrane (M) protein, envelope (E) protein, and nucleocapsid (N) protein. The S protein forms homotrimeric spikes on the viral surface and mediates receptor binding and membrane fusion. The M and E proteins play roles in virus assembly and budding, whereas the

N protein binds to the viral RNA genome, forming the ribonucleoprotein complex (Fehr & Perlman, 2015). NSPs are encoded by ORF1a and ORF1b and are involved in various aspects of the viral life cycle, including replication, transcription, and immune evasion (V'kovski et al., 2021).

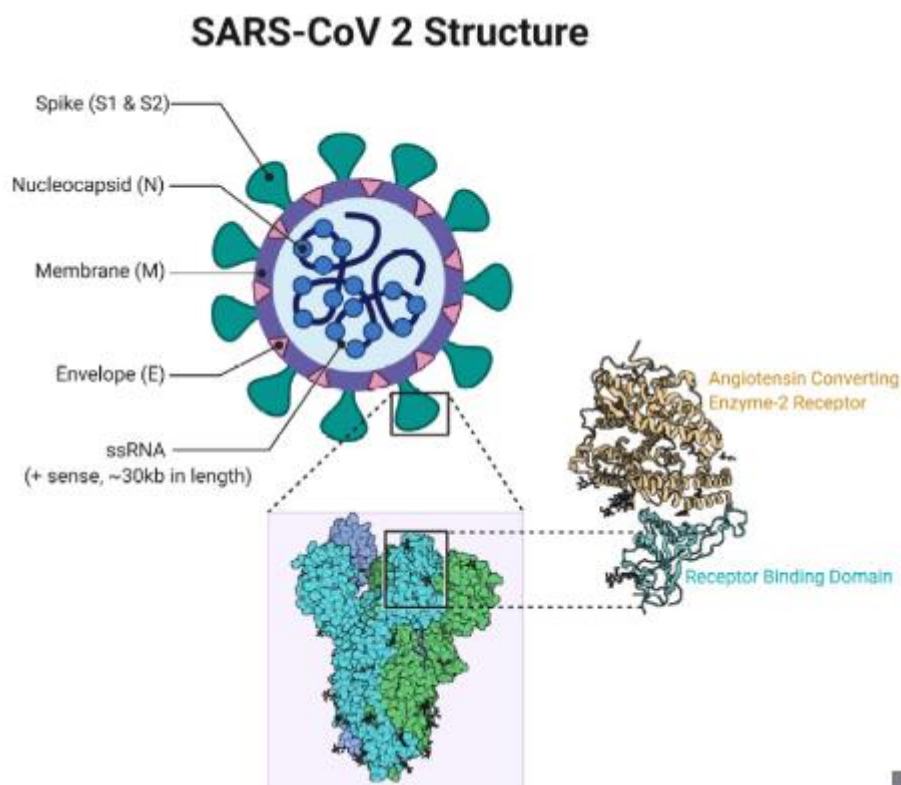


Figure 1.0: Schematic labeled diagram of SARS-CoV-2 structure.

(Figure contributed by Rohan Bir Singh, MD; Made with Biorender.com)

1.1.3 Public health significance

The emergence of SARS-CoV-2 has significantly influenced global public health. By July 2023, more than 600 million confirmed cases and 6.3 million deaths had been reported globally

(World Health Organization, 2023). The quick spread of the virus has posed unprecedented challenges for healthcare systems, economies, and societies. SARS-CoV-2 exhibits high transmissibility, with an approximated basic reproduction number (R_0) ranging from 2 to 4, indicating that each infected individual can infect 2-4 others in a susceptible population (Liu et al., 2020). The primary mode of transmission for the virus is through respiratory droplets and aerosols, and asymptomatic or presymptomatic transmission has been documented (Moghadas et al., 2020). The clinical spectrum of COVID-19 ranges from mild respiratory symptoms to severe pneumonia, acute respiratory distress syndrome (ARDS), and multi-organ failure. Elderly individuals, those with underlying health conditions, and immunocompromised patients face an increased risk of severe illness and mortality from COVID-19 (Wu & McGoogan, 2020). The pandemic has also exacerbated existing health disparities and disproportionately affected vulnerable populations (Bambra et al., 2020). The global focus has been on developing effective vaccines and therapeutics to contain the spread of SARS-CoV-2 and mitigate its impact on public health.

1.2 Molecular Biology and Characterization of SARS-CoV-2

1.2.1 Genomic functional domains

The SARS-CoV-2 genome is a single-stranded, positive-sense RNA of approximately 30 kilobases in length. It contains several functional domains that encode both structural and non-structural proteins (NSPs) essential for viral replication and pathogenesis (Wu et al., 2020).

The 5' two-thirds of the genome consists of two large open reading frames (ORFs), ORF1a and ORF1b, which encode the polyproteins pp1a and pp1ab. These polyproteins are cleaved by viral proteases, NSP3 (papain-like protease, PLpro) and NSP5 (main protease, Mpro), to produce 16 NSPs (NSP1-16) that form the replication-transcription complex (RTC) (V'kovski et al., 2021).

The 3' one-third of the genome contains ORFs encoding structural proteins, including S, E, M, and N proteins, as well as several accessory proteins (ORF3a, ORF6, ORF7a, ORF7b, ORF8, and ORF10) (Kim et al., 2020). These accessory proteins are believed to play roles in viral pathogenesis and immune evasion, although their precise functions are not yet fully understood (Yoshimoto, 2020).

The genome also contains untranslated regions (UTRs) at both the 5' and 3' ends, which are crucial for viral RNA replication and translation. The 5' UTR contains the leader sequence and the transcription regulation sequence (TRS), while the 3' UTR contains a stem-loop structure and a poly(A) tail (Yang & Leibowitz, 2015).

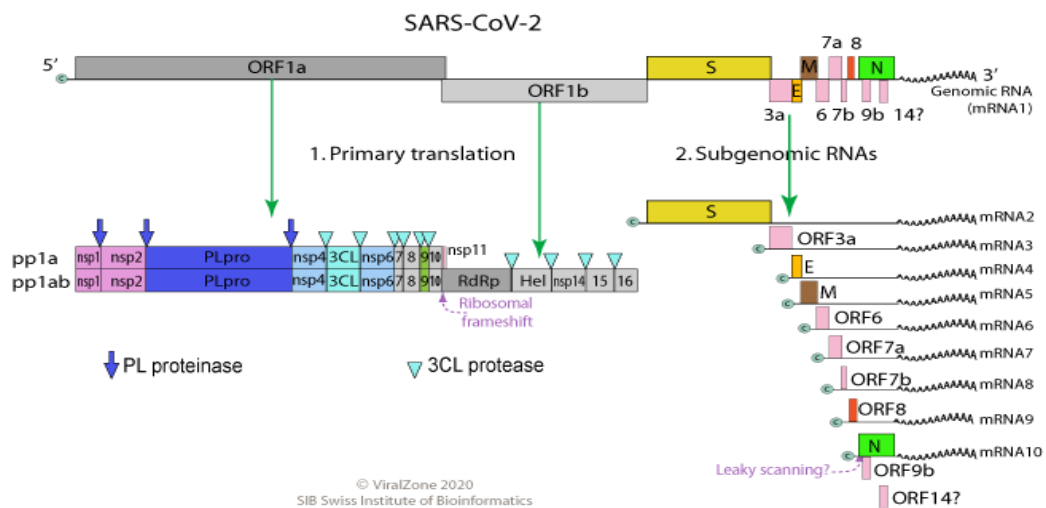


Figure 2.0: A graphical depiction of the ORFs, outlining each gene product of the ORFs.

1.2.2 Cell entry mechanisms

SARS-CoV-2 gains entry into host cells by binding the viral spike (S) protein to the cellular receptor, angiotensin-converting enzyme 2 (ACE2) (Hoffmann et al., 2020). The S protein consists of two subunits, S1 and S2. The S1 subunit contains the receptor-binding domain (RBD) that directly binds to ACE2, while the S2 subunit mediates the fusion of the viral envelope with the host cell membrane (Walls et al., 2020).

The binding of the S protein to ACE2 triggers a conformational change in the S protein, exposing the cleavage sites for host cell proteases, such as transmembrane protease serine 2 (TMPRSS2) and cathepsin L (Hoffmann et al., 2020). These proteases cleave the S protein at the S1/S2 boundary and the S2' site, leading to the release of the fusion peptide and the fusion of the viral envelope with the host cell membrane (Bestle et al., 2020).

In addition to the canonical ACE2-dependent entry pathway, SARS-CoV-2 can also enter cells through an alternative, ACE2-independent pathway involving the CD147 receptor (Wang et al., 2020). This process involves the interaction between the spike protein and CD147, leading to the cleavage of the spike protein by host cell proteases, followed by the fusion of the viral envelope with the host cell membrane.

1.2.3 Intracellular replication machinery

Upon entry into the host cell, the SARS-CoV-2 genome is released into the cytoplasm, where it serves as a template for both translation and replication. The ORF1a and ORF1b are translated to produce the polyproteins pp1a and pp1ab, which are cleaved by viral proteases to generate 16 NSPs that form the replication-transcription complex (RTC) (V'kovski et al., 2021).

The RTC is responsible for the synthesis of viral RNA, including genomic and subgenomic RNAs. The key components of the RTC include the RNA-dependent RNA polymerase (RdRp, NSP12), the helicase (NSP13), and the exonuclease (NSP14) (Gao et al., 2020; Jang et al., 2020; Ferron et al., 2018). The RdRp catalyzes the synthesis of new viral RNA, while the helicase unwinds double-stranded RNA and the exonuclease performs proofreading to maintain the fidelity of viral RNA synthesis.

The RTC is anchored to double-membrane vesicles (DMVs) derived from the endoplasmic reticulum (ER) (Wolff et al., 2020). These DMVs provide a protective environment for viral

RNA synthesis and help to evade host cell innate immune responses. The formation of DMVs is mediated by several NSPs, including NSP3, NSP4, and NSP6 (Hagemeijer et al., 2014).

In addition to the RTC, SARS-CoV-2 also encodes several other proteins that are essential for viral replication and transcription. For example, NSP15 is an endoribonuclease that cleaves viral RNA to evade host cell innate immune responses (Deng et al., 2019), while NSP16 is a 2'-O-methyltransferase that modifies the 5' cap structure of viral RNAs to mimic host cell mRNAs and avoid detection by the host cell innate immune system (Decroly et al., 2011).

1.2.4 Virion assembly and release

New SARS-CoV-2 virions are assembled and released in the ER-Golgi intermediate compartment (ERGIC) (Klein et al., 2020). The structural proteins S, E, and M are synthesized on the rough ER and transported to the ERGIC, where they interact with the N protein-encapsidated viral genome to form new virions (Siu et al., 2008).

The M protein assumes a central role in virion assembly by interacting with all other structural proteins and the viral genome (Neuman et al., 2011). The M protein also recruits the E protein to the site of virion assembly, which is essential for the formation of the viral envelope (Schoeman & Fielding, 2019).

The S protein is incorporated into the virion through its interaction with the M protein (Ujike et al., 2016). The S protein undergoes post-translational modifications, including glycosylation and proteolytic cleavage, which are critical for its function in viral entry (Walls et al., 2020).

After assembly, the newly formed virions are transported to the cell surface within smooth-walled vesicles and are then released via exocytosis (Ghosh et al., 2020). The release of new virions is facilitated by the action of the viral protein NSP3, which has been shown to disrupt the host cell secretory pathway and promote the release of virions (Reggiori et al., 2010).

In summary, the assembling and releasing new SARS-CoV-2 virions is a complex process that involves the coordinated action of both structural and non-structural viral proteins, as well as host cell factors. Understanding the molecular mechanisms underlying virion assembly and release may provide new targets for antiviral therapies and vaccines.

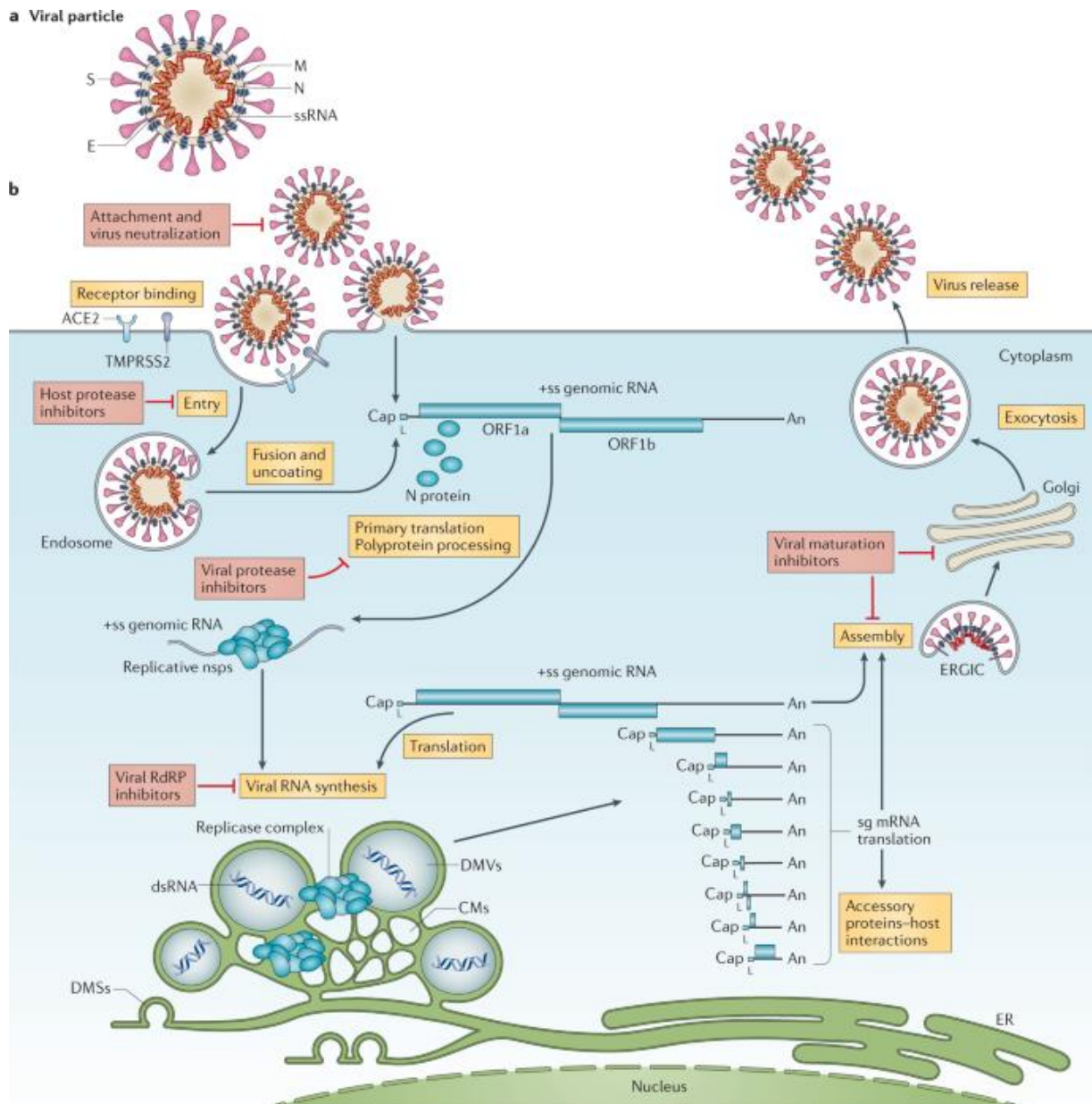


Figure 3.0 A picture detailing the molecular replication process of SARS-CoV-2. (a) shows the virus just before attachment to the ACE2 receptors while (b) depicts the cell entry mechanisms for replication and exocytosis of the virus.

(Figure contributed by V'kovski et al. <https://doi.org/10.1038/s41579-020-00468-6>)

1.3 Viral components supporting replication cycle

1.3.1 Structural proteins

The structural proteins of SARS-CoV-2 comprise the spike (S), envelope (E), membrane (M), and nucleocapsid (N) proteins. The S protein is a glycoprotein that mediates viral entry into host cells by binding to the angiotensin-converting enzyme 2 (ACE2) receptor (Hoffmann et al., 2020). The E and M proteins are involved in virus assembly and budding, while the N protein binds to the viral RNA genome, forming the ribonucleoprotein complex (Fehr & Perlman, 2015).

1.3.2 Non-structural proteins (NSPs)

SARS-CoV-2 encodes 16 non-structural proteins (NSP1-16) that are involved in various aspects of the viral life cycle, including replication, transcription, and immune evasion. These NSPs are encoded by the ORFs 1a and 1b, which occupy approximately two-thirds of the viral genome (V'kovski et al., 2021). The focus of this section will be on NSPs 1, 2, 3, 4, 6, 9, 13, 14, and 15.

1.3.3 Functions of key NSPs

NSP1: NSP1 is a potent suppressor of host gene expression and interferon (IFN) response. It binds to the 40S ribosomal subunit and blocks the mRNA entry channel, leading to translation inhibition (Thoms et al., 2020). The amino acid sequence of NSP1 is highly conserved among beta coronaviruses, with a length of 180 amino acids in SARS-CoV-2 (UniProt ID: P0DTD1). The C-terminal region of NSP1, particularly residues K164 and H165, is crucial for its binding to the 40S ribosomal subunit and its ability to inhibit host translation (Schubert et al., 2020).

NSP2: The function of NSP2 is not characterized well, but it is understood to play a role in viral RNA synthesis and modulating host cell survival pathways (Cornillez-Ty et al., 2009).

NSP2 of SARS-CoV-2 consists of 638 amino acids (UniProt ID: P0DTD1).

NSP3: NSP3 is a large, multifunctional protein with several domains, including a papain-like protease (PLpro) that cleaves the viral polyprotein and deubiquitinates host proteins, thereby inhibiting the host innate immune response (Shin et al., 2020). The SARS-CoV-2 NSP3 is 1,945 amino acids long (UniProt ID: P0DTD1). The PLpro domain of NSP3 contains a catalytic triad consisting of residues C111, H272, and D286, which are essential for its proteolytic and deubiquitinating activities (Shin et al., 2020).

NSP4: NSP4 is a transmembrane protein that plays a crucial role in the formation of double-membrane vesicles (DMVs), which serve as the site for viral RNA replication and transcription (Hagemeijer et al., 2014). The SARS-CoV-2 NSP4 consists of 500 amino acids (UniProt ID: P0DTD1).

NSP6: Like NSP4, NSP6 is a transmembrane protein involved in the formation of DMVs and the establishment of the viral replication complex (Hagemeijer et al., 2014). The SARS-CoV-2 NSP6 is 290 amino acids long (UniProt ID: P0DTD1).

NSP5: Although not a focus of this section, it is worth mentioning that the main protease (Mpro) domain of NSP5 has a catalytic dyad composed of residues H41 and C145, which are critical for its proteolytic activity (Zhang et al., 2020).

NSP9: NSP9 is a single-stranded RNA-binding protein that plays a role in viral replication and transcription (Sutton et al., 2004). The SARS-CoV-2 NSP9 consists of 113 amino acids (UniProt ID: P0DTD1).

NSP12: The RNA-dependent RNA polymerase (RdRp) domain of NSP12 contains conserved motifs A-G, with motif C (residues 753-767) being essential for its polymerase activity (Gao et al., 2020).

NSP13: NSP13 is a helicase that unwinds double-stranded RNA and DNA, facilitating viral replication and transcription (Jang et al., 2020). The SARS-CoV-2 NSP13 is 601 amino acids long (UniProt ID: P0DTD1). The helicase core of NSP13 consists of residues K288, S289, D374, E375, and Q404, which are involved in ATP binding and hydrolysis, while residues R337 and R339 are crucial for RNA binding (Jang et al., 2020).

NSP14: NSP14 is a bifunctional enzyme with 3'-to-5' exoribonuclease (ExoN) and guanine-N7-methyltransferase (N7-MTase) activities. The ExoN domain is involved in proofreading during viral RNA synthesis, while the N7-MTase domain is responsible for the formation of the 5' cap structure of viral RNAs (Ferron et al., 2018). The SARS-CoV-2 NSP14 consists of 527 amino acids (UniProt ID: P0DTD1). The ExoN domain of NSP14 contains a DEDDh motif (residues D90, E92, D243, D273, and H268) that is essential for its exoribonuclease activity and proofreading function (Ferron et al., 2018).

NSP15: NSP15 is an endoribonuclease that cleaves viral RNA at uridine sites, playing a role in evading host innate immune responses (Deng et al., 2019). The SARS-CoV-2 NSP15 is 346 amino acids long (UniProt ID: P0DTD1). The active site of the endoribonuclease domain of NSP15 consists of residues H234, H249, and K289, which are critical for its catalytic activity (Kim et al., 2020).

The NSPs highlighted in this section play crucial roles in various aspects of the SARS-CoV-2 life cycle. NSP1, NSP3, and NSP15 are involved in evading host innate immune responses by inhibiting translation, deubiquitinating host proteins, and cleaving viral RNA, respectively. NSP4 and NSP6 are essential for the formation of DMVs, which provide a protective environment for viral RNA synthesis. NSP9 is a single-stranded RNA-binding protein that

supports viral replication and transcription. NSP13 unwinds double-stranded RNA and DNA, facilitating the access of the RdRp to the viral genome. NSP14 ensures the fidelity of viral RNA synthesis through its proofreading activity and is involved in the formation of the 5' cap structure of viral RNAs. The functions of NSP2 remain largely unknown, but it is thought to contribute to viral RNA synthesis and host cell survival pathway modulation.

1.4 SARS-CoV-2 Evolution and Pathogenesis

1.4.1 Mutation rates and types

SARS-CoV-2, like other RNA viruses, exhibits a high mutation rate due to the lack of proofreading activity in its RNA-dependent RNA polymerase (RdRp) (Robson et al., 2020). The estimated mutation rate of SARS-CoV-2 is around 1×10^{-3} substitutions per base (30 nucleotides/genome) per year under neutral genetic drift conditions [85] or 1×10^{-5} – 1×10^{-4} substitutions per base in each transmission events, which is similar to other coronaviruses (Zhao et al., 2004). Mutations in the SARS-CoV-2 genome can be classified into three main types: substitutions, deletions, and insertions (Wang et al., 2020).

1. Effects on transmission and virulence

Mutations in the SARS-CoV-2 genome can have significant effects on the virus's transmission and virulence. For example, the D614G mutation in the S protein, which emerged early in the pandemic, has been associated with increased transmission efficiency and viral load in patients (Korber et al., 2020). Other mutations, such as N501Y and E484K, have been linked to increased binding affinity to the ACE2 receptor and reduced neutralization by antibodies, respectively (Greaney et al., 2021; Starr et al., 2020).

2. Spike protein adaptations

Mutations often target the spike protein of SARS-CoV-2 because of its vital function in viral entry and its exposure to the host immune system. Mutations in the receptor-binding domain (RBD) of the spike protein, such as N501Y, E484K, and K417N/T, have been identified in several variants of concern (VOCs) and are associated with increased transmissibility and immune escape (Garcia-Beltran et al., 2021). Other mutations, such as P681H and P681R, situated near the furin cleavage site, have been linked to enhanced spike protein cleavage and increased viral infectivity (Johnson et al., 2021).

3. NSP changes and viral phenotypes

In addition to mutations in the spike protein, changes in the NSPs of SARS-CoV-2 can also impact viral phenotypes and pathogenesis. NSPs play critical roles in viral replication, transcription, and immune evasion, making them potential targets for mutations that enhance viral fitness (Snijder et al., 2016).

a. Planned mutations in NSP1 for Omicron variants

NSP1 is a key virulence factor of SARS-CoV-2 that suppresses host gene expression and innate immune responses by targeting the 40S ribosomal subunit (Schubert et al., 2020). In this study, two mutations in the NSP1 protein of the Omicron variant: in the N-terminal region (G49S) and another in the XBB variant (K47R) will be introduced. These mutations were selected based on their predicted impact on the structure and function of NSP1, as well as their presence in other SARS-CoV-2 variants (Thorne et al., 2022).

b. Impact of mutations on NSP structure and function

The K47R and G49S mutations are situated in the N-terminal region of NSP1 that is critical for the protein's ability to suppress host gene expression (Narayanan et al., 2015). This mutation may alter the structure of the N-terminal domain and affect its interaction with the 40S ribosomal subunit, potentially modulating NSP1's host shutoff activity.

The C-terminal region of NSP1 has been implicated in the protein's stability and its interaction with other viral and host proteins (Shen et al., 2021). The substitution of residues may affect the local charge distribution and the conformation of the N and C-terminal domains. This change could influence NSP1's ability to interact with other proteins and modulate its role in immune evasion. Biochemical and biophysical assays, such as circular dichroism spectroscopy and isothermal titration calorimetry, could be employed to characterize the impact of this mutation on the stability and binding properties of NSP1.

By studying the effects of these mutations on the structure and function of NSP1, we aim to gain insights into the mechanisms by which SARS-CoV-2 evolves to enhance its replication efficiency and evade host immune responses. This knowledge can inform the development of targeted antiviral strategies and help predict the emergence of future variants with altered virulence and pathogenicity.

1.4.2 Immune evasion mechanisms

SARS-CoV-2 uses a variety of tactics to avoid the host's immune response, allowing it to establish a successful infection and spread within the host. These immune evasion mechanisms involve both innate and adaptive immunity.

1. Innate immune evasion SARS-CoV-2 has changed several mechanisms to counteract the host's innate immune response, particularly the interferon (IFN) system. The virus's NSPs play an important part in this process. For example, NSP1 suppresses host gene expression and interferes with the IFN response by preventing the translation of host mRNAs (Banerjee et al., 2020). NSP3, a papain-like protease, cleaves ISG15 (interferon-stimulated gene 15) from host proteins, thereby reducing their antiviral activity (Shin et al., 2020). Additionally, NSP6 inhibits the IFN-induced JAK-STAT

signaling pathway, further compromising the host's innate immune response (Xia et al., 2020).

2. Adaptive immune evasion SARS-CoV-2 also utilizes mechanisms to avoid the host's adaptive immune response, particularly antibody-mediated immunity. Mutations in the virus's spike protein, the primary target for neutralizing antibodies, have been acquired, diminishing its recognition by these antibodies. For example, the E484K mutation, identified in several variants of concern (VOCs), has been demonstrated to impart resistance to neutralizing antibodies. (Greaney et al., 2021). Furthermore, the high mutation rate of SARS-CoV-2 enables the rapid emergence of new variants with altered antigenic properties, posing challenges for the development of effective vaccines and therapeutic antibodies (Harvey et al., 2021).

1.4.3 Determinants of clinical severity

The clinical spectrum of COVID-19 ranges from asymptomatic infection to severe sickness, with some people developing life-threatening consequences such as acute respiratory distress syndrome (ARDS) and multi-organ failure. The clinical severity of COVID-19 is influenced by various factors:

Host factors: The likelihood of developing severe COVID-19 is significantly influenced by age, with those over 60 years of age having a higher chance of severe illness and death (Wu et al., 2020). Chronic lung disorders, diabetes, obesity, and cardiovascular diseases are among the conditions linked to a higher chance of developing severe COVID-19 (Guan et al., 2020). Genetic variables may potentially affect the severity and susceptibility to COVID-19. These variables include variations in the ACE2 gene and genes related to the innate immune response (Devaux et al., 2020).

1. ***Viral factors:*** The viral load and the infecting SARS-CoV-2 variant can impact the clinical severity of COVID-19. Higher viral loads in the upper respiratory tract have been linked to more severe disorders and heightened mortality (Fajnzyblber et al., 2020). Certain SARS-CoV-2 variants, such as the Alpha (B.1.1.7) and Delta (B.1.617.2) variants, have been linked to increased transmissibility and disease severity compared to the original Wuhan strain (Davies et al., 2021; Twohig et al., 2021).
2. ***Immune response:*** The degree of COVID-19 clinical severity is mostly determined by the host's immunological response to the SARS-CoV-2 infection. An inadequate or delayed innate immune response can lead to uncontrolled viral replication and the development of severe disease (Blanco-Melo et al., 2020). Conversely, an excessive or dysregulated immune response, characterized by a cytokine storm and hyperinflammation, can cause tissue damage and contribute to the pathogenesis of severe COVID-19 (Mehta et al., 2020).

1.4.4 Implications for diagnostics, interventions, therapeutics, vaccines

1. ***Diagnostics:*** Diagnostic test development and execution are impacted by the progression of SARS-CoV-2. As new variants emerge, it is crucial to ensure that existing diagnostic assays, such as RT-PCR and antigen tests, can detect these variants with high sensitivity and specificity. Genomic surveillance and the rapid characterization of emerging variants are essential for updating diagnostic protocols and maintaining the accuracy of testing (Vogels et al., 2021).
2. ***Interventions:*** The rise of SARS-CoV-2 variants displaying enhanced transmissibility and immune evasion characteristics might require adjustments to public health strategies. Non-pharmaceutical interventions, like social distancing, mask-wearing, and improved ventilation, remain critical in controlling the spread of new variants (Leung

et al., 2021). However, the effectiveness of these measures may need to be reassessed and adjusted based on the characteristics of the circulating variants.

3. **Therapeutics:** The evolution of SARS-CoV-2 presents hurdles for the development and utilization of antiviral treatments. Monoclonal antibodies, which have been used to treat COVID-19, may lose their efficacy against variants with mutations in the spike protein (Starr et al., 2021). Therefore, it is essential to monitor the susceptibility of emerging variants to existing therapeutics and to develop new therapies that target conserved regions of the virus or host factors involved in the viral life cycle (Zost et al., 2020).
4. **Vaccines:** The efficacy of COVID-19 vaccinations may be impacted by the appearance of SARS-CoV-2 variants with mutated antigenic characteristics. Vaccine-induced immunity may be less effective against variants with mutations in key epitopes recognized by neutralizing antibodies (Garcia-Beltran et al., 2021). To address this issue, vaccine manufacturers have begun developing updated vaccine formulations that incorporate sequences from emerging variants (Choi et al., 2021). Additionally, the development of universal coronavirus vaccines that target conserved regions of the virus could provide broader protection against future SARS-CoV-2 variants and other coronaviruses (Saunders et al., 2021).

In conclusion, the evolution of SARS-CoV-2 presents ongoing challenges for the diagnosis, treatment, and prevention of COVID-19. Continuous monitoring of emerging variants, coupled with the adaptation of diagnostic tools, therapeutic strategies, and vaccine designs, will be essential in effectively managing the COVID-19 pandemic and mitigating the impact of future variants of SARS-CoV-2.

CHAPTER TWO

Research Question and Hypothesis

Research question: Are there mutations in the SARS-CoV-2 genome that impact the viral replication machinery's proofreading ability?

Hypothesis: SARS-CoV-2 genome may have mutations that impact the ability of viral non-structural proteins that are involved in viral genome replication such that there is an increased rate of mismatches incorporated in the viral genome without compromising viability.

1. Mutations will enhance NSP functions controlling replication and transcription

Based on the initial analysis of early Omicron and XBB variants and the proposed computational analyses, the hypothesis is that the specific mutations introduced in the NSPs of SARS-CoV-2 will enhance their functions related to viral replication and transcription especially with regard to proofreading ability. There is some support by several evidence in the literature:

a. Crucial roles of NSPs in the viral cycle: The NSPs of SARS-CoV-2 play vital roles in various aspects of viral replication and transcription, including RNA synthesis, proofreading, and capping (V'kovski et al., 2021). Mutations in these proteins have the potential to modulate their activities, thereby influencing the efficiency of viral replication.

b. Adaptive mutations in other viruses: Research on other RNA viruses, like HIV and influenza, has shown that changes to viral proteins can improve their capabilities and raise the fitness of the virus (Bloom et al., 2010; Fernandes et al., 2016). These findings suggest that similar adaptive mutations could occur in SARS-CoV-2 NSPs.

This hypothesis will be further investigated through a comprehensive analysis that includes sequence analysis, comparative analysis, and structural studies. The integrated approach seeks to offer a thorough comprehension of how mutations affect the structure and function of NSPs, contributing to our knowledge of SARS-CoV-2 evolution and pathogenesis.

2. Structural and functional changes enable efficient viral replication

The hypothesis is further extended to propose that the structural and functional changes caused by the mutations in the NSPs will enable more efficient viral replication. This hypothesis is based on the following considerations:

a. Structure-function relationship: The structure of a protein is intimately linked to its function (Orengo & Thornton, 2005). Mutations that alter the structure of the NSPs, particularly in regions critical for their enzymatic activities or interactions with other viral or host factors, can significantly impact their function and, consequently, viral replication.

b. Evolutionary advantage: Mutations that enhance the efficiency of viral replication are likely to be selected for during the course of viral evolution (Grubaugh et al., 2020). The presence of the specific mutations in circulating SARS-CoV-2 strains suggests that they may confer an evolutionary advantage to the virus.

CHAPTER THREE

Research Methodology

3.1 Sequence Analysis

Obtain protein sequences of wild-type SARS-CoV-2 NSPs and mutated strains (Python, Biopython).

To obtain protein sequences of wild-type SARS-CoV-2 NSPs and mutated strains using Python and Biopython, the following steps will be taken:

Using the Biopython Entrez module, the whole genome sequence of the wild-type SARS-CoV-2 will be obtained from the NCBI GenBank database. The accession number for the SARS-CoV-2 reference genome is NC_045512.2 (Wu et al., 2020). Subsequently, the protein sequences of NSPs such as NSP1, NSP2, NSP3, NSP4, NSP6, NSP9, NSP13, NSP14, and NSP15 will be extracted from the genome sequence using the SeqIO module in Biopython.

Following that, protein sequences of mutated strains of SARS-CoV-2 will be obtained from public databases like GISAID (Shu & McCauley, 2017) and the NCBI Virus database (Hatcher et al., 2017). These databases offer a comprehensive collection of SARS-CoV-2 genomic sequences submitted globally. Specific mutated strains of interest, such as the Omicron variants, will be searched, and their sequences will be downloaded in FASTA format.

Using the SeqIO module in Biopython, the downloaded FASTA files will be parsed, and the protein sequences of NSPs from the mutated strains will be extracted. A subsequent step involves comparing these sequences to the wild-type NSP sequences to identify the specific mutations present in each strain.

Use bioinformatics tools to align sequences and identify differences (NCBI BLAST, Clustal Omega, MAFFT, JELVIEW)

After acquiring the protein sequences of wild-type and mutated SARS-CoV-2 NSPs, various bioinformatics tools will be employed to align the sequences and pinpoint the differences

between them. Sequence alignment is a critical step in comparative genomics, crucial for understanding evolutionary relationships and the functional implications of mutations (Pearson, 2013).

A widely used tool for sequence alignment is NCBI BLAST (Basic Local Alignment Search Tool) (Altschul et al., 1990). The blastp program, designed for protein sequence alignment, will be utilized to compare the mutated NSP sequences against the wild-type sequences. BLAST identifies conserved regions and regions with mutations, providing valuable insights. Clustal Omega (Sievers et al., 2011), another popular tool for multiple sequence alignment, will be employed. By using HMM profile-profile methods and seeded guide trees, Clustal Omega generates alignments and highlights differences between the wild-type and mutated NSP sequences. The tool offers a graphical representation for easy visualization of mutations. MAFFT (Multiple Alignment using Fast Fourier Transform) (Kato & Standley, 2013), known for its speed and accuracy, will be used for sequence alignment. The MAFFT-DASH web server (Rozewicki et al., 2019) will facilitate alignments and visualization. Employing multiple alignment tools ensures cross-validation and accuracy in identifying mutations.

In summary, the sequence analysis methodology involves obtaining protein sequences of wild-type and mutated SARS-CoV-2 NSPs using Python and Biopython. Subsequently, bioinformatics tools such as NCBI BLAST, Clustal Omega, and MAFFT will be used to align the sequences, revealing differences. This approach establishes a comprehensive understanding of NSP mutations, laying the groundwork for further structural and functional analyses.

3.2 Comparative Analysis

Compare properties of wild-type and mutated NSPs

After aligning the sequences of the wild-type and mutated NSPs, a comprehensive comparative analysis will be conducted to identify potential differences in their properties. This analysis

involves examining various physicochemical characteristics, such as molecular weight, isoelectric point, hydrophobicity, and amino acid composition (Gasteiger et al., 2005).

To compute these properties for each NSP sequence, bioinformatics tools such as ProtParam (Gasteiger et al., 2005) and ProtScale (Gasteiger et al., 2005) from the ExPasy server will be utilized. Numerous physicochemical parameters are available through ProtParam, such as the grand average of hydropathicity (GRAVY), extinction coefficient, predicted half-life, theoretical pI, molecular weight, atomic composition, aliphatic index, and extinction coefficient. ProtScale generates amino acid scale profiles for a given protein sequence, facilitating the analysis of hydrophobicity, hydrophilicity, and other scale-based properties.

3.3 Structure Prediction and Visualization

Utilize protein structure prediction tools to model 3D structures of NSPs (MODELLER, SWISS-MODEL, I-TASSER, RAPTORX)

To understand the structural basis of the functional changes caused by mutations in the NSPs, it is essential to have accurate 3D models of both the wild-type and mutated proteins. In cases where experimental structures are not available, state-of-the-art protein structure prediction tools will be employed to generate high-quality models.

MODELLER is a frequently utilized resources for modeling protein structure (Webb & Sali, 2016). This comparative modeling tool utilizes sequence alignment and template structures to generate 3D models of the target protein. Supposing that proteins with comparable sequences frequently possess comparable structures, MODELLER will use the wild-type NSP sequences to search for appropriate template structures in the Protein Data Bank (PDB) (Berman et al., 2000). Subsequently, these templates will be used as input for MODELLER to generate multiple models, which will be ranked and selected based on their quality scores (e.g., DOPE score).

Another widely-used web-based platform for protein structure prediction is SWISS-MODEL (Waterhouse et al., 2018). SWISS-MODEL employs a fully automated homology modeling pipeline to generate 3D models of proteins. The platform searches for suitable template structures in the SWISS-MODEL Template Library (SMTL), selecting the best templates based on sequence identity and quality. The models are then built using ProMod3 and evaluated using various quality estimation tools, such as QMEAN (Benkert et al., 2011) and ModFOLD (Maghrabi & McGuffin, 2017).

For challenging targets where homology modeling may not be sufficient, I-TASSER (Yang et al., 2015) will be employed. A hierarchical method for predicting protein structure and function is called I-TASSER which combines *ab initio* folding, fragment assembly, and template-based modeling to generate 3D models. It also includes additional steps, such as structure-based functional annotations and ligand-binding site predictions. I-TASSER has consistently performed well in the Critical Assessment of protein Structure Prediction (CASP) experiments (Moult et al., 2018).

After generating the 3D models for the wild-type and mutated NSPs, their quality will be assessed using various evaluation tools. These include Verify3D (Bowie et al., 1991), which assesses how well the 3D model matches the amino acid sequence, and PROCHECK (Laskowski et al., 1993), which analyzes the stereochemical quality of the model. Models with high quality scores will be selected for further analysis.

Assess if mutations lead to significant structural changes (PyMOL)

To visualize and compare the 3D structures of the wild-type and mutated NSPs, we will use PyMOL (Schrödinger, 2021), a powerful molecular visualization system. PyMOL provides a user-friendly interface for rendering and manipulating 3D structures of proteins and other biomolecules.

We will load the 3D models of the wild-type and mutated NSPs into PyMOL and perform structural alignments to identify any significant deviations caused by the mutations. The aligned structures will be visually inspected to locate the mutated residues and assess their impact on the local and global structure of the protein.

PyMOL's powerful selection and visualization tools will be used to highlight the mutated residues and their surrounding regions. We will analyze the changes in secondary structure elements (α -helices, β -sheets, or loops) and the potential disruption of important intramolecular interactions (hydrogen bonds, salt bridges, or hydrophobic contacts) caused by the mutations. In addition to visual inspection, we will use PyMOL's measurement tools to quantify the structural differences between the wild-type and mutated NSPs. This includes calculating the separations between particular residues or functional groups as well as the root-mean-square deviation (RMSD) of the backbone atoms.

If the mutations are found to cause significant structural changes, we will investigate the potential functional implications of these alterations. For example, the function of a protein can be significantly affected by mutations that alter the binding interface of a protein-protein interaction or the geometry of an enzyme's active site.

The structural insights obtained from the 3D models and PyMOL analysis will complement the results from the sequence analysis, comparative analysis, and molecular dynamics simulations. By integrating these findings, we will gain a comprehensive understanding of how the mutations affect the structure and function of the NSPs and their role in SARS-CoV-2 evolution and pathogenesis.

In summary, we will use protein structure prediction tools (MODELLER, SWISS-MODEL, and I-TASSER) to generate high-quality 3D models of the wild-type and mutated NSPs. The models will be evaluated using various quality assessment tools. PyMOL will be used to visualize and compare the 3D structures, evaluate how mutations affect the structure of protein,

and investigate the potential functional implications of these changes. The structural insights will be integrated with other computational analyses to provide a thorough comprehension of the effects of NSP mutations on SARS-CoV-2 biology.

Identify potential differences in activity or function

To identify potential differences in the activity or function of the wild-type and mutated NSPs, various computational tools and databases will be employed. One such resource is the Pfam database (El-Gebali et al., 2019), which contains a large collection of protein families and domains. Searching the Pfam database with the NSP sequences will help identify conserved domains and motifs crucial for the function of each NSP.

The Pfam domain architecture of the wild-type and mutated NSPs will be compared to determine if any mutations have occurred within these functional domains. Mutations within or near conserved domains may significantly impact the activity or function of the NSP (Forni et al., 2017). For example, mutations in the catalytic site of an enzyme or the substrate-binding region of a protein could alter its function.

Another useful resource for identifying functional differences is the Gene Ontology (GO) database (Ashburner et al., 2000). For the purpose of characterizing the biological processes, molecular operations, and cellular constituents connected to genes and proteins, GO offers a standardized language. Tools such as InterProScan (Jones et al., 2014) will be used to assign GO terms to the wild-type and mutated NSPs based on their sequence features. By comparing the GO annotations between the wild-type and mutated NSPs, any changes in the predicted functions or biological processes associated with the mutations can be identified.

3.4 Look for changes in enzymatic activity, substrate binding, or other relevant functional aspects

To ascertain whether mutations could have an effect on the functional aspects of SARS-CoV-2 NSPs, a combination of sequence-based and structure-based methods will be employed. For NSPs with known enzymatic functions, such as the papain-like protease (PLpro) domain of NSP3 or the RNA-dependent RNA polymerase (RdRp) of NSP12, tools like PROSITE (Hulo et al., 2006) and PRINTS (Attwood et al., 2003) will be used to identify conserved catalytic motifs and active site residues. Comparing these motifs and residues between the wild-type and mutated NSPs will help predict if mutations have occurred in regions critical for enzymatic activity.

For NSPs involved in substrate binding, such as the RNA-binding domain of NSP9 or the nucleic acid-binding domain of NSP14, tools like BindN+ (Wang et al., 2010) and RNABindRPlus (Walia et al., 2014) will be utilized to predict RNA-binding residues based on the protein sequence. By comparing the predicted RNA-binding residues between the wild-type and mutated NSPs, potential changes in substrate binding affinity or specificity can be identified.

In addition to sequence-based methods, structure-based approaches will be used to predict the impact of mutations on NSP function. Molecular docking and molecular dynamics simulations will be utilized to examine the interactions between NSPs and their substrates or binding partners where those NSPs have accessible experimental structures or high-quality homology models. Tools like AutoDock Vina (Trott & Olson, 2010) and GROMACS (Abraham et al., 2015) will be used for these analyses. Comparing the binding energies, conformational changes, and stability of the wild-type and mutated NSP-substrate complexes will provide further understanding into the potential functional consequences of the mutations.

In summary, the comparative analysis involves a multi-faceted approach to identify differences between the wild-type and mutated NSPs. By examining physicochemical properties,

conserved domains, GO annotations, enzymatic motifs, and substrate binding residues, predictions regarding the potential impact of mutations on NSP function can be made. Additionally, structure-based methods like molecular docking and dynamics simulations will offer a more detailed understanding of the functional consequences of the mutations. This comprehensive analysis will guide further experimental validation and help elucidate the role of NSP mutations in SARS-CoV-2 evolution and pathogenesis.

3.5 Experimental Procedure

Whole-genome sequences of the XBB and early Omicron variants were obtained from GISAID and imported into Nextstrain (Hadfield et al., 2018) to translate them into protein sequences. Using a Python script with SeqIO, NSPs were extracted from each variant type. Jalview facilitated alignment of these NSPs with those of the Wuhan strain to identify sequence similarities. Subsequently, each NSP sequence was modeled using I-TASSER, employing either homology or *ab initio* methods based on available X-ray crystallography structures. The resulting protein structures with the highest confidence scores were then compared to those of the Wuhan strain in Pymol to assess mutation-induced structural effects. In addition to structural analysis, we examined the physicochemical properties of mutated proteins using ExPASy's Protparam. Characteristics including molecular weight, theoretical isoelectric point (pI), amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index, and grand average of hydropathicity were evaluated. To investigate the impact of mutations on SARS-CoV-2's proofreading ability, we gathered 17,928 XBB and early Omicron sequences from GISAID, spanning December 19-31, 2022. These sequences harbored the mutations of interest (Table 1.0). The reference sequence NC_045512.2 was obtained from NCBI and translated into protein sequences, from which NSPs were extracted. For NSPs lacking X-ray crystallography structures, *ab initio* modeling with I-

TASSER was utilized. Accuracy of the I-TASSER models was verified by comparing the NSP1 X-ray structure to that of the NC_045512.2 reference sequence. Homology modeling was employed to determine whether mutations affected protein structure or core regions, with a focus on identifying any induced conformational changes. Mutations in proteins can lead to various effects, ranging from significant structural alterations to more subtle changes in stability and dynamics (Rana et al., 2015). Therefore, we comprehensively analyzed the physicochemical properties of each mutation in comparison to the reference sequences.

Table 1.0: Mutations of interest of the various SARS-CoV-2 variants.

NSP	Mutation	Covid-19 Variant
NSP1	G49S	Omicron
	K47R	XBB
NSP2	G118S	Omicron
	G265V	Omicron
	A510V	XBB
NSP3	Y129H	Omicron
NSP4	A380V	XBB
NSP6	N156S	Omicron
NSP9	T35I	XBB
NSP13	S36P	XBB
NSP14	D222Y	XBB
	D324Y	Omicron
NSP15	D282N	Omicron

3.5.1 Sequence Alignment of XBB and early omicron variants.

After conducting a thorough analysis of the XBB and omicron protein sequences, we were able to perform a sequence alignment using MAFFT integrated into Jalview, which showed that most of the sequences downloaded from GISAID were consensus sequences (Figure 4.0). Subsequently, we grouped the consensus sequences and selected a representative sequence along with its GISAID EPI identification number for further analysis. Following this, we identified and extracted the various NSPs based on specific mutations (Table 1.0).

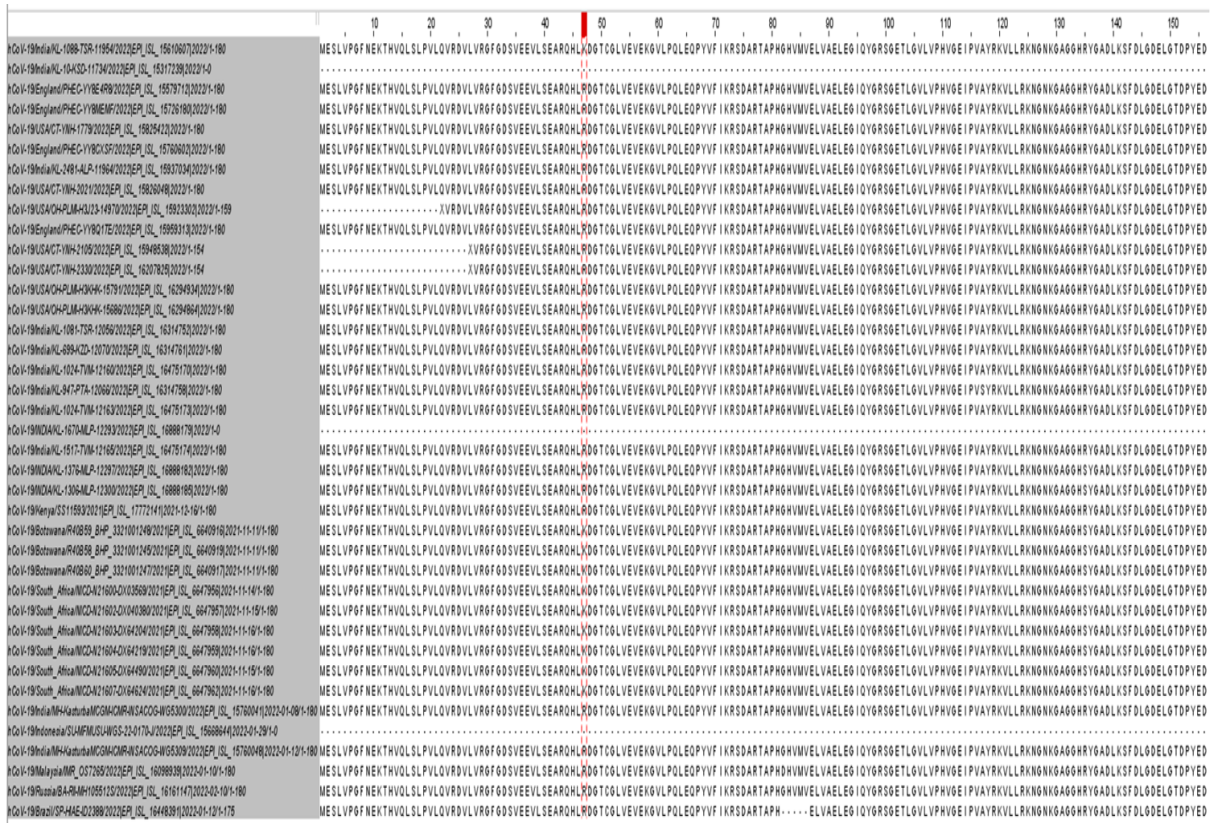


Figure 4.0: Sequence alignment of XBB and Omicron variants.

3.5.2 Structure Modeling of SARS-CoV-2 NSP1 and I-TASSER Accuracy Assessment.

The crystal structure of NSP1 from SARS-CoV-2 (7K3N), determined through X-ray diffraction, revealing the presence of amino acid residues from 13 to 127 at a resolution of 1.65 Å was overlaid with the I-TASSER wildtype model, which falls short of the typical 180 amino acids found in NSP1s (Semper et al., 2021). Due to this incomplete representation of the full-length NSP1 (as well as the other NSPs) protein with its 180 amino acid composition, comparative modeling based on this determined structure may lead to inaccuracies in subsequent analyses. Therefore, we employed the I-TASSER software to predict the full 180 amino acid structure of NSP1. To ensure the accuracy of the predicted model, we overlaid it with the X-ray structure (PDB ID; 7k3N) and assessed its confidence level. Structural analysis of the two protein structures showed that the predicted model was very accurate and had extra loops [Figure 5.0 (A)], an indication of a structure with more amino acids.

The global accuracy of an I-TASSER model is represented by C-score values, a model with accurate global topology is indicated by a score larger than -1.5 (Zhou et al., 2022). The closer the C-score is to 2, the more accurate the generated structure is. The C-scores generated for the NSP1 were 2.0.

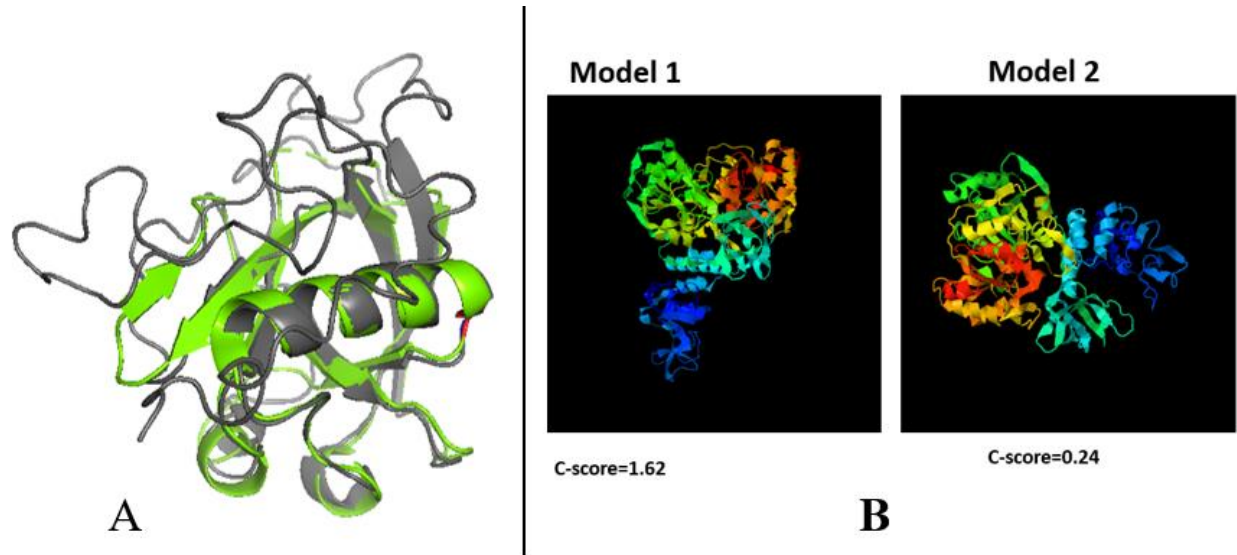


Figure 5.0: Predicted models of SARS-CoV-2 NSP1 from Wuhan-Hu-1 reference strain. (A) Ribbon models showing structural alignments of 7K3N and NSP1 generated by I-TASSER. Structure coded green is the X-ray structure (PDB ID; 7k3N) of the SARS-CoV-2 NSP1 (from 3 to 127 amino acid residues) and the structure coded ash is the predicted model of SARS-CoV-2 NSP1 (from 1 to 180 amino acid residues) from I-TASSER (model 1). (B) Ribbon models showing confidence (C-score) levels of an I-TASSER predicted structure of SARS-CoV-2 NSP.

CHAPTER FOUR

Results and Discussion

4.1 Homology Modeling NSP1

NSP1, also known as the leader protein or host translational inhibitor, is known to inhibit host mRNA translation by binding to the 40S ribosomal subunit *via* its C-terminal domain and promoting mRNA degradation (Mendez et al., 2021). Its mechanism of action involves two main steps:

1. **Binding to 40S Ribosomal Subunit:** The C-terminus of NSP1 binds to the 40S ribosome, entering the mRNA entry tunnel through the A and P sites to obstruct mRNA binding. (Zhao et al., 2021; Schubert et al., 2020). Simultaneously, The N-terminus of NSP1 is directed towards the ribosomes, serving as an obstacle to impede the entry of mRNA into the channel. (Zhao et al., 2021).
2. **Enhanced Host Shutoff:** The N-terminal region and adjacent residues of NSP1 stabilize the interaction between NSP1 and the 40S subunit, boosting its host shutoff functions (Mendez et al., 2021).

RNA Binding Domain (RBD) and amino acids involved:

- The RNA-binding domain (RBD) found in NSP1 is tasked with its interaction with RNA.
- **Amino Acids:** Arginine residues: R124, R125, and R126: Identified as important for RNA binding. Lysine residues: K164: Implicated in RNA interaction. Tryptophan residues: W187 and W189: Thought to be involved in RNA interactions. Other residues: F70 and Y99: Mentioned in some reports for their potential role in RNA binding. Consequently, the mutations of interest, G49S and K47R for the omicron and XBB variants, respectively, are located within the N-terminal region (Selvavinayagam et al., 2023). Therefore, we examined the consequences of these mutations in NSP1 by comparing the wild type and mutant *de novo*

structures using PyMol, we appended the wild type onto the K47R and G49S mutations (Figure 6) to identify conformational changes and examine a root mean squared deviation (RMSD) between the appended structures. Interestingly, both mutations did not induce any conformational changes in the critical regions of their respective variants when their structures were compared. K47R mutation yielded a root mean squared deviation (RMSD) of 0.542 whilst that of G49S mutation yielded an RMSD of 0.589.

4.1.1 Mutation Impact: Glycine to Serine (G49S):

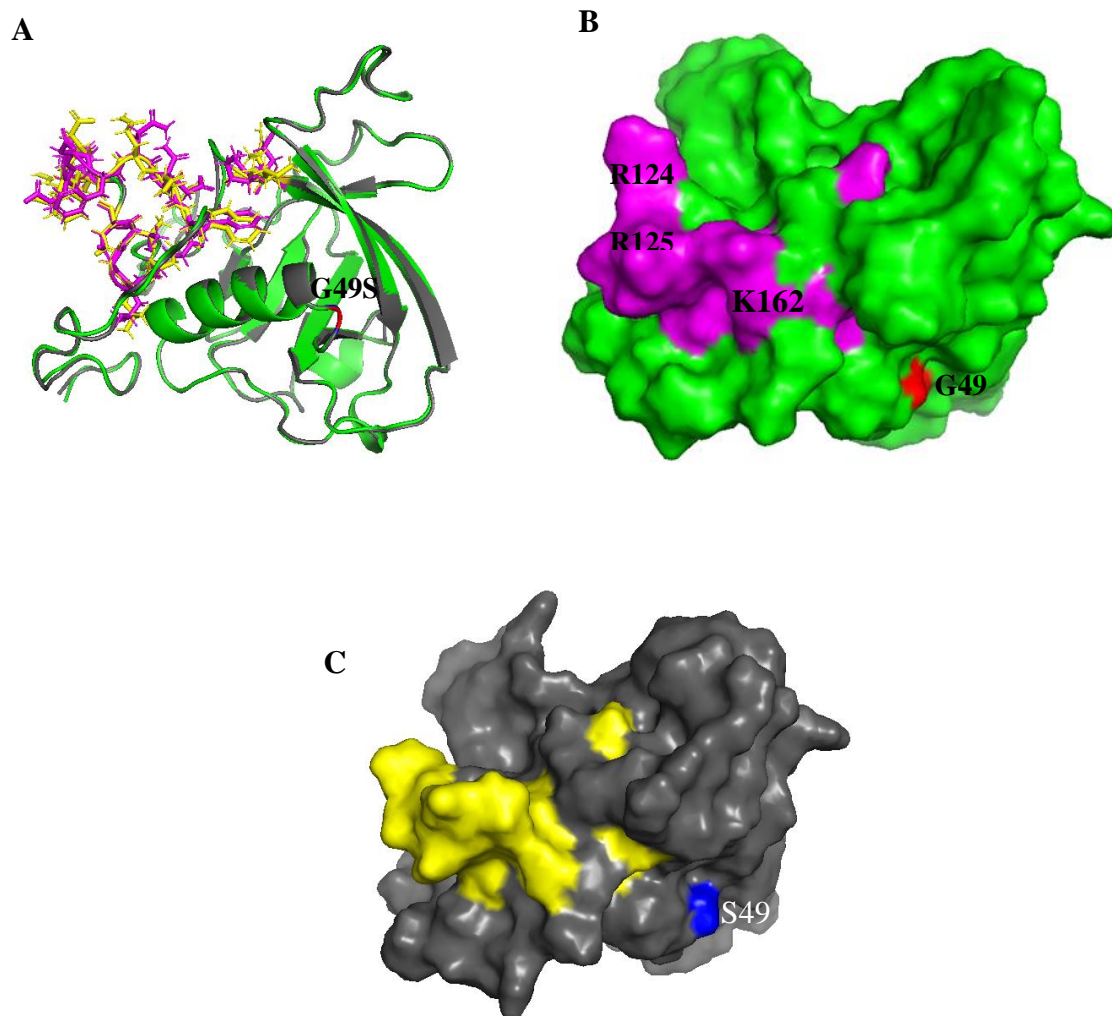


Figure 6.0: Homology model of NSP1 wildtype and G49S. (A) NSP1 wildtype and mutant superimposed. Color code, Green represents the wildtype and ash, the mutant. Critical binding sites for RNA on NSP1 wildtype (magenta) and G49S mutant (yellow). (B) Surface structure of wildtype showing binding sites and amino acid G. (C) Surface structure of mutant showing RNA binding sites and mutated amino acid.

While the specific impact of the mutation needs to be experimentally verified, the substitution of glycine with serine at the 49th position in NSP1 did not significantly change its structure. This may be due to the similarities between the two amino acids, the location of the mutation, and the potential for compensatory changes. However, detailed structural and functional studies would be needed to confirm the actual effect of this mutation on NSP1.

Glycine (G) is the smallest amino acid and is often found in regions where flexibility is required due to its lack of a side chain. Serine (S) is slightly larger and has a side chain, but it is also relatively small and flexible. These two amino acids have similar properties (conservative mutation). The substitution of glycine by serine introduces a hydroxyl group (-OH) in the side chain, which might not drastically alter the overall structure. Conservative mutations often have minimal effects on protein structure and function, especially if they occur in flexible or non-critical regions.

The position of the substitution in the protein is also essential. The 49th position in the NSP1 protein is located in a region that is not directly involved in the protein's active site, binding interfaces, or critical structural elements. The position, as depicted in figure 4A, resides within a loop or a flexible region of the protein. Mutations at this site may not exert a significant influence on the protein's overall structure. Some proteins are inherently tolerant to amino acid substitutions at certain positions without affecting their overall function. If the 49th position in NSP1 is not crucial for its activity or stability, a mutation here may not lead to significant changes.

Proteins can sometimes accommodate mutations by making compensatory changes in other parts of their structure. The protein might adjust locally to accommodate the serine residue without altering its overall structure or function.

4.1.2 Mutation Impact: Lysine to Arginine (K47R):

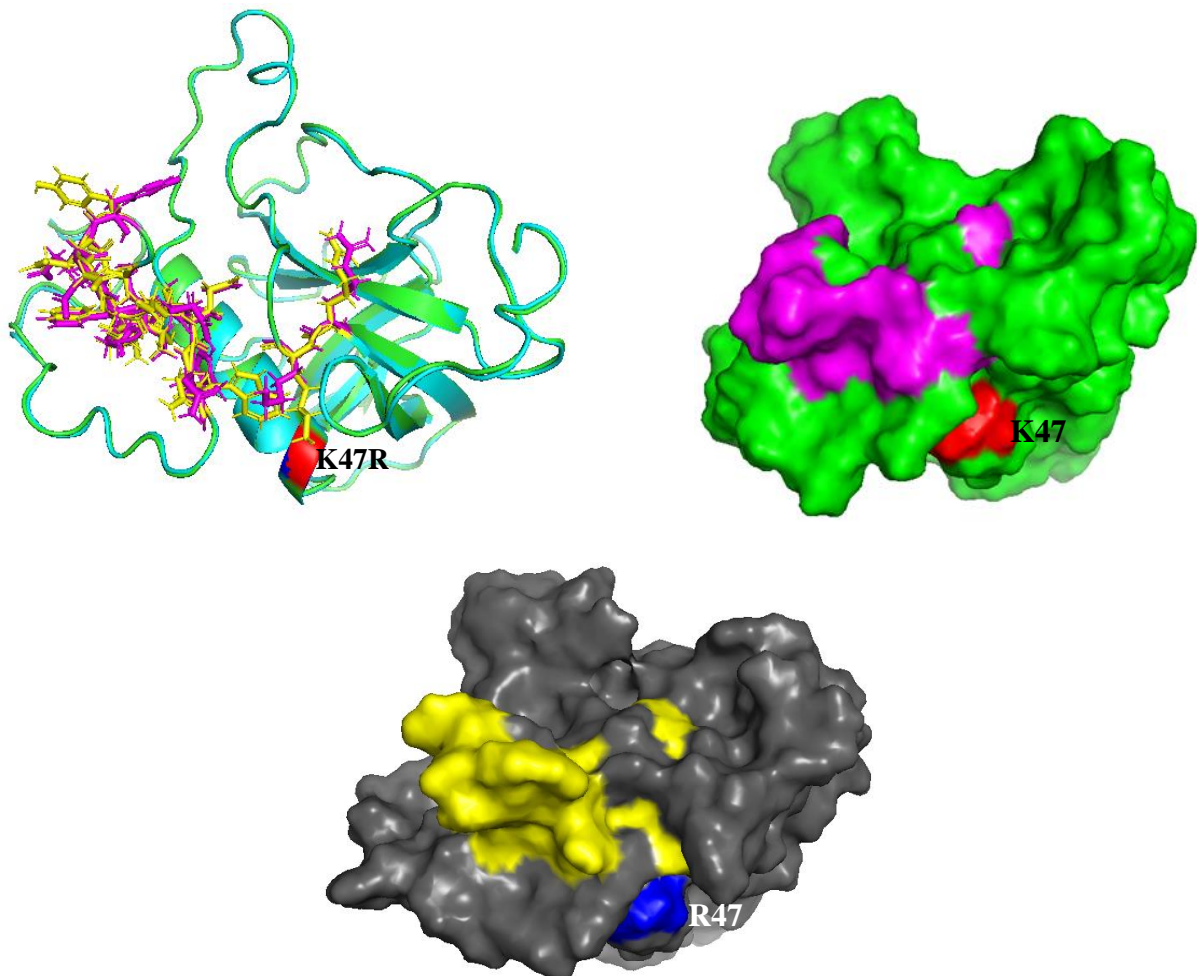


Figure 7.0: Homology model of NSP1 wildtype and K47R. (A) NSP1 wildtype and mutant superimposed. Color code, Green represents the wildtype and cyan, the mutant. magenta are the critical binding sites for RNA on NSP1 wildtype, yellow same on mutant (B) Surface Structure of wildtype showing RNA binding sites and amino acid of interest. (C) Surface Structure of mutant showing RNA binding sites and mutated amino acid.

Similarity and Conservation: Lysine (K) and arginine (R) share positive charge and size. K47R is a conservative mutation, often tolerable in non-critical regions. The 47th position is not directly involved in critical functions. Even though as seen above the K47R region is in an alpha helix, the mutation had minimal impact on the overall NSP1 structure

4.2 Homology Modeling of NSP2

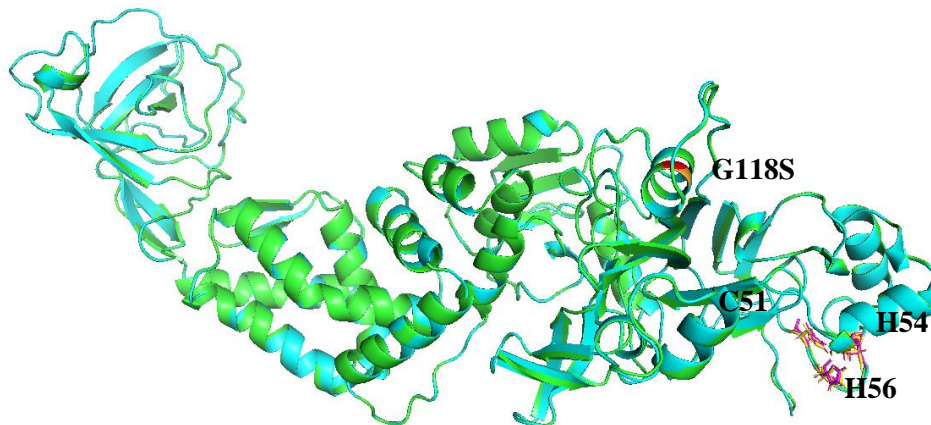
NSP2 promotes the creation of the 4EHP-GIGYF2 protein complex, which suppresses the translation of IFN1 β mRNA (Korneeva et al., 2023). NSP2 engages with the 4EHP-GIGYF2 complex through its N-terminal region, which includes the conserved zinc finger domain, to disrupt the function of this complex. The 4EHP-GIGYF2 protein complex participates in miRNA-mediated translational suppression and the degradation of tristetraprolin-targeted mRNAs. (Korneeva et al., 2023). The interaction between NSP2 and 4EHP-GIGYF2 results in a complex mode of binding, impacting the cap-binding pocket of 4EHP or influencing the recruitment of GIGYF2 co-factors. NSP2 disrupts the activity of 4EHP-GIGYF2, potentially disturbing the production of IFN- β and causing dysregulation in sustained cytokine production (Zou et al., 2022).

The N-terminal region of the 638 amino acid protein NSP2 in SARS-CoV-2 has three zinc fingers (ZnFs). The ZnFs in SARS-CoV-2 NSP2 may not bind to nucleic acids directly, despite sharing structural similarities with RNA binding proteins (Ma et al., 2021). Its function is linked with transcription, inhibition of host protein synthesis, and viral replication; however, its specific function has not yet been discovered (Ma et al., 2021). NSP2 is involved in the replication of the viral genome and the assembly of viral particles. It likely interacts with other viral proteins and host factors to facilitate these processes. NSP2 may also play a role in modulating the host cell environment to benefit viral replication and evade host immune

responses. It's been proposed to disrupt host cell transcription, translation, and immune signalling pathways.

Some studies propose that NSP2, along with other viral proteins, contributes to the rearrangement of host cell membranes to form viral replication organelles. These organelles create a favourable environment for viral RNA replication. While its specific functions and mechanism of action are not yet fully understood, as research on this protein is ongoing, specific residues involved in the binding of NSP2 to other proteins or structures are still being investigated. We examine the impact of our mutations of interest (Table 1.0), We specifically looked at whether there are any conformational changes induced by the mutations on the proposed critical binding regions (C18, C51, H54, H56). Structural analysis of the mutations compared to their reference wildtype revealed no changes in conformation of structure as well as the critical binding sites of these proteins (Figure 8.0). Furthermore, all three structures had an RMSD value of 0.260.

A



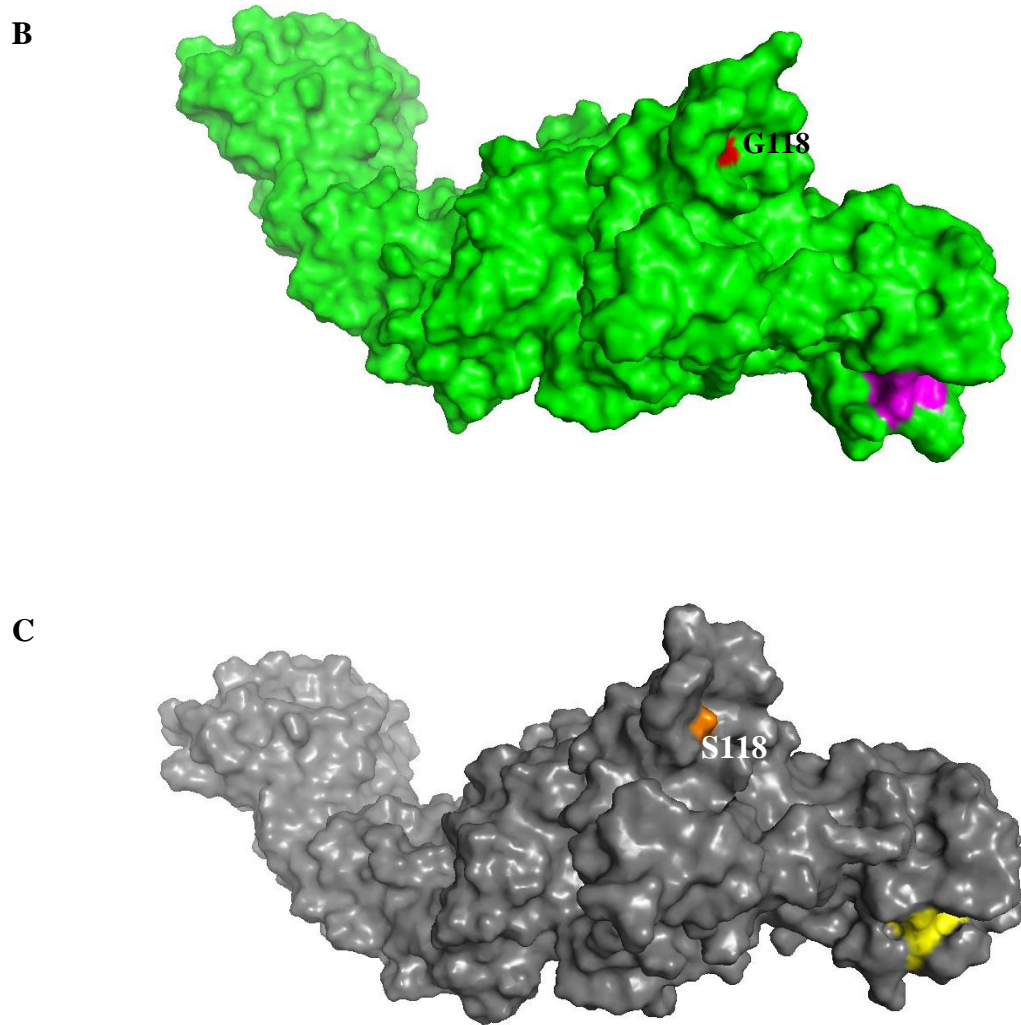


Figure 8.0: Homology model of NSP2 wildtype and G118S. (A) NSP2 wildtype and mutant superimposed. Color code, Green represents the wildtype and cyan, the mutant. Magenta are the critical prohibitin binding sites on NSP2 wildtype, yellow same on mutant (B) Surface Structure of wildtype showing prohibitin binding sites and wildtype amino acid of interest. (C) Surface Structure of mutant showing binding sites for prohibitin and mutated amino acid.

4.3 Homology Modeling of NSP13

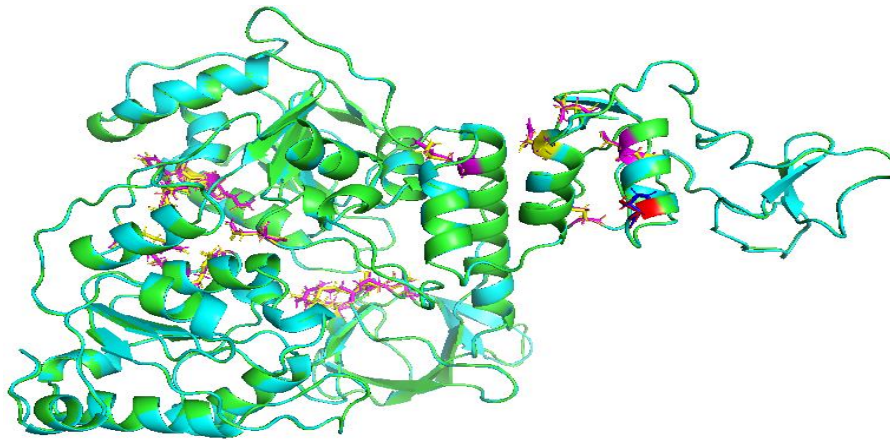
An enzyme or helicase, crucial for coronaviruses, facilitates the hydrolysis of the 5' γ -phosphate of the nascent mRNA transcript. This process is essential for generating the 5'-ppN ends, which are required for transferring guanosine monophosphate by RNA guanylyltransferase (GTase), a crucial step in forming the primary cap structure (GpppN). NSP13 collaborates with NSP10, NSP14, and NSP16 to ensure the efficient mRNA capping of SARS-CoV-2 (Low et al., 2022). NSP13 is a multifunctional enzyme that plays a critical role in viral replication by unwinding double-stranded RNA during the replication process. NSP13 possesses multiple functional domains, each with distinct roles, and certain crucial amino acids play roles in the binding and catalytic activities within these domains. NSP13 is a triangular pyramid-shaped, 603-amino acid helicase with five domains: Stalk, 1B, 1A(RecA-like), Zinc binding domain (ZBD), and 2A(RecA-like 2A) (Newman et al., 2021). The nucleotide-binding site is located within the crevice between the 1A and 2A domains. The zinc-binding domain (ZBD) has the capacity to bind a minimum of three Zn^{2+} ions by utilizing twelve conserved C/H residues. NSP13 contributes to the early stage of RNA capping by cleaving the γ -phosphate group of the 5'-terminus of viral RNA, thereby impeding detection by the host immune system (Justo Arevalo et al., 2023a). Hel1 domain contains the ATPase and RNA/DNA helicase activities. Hel2 Domain is important for coupling ATPase activity with the translocation along the nucleic acid. ATP Binding and Hydrolysis (Hel1 Domain) contains Motif I (Walker A) which contains the sequence GXXXXGKT/S, which is involved in ATP binding and hydrolysis. Motif II (Walker B): Contains the sequence DExx, which coordinates the magnesium ion for ATP hydrolysis. Motif VI: Contains the sequence QxxR, which is involved in RNA/DNA binding. Motif IV: Contains the sequence DEAH, involved in RNA unwinding. RNA/DNA Binding (Hel1 Domain) has several arginine residues that are critical

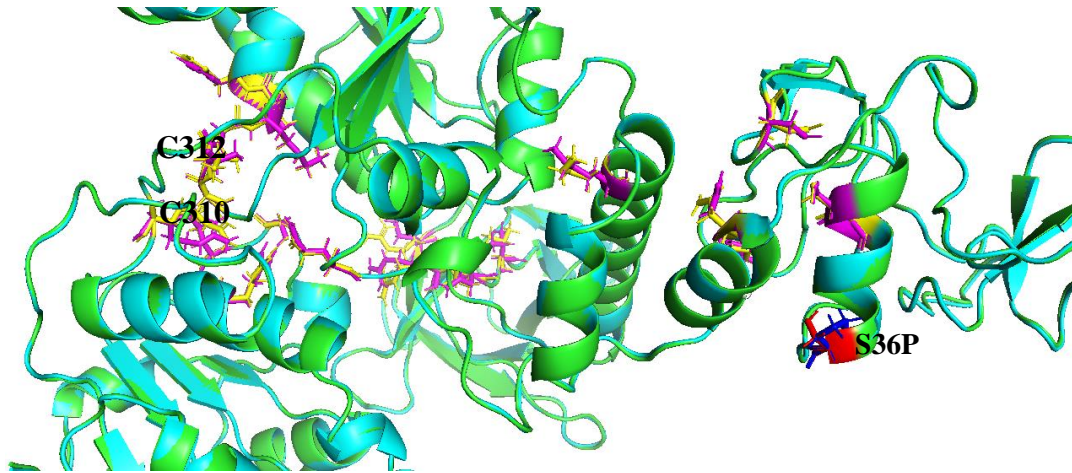
for RNA binding and unwinding. For example, R367, R441, and R443 have been identified as important for both RNA binding and helicase activity.

The Zinc-Binding Domain has cysteine and histidine residues: These residues are typically involved in coordinating Zn^{2+} ions in ZBD; they are C310, C312, H366, and H370. These residues are important for RNA binding and possibly helicase activity. It is also believed that the protein contributes to the inhibition of host protein production.

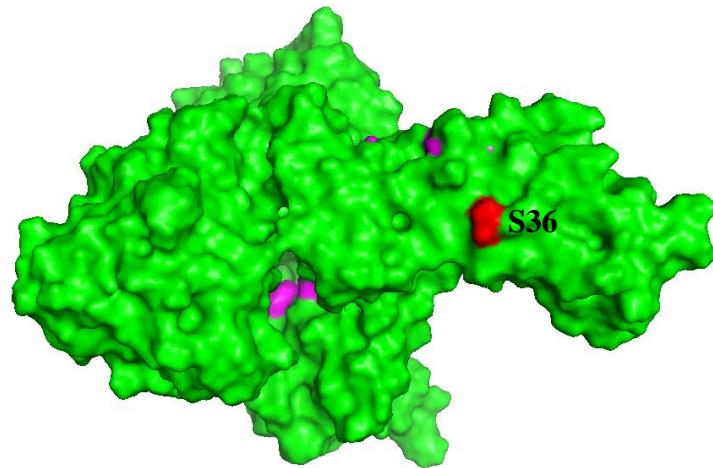
To assess the impact of mutation S36P (Table 1.0) on the structure and function of the NSP13 of XBB strains, we appended the NSP13 structure from the reference strain onto the representative XBB strain with the mutation. Structural comparisons showed no conformational change in either the structure of the XBB or the critical binding sites (Figure 9.0). RMSD of 0.290 was recorded from appending the two structures.

A





B



C

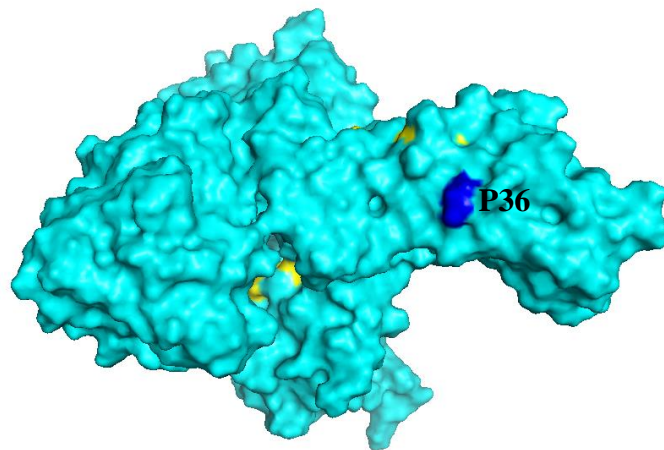


Figure 9.0: Homology model of NSP13 wildtype and S36P mutant. (A) NSP13 wildtype and mutant superimposed. Color code, Green represents the wildtype and cyan, the mutant. Magenta are the critical nucleotide and zinc binding sites on NSP13 wildtype, yellow same on mutant (B) Surface Structure of wildtype showing zinc and nucleotide binding sites and

wildtype amino acid of interest. (C) Surface Structure of mutant showing binding for zinc and nucleotide sites and mutated amino acid in blue.

4.3.1 Mutation Impact: Serine to Proline Mutation (S36P)

Proline (P) is unique among amino acids due to its cyclic side chain, which limits its conformational flexibility. Proline residues often introduce kinks or bends in protein structures due to their rigid nature. This rigidity can disrupt regular secondary structure elements like alpha-helices.

The 36th position in this protein may not be directly involved in critical structural elements or active sites as it is located the N-terminus of the protein, which is typically more flexible and less critical for maintaining the overall protein fold and structure. The substitution of serine with proline introduces rigidity due to proline's unique structure. This rigidity could potentially stabilize local conformations or loops in the protein. The 36th position might be in a region where proline's kinking effect is accommodated without disrupting the overall structure. The region surrounding the 36th position may have a conserved structure or function that requires proline's unique properties. Proline's introduction may cause minor local structural changes, but if the surrounding structure can accommodate this, the overall fold may remain unchanged. Some proteins are tolerant to amino acid substitutions at certain positions without affecting their overall function. While the specific impact of a mutation needs to be experimentally verified, the substitution of serine with proline at the 36th position in NSP13 might not significantly change its overall structure.

4.4 Homology Modeling of NSP14

The SARS-CoV-2 NSP14 is a dual-function enzyme found in *Coronaviridae*, consisting of an N-terminal exonuclease domain (ExoN) and a C-terminal N7-methyltransferase domain

(Imprachim et al., 2023). The ExoN domain serves as a proofreading exoribonuclease, improving RNA fidelity by eliminating mismatched nucleotides during RNA synthesis. On the contrary, the N7-MTase domain is involved in capping activity, using S-adenosyl-L-methionine (SAM) to create a cap 0 structure (Chen et al., 2013). NSP14 functions as a S-adenosylmethionine (SAM)-dependent methyltransferase (MTase), Utilizing SAM as a methyl donor is essential for the viral life cycle. Meanwhile, ExoN is an enzyme tasked with RNA degradation by eliminating nucleotides from either end of an RNA structure, a critical step in generating various RNAs from the RNA template in RNA viruses. In SARS-CoV-2, the N-terminal ExoN domain is anticipated to have proofreading capabilities, eliminating incorrectly matched nucleotides inserted by the RdRp. When NSP10 acts as a cofactor, it combines with ExoN to form an NSP10-NSP14 complex, which boosts ExoN activity. (Low et al., 2022).

Exonuclease Activity (ExoN Domain): The ExoN domain of NSP14 serves as a proofreading enzyme, rectifying errors made by the viral RNA polymerase during replication. It removes nucleotides from the 3' end of RNA molecules, enhancing the fidelity of viral RNA replication.

N7-Methyltransferase Activity (N7-MTase Domain): The N7-MTase domain is responsible for methylating the 5' cap structure of viral RNA. This modification is essential for evading the host immune response and ensuring efficient translation of viral proteins.

The ExoN domain contains metal-binding residues critical for its exonuclease activity. D90, D91, E191, and D273: These residues coordinate divalent metal ions (e.g., zinc or magnesium) necessary for catalysis. D331 and D333 residues are part of the ExoN active site and are involved in catalyzing the removal of incorrect nucleotides. H268 helps to position the RNA substrate for catalysis.

The C-terminal N7-Methyltransferase Domain (N7-MTase): it has the SAM Binding Pocket: K58, K61, R95, R149, and N166: these residues are involved in binding to SAM, the methyl

donor molecule. It also has a cap-binding pocket composed by residues W689, H710, and F713: These residues are involved in recognizing and binding the 5' cap structure of RNA.

With reference to SARS-CoV-2, the DDED motif is crucial for the exoribonuclease activity of the non-structural protein 14 (NSP14). The two aspartate (D) residues followed by a glutamate (E) residue and another aspartate (D) residue play a crucial role in coordinating metal ions required for the catalytic activity of the protein. Substituting one or more of these aspartate residues with alanine as done in previous research disrupted the coordination of metal ions necessary for the enzymatic activity. Alanine is a smaller, non-polar amino acid compared to aspartate, which is polar and negatively charged. This substitution could alter the protein's structure and did affect its ability to bind metal ions effectively, leading to reduced or loss of enzymatic activity. As a result, the exoribonuclease activity of NSP14 was impaired, which could impact viral replication and survival.

To assess the effect of mutations, D222Y and D324Y on the function of XBB and omicron NSP14 (Table 1.0), we model the NSP14 of both variants and compare them to that of the wildtype reference strain. We found out that the mutations did not result in a conformational change as expected and this was confirmed by the executive RMSD value of 0.330 for both mutations and variants.

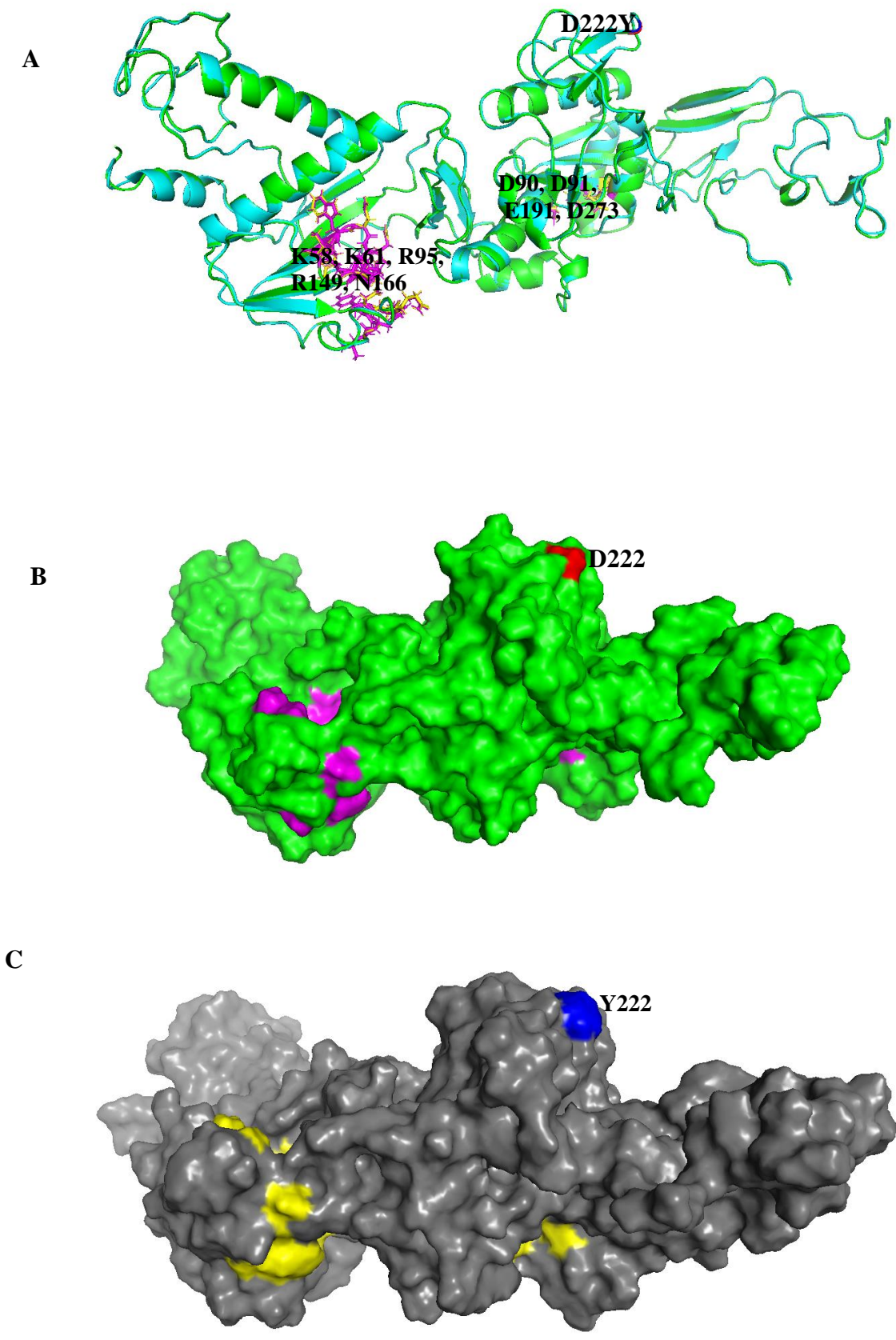


Figure 10.0: Homology model of NSP14 wildtype and D222Y. (A) NSP14 wildtype and mutant superimposed. Color code, Green represents the wildtype and cyan, the mutant. Magenta are the critical RNA and Zinc binding sites on NSP14 wildtype, yellow same on mutant (B) Surface

Structure of wildtype showing RNA and Zinc binding sites and wildtype amino acid of interest. (C) Surface structure of mutant showing RNA and Zinc binding sites and mutated amino acid in blue.

Tyrosine is a polar, uncharged amino acid with a big aromatic side chain, whereas aspartic acid is a negatively charged amino acid with a carboxylate group in its side chain. Although both amino acids differ in their properties, the 222nd position is located in a region of the protein that is not critical for its function or does not participate in key interactions, a mutation at this position may not have a substantial impact on the protein's structure or function. As discussed previously, some positions in a protein can tolerate a variety of amino acid substitutions without significant changes to the protein's structure or function. This is often the case for positions that are not directly involved in the protein's active site or in key structural elements. While aspartic acid and tyrosine have different charges, they are similar in size. Therefore, a mutation might not cause significant steric hindrance or dramatic changes in the protein's overall shape. However, it's important to note that even if a mutation does not significantly change the protein's structure, it could still affect its function. For instance, if the 222nd position is involved in interactions with other molecules, a mutation could alter these interactions and thus affect the protein's function.

4.5 Homology Modeling of NSP15

NSP15 is a uridine-specific endoribonuclease composed of 347 amino acids, with three domains: N-terminal domain (NTD), middle domain (MD), and a C-terminal domain (CTD) (Wilson et al., 2022a). NSP15 forms hexameric structures, with the N-terminal domain (NTD) contributing to stability, the C-terminal domain (CTD) housing the catalytic site, and the middle domain (MD) acting as a connector. Its function involves processing viral RNA, particularly by cleaving RNA substrates at the 3' end of uridines, thus assisting the virus in

avoiding the innate immune system's identification (Justo Arevalo et al., 2023a). NSP15 functions as a broad-spectrum endoribonuclease, primarily targeting its cleavage sites through the recognition of a single uridine (Wilson et al., 2022a). Its nuclease activity is vital for preventing the buildup of viral double-stranded RNA.

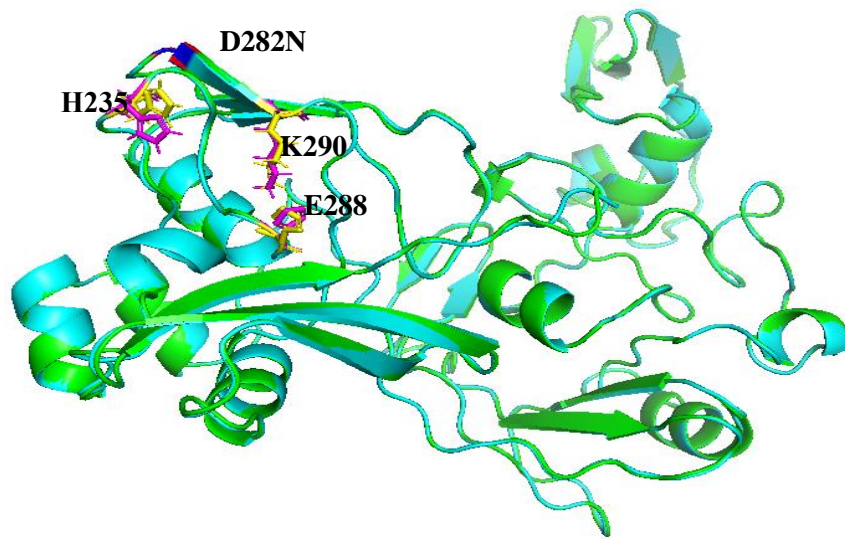
Endoribonuclease Activity (NendoU Domain): NSP15's primary function is to cleave viral RNA at specific sites to evade the host immune response and facilitate viral replication. The NendoU domain is specialized in cleaving double-stranded RNA (dsRNA) and potentially some single-stranded RNA (ssRNA) structures. The catalytic residues are H235, which is a key catalytic residue, acting as a general acid-base during the cleavage of RNA. The K290 and E288 residues are also involved in catalysis and stabilization of the transition state. The H235, K290, and E288 form the catalytic triad essential for endoribonuclease activity.

RNA-binding residues: R17, R19, R21, K22, R37, K38, R39, and R50 are involved in RNA binding and substrate recognition.

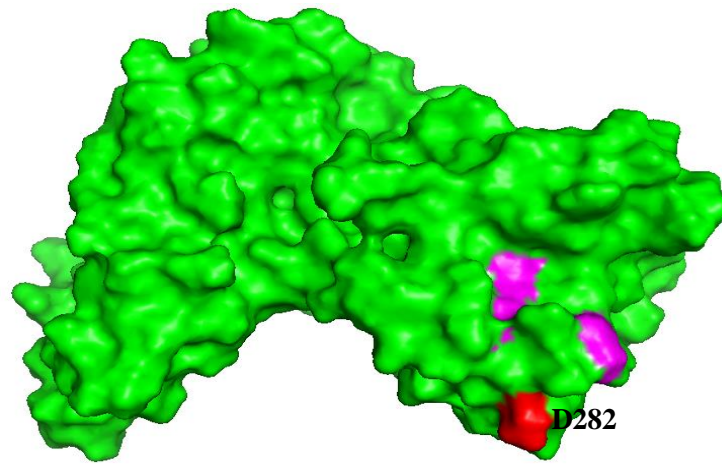
Metal-binding residues: H235, H250, and H263 are involved in coordinating divalent metal ions (e.g., zinc or magnesium) necessary for catalysis.

The effect of the mutation, D282N was assessed on the structure of NSP15 in the Omicron strain. Structural comparison revealed that the mutant NSP and the reference were similar without a conformational change. This was confirmed by the root mean square deviation value of 0.290.

A



C



C

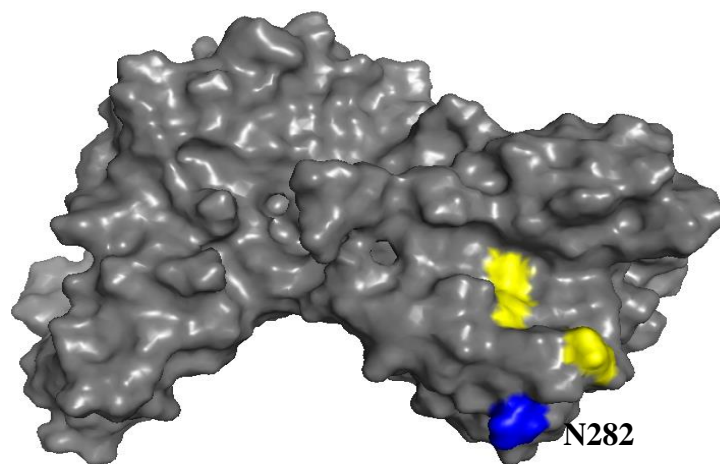


Figure 11.0: Homology model of NSP15 wildtype and D282N. (A) NSP15 wildtype and mutant superimposed. Color code, Green represents the wildtype and cyan, the mutant. Magenta are the critical RNA binding sites on NSP15 wildtype, yellow same on mutant (B) Surface Structure of wildtype showing RNA binding sites and wildtype amino acid of interest. (C) Surface Structure of mutant showing RNA binding sites and mutated amino acid in blue.

Aspartic acid (D) and asparagine (N) share some similarities in their chemical properties. Both are polar amino acids, with asparagine also having an amide group in its side chain. Asparagine can form hydrogen bonds and participate in interactions similar to aspartic acid. The 282nd position in a protein may not be directly involved in critical structural elements or active sites. If this position occurs within a loop or a flexible region of the protein, mutations in this area may not substantially affect the protein's overall structure.

Summary:

Mutations in NSP1, specifically G49S and K47R, occur in the N-terminal region, which is essential for binding to the 40S ribosomal subunit and obstructing the mRNA entry channel. As a result, there is reduction in translation and a rise in viral replication. Interestingly, the structures of the mutated NSP1 variants did not exhibit any conformational changes, suggesting that these mutations may not impact the protein's ability to bind to the ribosomal subunit and inhibit translation (Justo Arevalo et al., 2023b). However, the mutations in NSP2, NSP13, NSP14, and NSP15 did not result in significant conformational changes, suggesting that the structure or functionality of these proteins may not be greatly impacted by these changes.

The RMSD value is a crucial metric in homology modeling, as it enables the comparison protein structures and facilitates the detection of conformational changes that may be triggered by mutations (Kufareva & Abagyan, 2012). The significance of RMSD values in homology modeling is further demonstrated in a study by (Mirza & Froeyen, 2020), which examined the structure and operation of the SARS-CoV-2 RNA-dependent RNA polymerase using homology modeling (RdRp, NSP12). The study found that the RdRp contains conserved motifs

A-G, with motif C being essential for its polymerase activity. The RMSD value was used to compare the structures of the RdRp from different coronaviruses, revealing that the SARS-CoV-2 RdRp shares a high degree of structural similarity with other coronaviruses.

In this study, the RMSD values obtained for NSP1 (0.542 and 0.589) were close to 1, suggesting that the mutations G49S and K47R may not have functional consequences on the protein's ability to bind to the 40S ribosomal subunit and inhibit translation. On the other hand, the RMSD values for the mutations in NSP2, NSP13, NSP14, and NSP15 ranged from 0.260 to 0.277, which suggests that these mutations do not substantial effect the overall structure of these proteins, as RMSD values below 1.0 signifies a significant level of structural similarity. It is crucial to emphasize, however, that even small changes in protein structure can have functional consequences, such as altering protein-protein interactions or enzymatic activity (Vihinen, 2021). In the case of NSP2, the mutations did not result in a conformational change in the critical binding regions (C18, C51, H54, H56, and G118), suggesting that the structure or function of this protein may not be greatly affected by these changes. However, to completely understand the effects of these mutations on NSP2 function, more research is necessary. Similarly, the mutations in NSP13 did not result in a conformational change in the structure or the critical binding sites, showing that these mutations may not necessarily impact the structure or function of this protein. However, NSP13 plays a role in the initial phase of RNA capping by hydrolyzing the γ -phosphate group of the 5'-terminus of viral RNA. This action helps in shielding the viral RNA from identification by the host immune system. Therefore, even small changes in NSP13 function could have significant implications for viral replication and pathogenesis (Justo Arevalo et al., 2023b).

The mutations in NSP14, specifically the D222Y and D324Y variants, did not result in any conformational changes as expected, as evidenced by the RMSD value of 0.330 for both. NSP14 is a dual-function enzyme present in *Coronaviridae*, comprising ExoN and a C-terminal

N7-methyltransferase domains. The ExoN domain functions as a proofreading exoribonuclease, enhancing RNA fidelity by eliminating mismatched nucleotides during RNA synthesis. In contrast, the N7-MTase domain is involved in capping activity, utilizing SAM to create a cap 0 structure (Justo Arevalo et al., 2023b). Any changes in NSP14 function could thus have significant implications for viral replication and pathogenesis.

Similarly, the mutation in NSP15, specifically the D282N variation, did not result in any conformational change, as indicated by the RMSD value of 0.290. NSP15 is an endoribonuclease specific to uridine, consisting of 347 amino acids and comprising three domains: the N-terminal domain (NTD), middle domain (MD), and C-terminal domain (CTD). It is responsible for processing viral RNA, with a particular ability to cleave RNA substrates at the 3' end of uridines, thus helping the virus in avoiding detection by the innate immune system. Any changes in NSP15 function could thus have significant implications for viral replication and pathogenesis (Wilson et al., 2022b).

It is important to note that mutations can also affect protein function through other mechanisms, such as altering protein-protein interactions or enzymatic activity. Consequently, further study is needed to completely understand the impacts of these mutations on the functionality of these proteins.

CHAPTER FIVE

Conclusions and Recommendations

In this study, a comprehensive computational approach is proposed to examine how mutations influence the NSPs of SARS-CoV-2. By combining sequence analysis, comparative analysis and structure prediction and visualization, the aim is to elucidate the structural and functional consequences of these mutations and their potential impact on viral replication and pathogenesis.

The research plan involved:

- Obtaining and aligning the sequences of wild-type and mutated NSPs using bioinformatics tools.
- Comparing the physicochemical properties, conserved domains, and functional motifs of the NSPs to identify potential changes induced by the mutations.
- Utilizing protein structure prediction tools to model the 3D structures of the NSPs and assess the impact of mutations on their structural integrity.

The hypothesis posits that the specific mutations in the NSPs will enhance their functions related to viral replication and transcription, and the alterations in structure and function resulting from these mutations will facilitate enhanced viral replication.

While significant attention has been directed towards mutations in the structural proteins of SARS-CoV-2, notably the spike protein, this study aimed to emphasize the importance of investigating the NSPs. These proteins are essential for key functions in the viral life cycle, and mutations in these proteins can have significant implications for viral replication, pathogenesis, and evolution (Pachetti et al., 2020).

Understanding the impacts of mutations on the structure and function of NSPs is pivotal for several reasons:

- Identifying novel antiviral targets: By elucidating the structural and functional consequences of NSP mutations, new targets for antiviral therapies that are less likely to be affected by viral evolution may be identified (Shin et al., 2020).
- Predicting viral evolution: Characterizing the adaptive mutations in NSPs can help predict the emergence of increasingly transmissible or virulent SARS-CoV-2 strains, enabling proactive public health measures (Plante et al., 2021).
- Improving diagnostic and surveillance tools: Incorporating knowledge of NSP mutations into the design of diagnostic tests and surveillance strategies can enhance their accuracy and effectiveness in detecting and monitoring emerging SARS-CoV-2 variants (Wang et al., 2021).

The results outline in this study do not corroborate the proposed hypothesis. Should the hypotheses be validated, and if the identified mutations within the NSPs are indeed discovered to bolster viral replication and transcription, the outcomes of this study could carry significant implications for combatting emerging SARS-CoV-2 variants:

- Informing vaccine design: Understanding the structural and functional impacts of NSP mutations can direct the design of vaccines that are better equipped to withstand viral evolution. By targeting conserved regions of the NSPs or designing multivalent vaccines that account for potential mutations, the long-term efficacy of vaccination strategies may be improved (Dai & Gao, 2021).
- Developing targeted antivirals: Identifying the specific mechanisms by which NSP mutations enhance viral replication can facilitate the development of targeted antiviral compounds that inhibit these processes. By focusing on the most conserved and functionally important regions of the NSPs, antivirals that are less susceptible to resistance mutations may be developed (Zhu et al., 2021).

- Enhancing pandemic preparedness: Analyzing the adaptive mutations within SARS-CoV-2 NSPs can enhance our comprehension of the virus's evolutionary capabilities and guide the development of strategies for pandemic readiness. Anticipating the appearance of new variants that exhibit heightened replication or transmission abilities enables the creation of proactive and efficient public health interventions. (Kaushal et al., 2020).

Following the completion of structural modeling, our intended research trajectory involved conducting molecular dynamics simulations. However, this phase was not feasible within the allocated time frame. To strengthen our understanding of these findings, it will be crucial to validate them using additional de novo homology modeling tools. These tools can provide alternative perspectives and enhance the reliability of our conclusions.

Molecular dynamics (MD) simulations should be conducted to investigate the dynamic behavior of both wild-type and mutated NSPs. These simulations will offer valuable insights into the conformational flexibility, stability, and interactions caused by the mutations. The results will provide valuable insights into the structural basis of the functional alterations and guide further experimental. Moreover, while some mutations may not visibly alter the structure of the protein, they could still impact NSP function. To investigate this further, conducting assays to directly test the functionality of the protein will be essential. These assays will help uncover subtle yet significant changes in protein behavior that may influence viral infectivity and response to treatment.

In summary, the objective of this research is to offer a thorough comprehension of how mutations impact the structure and functionality of SARS-CoV-2 NSPs. By integrating computational approaches, the goal is to elucidate the mechanisms by which these mutations enhance viral replication and lead to the evolution and pathogenesis of SARS-CoV-2. The

findings of this study have the potential to inform the development of more effective vaccines, antivirals, and public health initiatives to counter the ongoing COVID-19 pandemic and subsequent outbreaks of coronavirus.

References

- Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., & Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1, 19-25.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), 403-410.
- Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., De Castro, E., ... & Stockinger, H. (2012). ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Research*, 40(W1), W597-W603.
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., ... & Sherlock, G. (2000). Gene ontology: tool for the unification of biology. *Nature Genetics*, 25(1), 25-29.
- Attwood, T. K., Bradley, P., Flower, D. R., Gaulton, A., Maudling, N., Mitchell, A. L., ... & Zygouri, C. (2003). PRINTS and its automatic supplement, prePRINTS. *Nucleic Acids Research*, 31(1), 400-402.
- Attwood, T. K., Coletta, A., Muirhead, G., Pavlopoulou, A., Philippou, P. B., Popov, I., ... & Mitchell, A. L. (2012). The PRINTS database: a fine-grained protein sequence annotation and analysis resource—its status in 2012. *Database*, 2012.
- Bambra, C., Riordan, R., Ford, J., & Matthews, F. (2020). The COVID-19 pandemic and health inequalities. *Journal of Epidemiology and Community Health*, 74(11), 964-968.
- Banerjee, A. K., Blanco, M. R., Bruce, E. A., Honson, D. D., Chen, L. M., Chow, A., ... & Quinlan, B. D. (2020). SARS-CoV-2 disrupts splicing, translation, and protein trafficking to suppress host defenses. *Cell*, 183(5), 1325-1339.e21.
- Benkert, P., Biasini, M., & Schwede, T. (2011). Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics*, 27(3), 343-350.

Benson, N. C., & Daggett, V. (2012). A comparison of multiscale methods for the analysis of molecular dynamics simulations. *The Journal of Physical Chemistry B*, 116(29), 8722-8731.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., ... & Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Research*, 28(1), 235-242.

Bestle, D., Heindl, M. R., Limburg, H., Van Lam van, T., Pilgram, O., Moulton, H., ... & Böttcher-Friebertshäuser, E. (2020). TMPRSS2 and furin are both essential for proteolytic activation of SARS-CoV-2 in human airway cells. *Life Science Alliance*, 3(9), e202000786.

Blanco-Melo, D., Nilsson-Payant, B. E., Liu, W. C., Uhl, S., Hoagland, D., Møller, R., ... & tenOever, B. R. (2020). Imbalanced host response to SARS-CoV-2 drives development of COVID-19. *Cell*, 181(5), 1036-1045.e9.

Bloom, J. D., Gong, L. I., & Baltimore, D. (2010). Permissive secondary mutations enable the evolution of influenza oseltamivir resistance. *Science*, 328(5983), 1272-1275.

Bowie, J. U., Lüthy, R., & Eisenberg, D. (1991). A method to identify protein sequences that fold into a known three-dimensional structure. *Science*, 253(5016), 164-170.

Choi, A., Koch, M., Wu, K., Dixon, G., Oestreicher, J., Legault, H., ... & Edwards, D. K. (2021). Serum neutralizing activity of mRNA-1273 against SARS-CoV-2 variants. *bioRxiv*. <https://doi.org/10.1101/2021.06.28.449914>

Cock, P. J., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., ... & De Hoon, M. J. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, 25(11), 1422-1423.

Cornillez-Ty, C. T., Liao, L., Yates, J. R., Kuhn, P., & Buchmeier, M. J. (2009). Severe acute respiratory syndrome coronavirus nonstructural protein 2 interacts with a host protein complex involved in mitochondrial biogenesis and intracellular signaling. *Journal of Virology*, 83(19), 10314-10318.

- Dai, L., & Gao, G. F. (2021). Viral targets for vaccines against COVID-19. *Nature Reviews Immunology*, 21(2), 73-82.
- Darden, T., York, D., & Pedersen, L. (1993). Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *The Journal of Chemical Physics*, 98(12), 10089-10092.
- Davies, N. G., Abbott, S., Barnard, R. C., Jarvis, C. I., Kucharski, A. J., Munday, J. D., ... & Edmunds, W. J. (2021). Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science*, 372(6538), eabg3055.
- Decroly, E., Debarnot, C., Ferron, F., Bouvet, M., Coutard, B., Imbert, I., ... & Canard, B. (2011). Crystal structure and functional analysis of the SARS-coronavirus RNA cap 2'-O-methyltransferase NSP10/NSP16 complex. *PLoS Pathogens*, 7(5), e1002059.
- Deng, X., Hackbart, M., Mettelman, R. C., O'Brien, A., Mielech, A. M., Yi, G., ... & Baker, S. C. (2019). Coronavirus nonstructural protein 15 mediates evasion of dsRNA sensors and limits apoptosis in macrophages. *Proceedings of the National Academy of Sciences*, 116(19), 9433-9438.
- Devaux, C. A., Rolain, J. M., & Raoult, D. (2020). ACE2 receptor polymorphism: Susceptibility to SARS-CoV-2, hypertension, multi-organ failure, and COVID-19 disease outcome. *Journal of Microbiology, Immunology and Infection*, 53(3), 425-435.
- El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., ... & Finn, R. D. (2019). The Pfam protein families database in 2019. *Nucleic Acids Research*, 47(D1), D427-D432.
- Fajnzylber, J., Regan, J., Coxen, K., Corry, H., Wong, C., Rosenthal, A., ... & Li, J. Z. (2020). SARS-CoV-2 viral load is associated with increased disease severity and mortality. *Nature Communications*, 11(1), 1-9.
- Fehr, A. R., & Perlman, S. (2015). Coronaviruses: an overview of their replication and pathogenesis. *Coronaviruses*, 1-23.

Fernandes, J. D., Faust, T. B., Strauli, N. B., Smith, C., Crosby, D. C., Nakamura, R. L., ... & Frankel, A. D. (2016). Functional segregation of overlapping genes in HIV. *Cell*, 167(7), 1762-1773.

Ferron, F., Subissi, L., De Morais, A. T. S., Le, N. T. T., Sevajol, M., Gluais, L., ... & Imbert, I. (2018). Structural and molecular basis of mismatch correction and ribavirin excision from coronavirus RNA. *Proceedings of the National Academy of Sciences*, 115(2), E162-E171.

Forni, D., Cagliani, R., Clerici, M., & Sironi, M. (2017). Molecular evolution of human coronavirus genomes. *Trends in Microbiology*, 25(1), 35-48.

Gao, Y., Yan, L., Huang, Y., Liu, F., Zhao, Y., Cao, L., ... & Rao, Z. (2020). Structure of the RNA-dependent RNA polymerase from COVID-19 virus. *Science*, 368(6492), 779-782.

Garcia-Beltran, W. F., Lam, E. C., St. Denis, K., Nitido, A. D., Garcia, Z. H., Hauser, B. M., ... & Balazs, A. B. (2021). Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity. *Cell*, 184(9), 2372-2383.e9.

Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server. *The Proteomics Protocols Handbook*, 571-607.

Ghosh, S., Dellibovi-Ragheb, T. A., Kerviel, A., Pak, E., Qiu, Q., Fisher, M., ... & Altan-Bonnet, N. (2020). β -Coronaviruses use lysosomes for egress instead of the biosynthetic secretory pathway. *Cell*, 183(6), 1520-1535.

Gowers, R. J., Linke, M., Barnoud, J., Reddy, T. J. E., Melo, M. N., Seyler, S. L., ... & Beckstein, O. (2016). MDAnalysis: a Python package for the rapid analysis of molecular dynamics simulations. *Proceedings of the 15th Python in Science Conference*, 98-105.

Greaney, A. J., Loes, A. N., Crawford, K. H., Starr, T. N., Malone, K. D., Chu, H. Y., & Bloom, J. D. (2021). Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding

domain that affect recognition by polyclonal human plasma antibodies. *Cell Host & Microbe*, 29(3), 463-476.e6.

Greaney, A. J., Starr, T. N., Gilchuk, P., Zost, S. J., Binshtein, E., Loes, A. N., ... & Bloom, J. D. (2021). Complete mapping of mutations to the SARS-CoV-2 spike receptor-binding domain that escape antibody recognition. *Cell Host & Microbe*, 29(1), 44-57.e9.

Grubaugh, N. D., Hanage, W. P., & Rasmussen, A. L. (2020). Making sense of mutation: what D614G means for the COVID-19 pandemic remains unclear. *Cell*, 182(4), 794-795.

Guan, W. J., Liang, W. H., Zhao, Y., Liang, H. R., Chen, Z. S., Li, Y. M., ... & He, J. X. (2020). Comorbidity and its impact on 1590 patients with COVID-19 in China: a nationwide analysis. *European Respiratory Journal*, 55(5), 2000547.

Hadfield J., Megill C., Bell S.M., Huddleston J., Potter B., Callender C., Sagulenko P., Bedford T., Neher R. A., Nextstrain: real-time tracking of pathogen evolution, *Bioinformatics*, Volume 34, Issue 23, December 2018, Pages 4121–4123, <https://doi.org/10.1093/bioinformatics/bty407>

Hagemeyer, M. C., Monastyrska, I., Griffith, J., van der Sluijs, P., Voortman, J., van Bergen en Henegouwen, P. M., ... & de Haan, C. A. (2014). Membrane rearrangements mediated by coronavirus nonstructural proteins 3 and 4. *Virology*, 458, 125-135.

Harvey, W. T., Carabelli, A. M., Jackson, B., Gupta, R. K., Thomson, E. C., Harrison, E. M., ... & Robertson, D. L. (2021). SARS-CoV-2 variants, spike mutations and immune escape. *Nature Reviews Microbiology*, 19(7), 409-424.

Hatcher, E. L., Zhdanov, S. A., Bao, Y., Blinkova, O., Nawrocki, E. P., Ostapchuck, Y., ... & Brister, J. R. (2017). Virus Variation Resource—improved response to emergent viral outbreaks. *Nucleic Acids Research*, 45(D1), D482-D490.

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., ... & Pöhlmann, S. (2020). SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell*, 181(2), 271-280.

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

Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., de Groot, B. L., ... & MacKerell, A. D. (2017). CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature Methods*, 14(1), 71-73.

Hulo, N., Bairoch, A., Bulliard, V., Cerutti, L., De Castro, E., Langendijk-Genevaux, P. S., ... & Sigrist, C. J. (2006). The PROSITE database. *Nucleic Acids Research*, 34(suppl_1), D227-D230.

Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: visual molecular dynamics. *Journal of Molecular Graphics*, 14(1), 33-38.

Jang, K. J., Jeong, S., Kang, D. Y., Sp, N., Yang, Y. M., & Kim, D. E. (2020). A high ATP concentration enhances the cooperative translocation of the SARS coronavirus helicase NSP13 in the unwinding of duplex RNA. *Scientific Reports*, 10(1), 1-13.

Johnson, B. A., Xie, X., Bailey, A. L., Kalveram, B., Lokugamage, K. G., Muruato, A., ... & Menachery, V. D. (2021). Loss of furin cleavage site attenuates SARS-CoV-2 pathogenesis. *Nature*, 591(7849), 293-299.

Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., ... & Hunter, S. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics*, 30(9), 1236-1240.

Kabsch, W., & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers: Original Research on Biomolecules*, 22(12), 2577-2637.

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, 30(4), 772-780.

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

Kim, D., Lee, J. Y., Yang, J. S., Kim, J. W., Kim, V. N., & Chang, H. (2020). The architecture of SARS-CoV-2 transcriptome. *Cell*, 181(4), 914-921.

Kim, Y., Jedrzejczak, R., Maltseva, N. I., Wilamowski, M., Endres, M., Godzik, A., ... & Joachimiak, A. (2020). Crystal structure of NSP15 endoribonuclease NendoU from SARS-CoV-2. *Protein Science*, 29(7), 1596-1605.

Kirkwood, J. G. (1935). Statistical mechanics of fluid mixtures. *The Journal of Chemical Physics*, 3(5), 300-313.

Klein, S., Cortese, M., Winter, S. L., Wachsmuth-Melm, M., Neufeldt, C. J., Cerikan, B., ... & Chlanda, P. (2020). SARS-CoV-2 structure and replication characterized by in situ cryo-electron tomography. *Nature Communications*, 11(1), 1-10.

Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., ... & Cheatham, T. E. (2000). Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Accounts of Chemical Research*, 33(12), 889-897.

Korber, B., Fischer, W. M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., ... & Montefiori, D. C. (2020). Tracking changes in SARS-CoV-2 spike: Evidence that D614G increases infectivity of the COVID-19 virus. *Cell*, 182(4), 812-827.e19.

Korneeva, N., Khalil, M. I., Ghosh, I., Fan, R., Arnold, T., & De Benedetti, A. (2023). SARS-CoV-2 viral protein Nsp2 stimulates translation under normal and hypoxic conditions. *Virology Journal*, 20(1), 55. <https://doi.org/10.1186/s12985-023-02021-2>.

- Lam, T. T. Y., Jia, N., Zhang, Y. W., Shum, M. H. H., Jiang, J. F., Zhu, H. C., ... & Cao, W. C. (2020). Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. *Nature*, 583(7815), 282-285.
- Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, 26(2), 283-291.
- Lei, J., Kusov, Y., & Hilgenfeld, R. (2018). Nsp3 of coronaviruses: Structures and functions of a large multi-domain protein. *Antiviral Research*, 149, 58–74. <https://doi.org/10.1016/j.antiviral.2017.11.001>.
- Leung, K., Shum, M. H., Leung, G. M., Lam, T. T., & Wu, J. T. (2021). Early transmissibility assessment of the N501Y mutant strains of SARS-CoV-2 in the United Kingdom, October to November 2020. *Eurosurveillance*, 26(1), 2002106.
- Liu, Y., Gayle, A. A., Wilder-Smith, A., & Rocklöv, J. (2020). The reproductive number of COVID-19 is higher compared to SARS coronavirus. *Journal of Travel Medicine*, 27(2), taaa021.
- Low, Z. Y., Zabidi, N. Z., Yip, A. J. W., Puniyamurti, A., Chow, V. T. K., & Lal, S. K. (2022). SARS-CoV-2 Non-Structural Proteins and Their Roles in Host Immune Evasion. *Viruses*, 14(9), 1991. <https://doi.org/10.3390/v14091991>.
- Maghrabi, A. H., & McGuffin, L. J. (2017). ModFOLD6: an accurate web server for the global and local quality estimation of 3D protein models. *Nucleic Acids Research*, 45(W1), W416-W421.
- Mehta, P., McAuley, D. F., Brown, M., Sanchez, E., Tattersall, R. S., & Manson, J. J. (2020). COVID-19: consider cytokine storm syndromes and immunosuppression. *The Lancet*, 395(10229), 1033-1034.

- Mendez, A. S., Ly, M., González-Sánchez, A. M., Hartenian, E., Ingolia, N. T., Cate, J. H., & Glaunsinger, B. A. (2021). The N-terminal domain of SARS-CoV-2 nsp1 plays key roles in suppression of cellular gene expression and preservation of viral gene expression. *Cell Reports*, 37(3), 109841. <https://doi.org/10.1016/j.celrep.2021.109841>.
- Moghadas, S. M., Fitzpatrick, M. C., Sah, P., Pandey, A., Shoukat, A., Singer, B. H., & Galvani, A. P. (2020). The implications of silent transmission for the control of COVID-19 outbreaks. *Proceedings of the National Academy of Sciences*, 117(30), 17513-17515.
- Moult, J., Fidelis, K., Kryshtafovych, A., Schwede, T., & Tramontano, A. (2018). Critical assessment of methods of protein structure prediction (CASP)—Round XII. *Proteins: Structure, Function, and Bioinformatics*, 86, 7-15.
- Narayanan, K., Ramirez, S. I., Lokugamage, K. G., & Makino, S. (2015). Coronavirus nonstructural protein 1: Common and distinct functions in the regulation of host and viral gene expression. *Virus Research*, 202, 89-100.
- Neuman, B. W., Kiss, G., Kunding, A. H., Bhella, D., Baksh, M. F., Connelly, S., ... & Buchmeier, M. J. (2011). A structural analysis of M protein in coronavirus assembly and morphology. *Journal of Structural Biology*, 174(1), 11-22.
- Orengo, C. A., & Thornton, J. M. (2005). Protein families and their evolution—a structural perspective. *Annual Review of Biochemistry*, 74, 867-900.
- Oude Munnink, B. B., Sikkema, R. S., Nieuwenhuijse, D. F., Molenaar, R. J., Munger, E., Molenkamp, R., ... & Koopmans, M. P. (2021). Transmission of SARS-CoV-2 on mink farms between humans and mink and back to humans. *Science*, 371(6525), 172-177.
- Pachetti, M., Marini, B., Benedetti, F., Giudici, F., Mauro, E., Storici, P., ... & Ippodrino, R. (2020). Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. *Journal of Translational Medicine*, 18(1), 1-9.

- Pearson, W. R. (2013). An introduction to sequence similarity ("homology") searching. *Current Protocols in Bioinformatics*, 42(1), 3-1.
- Plante, J. A., Mitchell, B. M., Plante, K. S., Debbink, K., Weaver, S. C., & Menachery, V. D. (2021). The variant gambit: COVID-19's next move. *Cell Host & Microbe*, 29(4), 508-515.
- Reggiori, F., Monastyrska, I., Verheije, M. H., Calì, T., Ulasli, M., Bianchi, S., ... & Molinari, M. (2010). Coronaviruses hijack the LC3-I-positive EDEMosomes, ER-derived vesicles exporting short-lived ERAD regulators, for replication. *Cell Host & Microbe*, 7(6), 500-508.
- Robson, F., Khan, K. S., Le, T. K., Paris, C., Demirbag, S., Barfuss, P., ... & Ng, W. L. (2020). Coronavirus RNA proofreading: Molecular basis and therapeutic targeting. *Molecular Cell*, 79(5), 710-727.
- Rozewicki, J., Li, S., Amada, K. M., Standley, D. M., & Katoh, K. (2019). MAFFT-DASH: integrated protein sequence and structural alignment. *Nucleic Acids Research*, 47(W1), W5-W10.
- Sahu, U., Biswas, D., Singh, A. K., & Khare, P. (2021). Mechanism involved in the pathogenesis and immune response against SARS-CoV-2 infection. *VirusDisease*, 32(2), 211–219. <https://doi.org/10.1007/s13337-021-00687-2>.
- Sammond, D. W., Eletr, Z. M., Purbeck, C., Kimple, R. J., Siderovski, D. P., & Kuhlman, B. (2007). Structure-based protocol for identifying mutations that enhance protein-protein binding affinities. *Journal of Molecular Biology*, 371(5), 1392-1404.
- Saunders, K. O., Lee, E., Parks, R., Martinez, D. R., Li, D., Chen, H., ... & Haynes, B. F. (2021). Neutralizing antibody vaccine for pandemic and pre-emergent coronaviruses. *Nature*, 594(7864), 553-559.
- Schoeman, D., & Fielding, B. C. (2019). Coronavirus envelope protein: current knowledge. *Virology Journal*, 16(1), 1-22.
- Schrödinger, LLC. (2021). The PyMOL Molecular Graphics System, Version 2.5.

- Schubert, K., Karousis, E. D., Jomaa, A., Scaiola, A., Echeverria, B., Gurzeler, L. A., ... & Ban, N. (2020). SARS-CoV-2 NSP1 binds the ribosomal mRNA channel to inhibit translation. *Nature Structural & Molecular Biology*, 27(10), 959-966.
- Shen, Z., Ye, G., Deng, F., Wang, G., Cui, M., Fang, L., ... & Xiao, S. (2021). Structural basis for the inhibition of host gene expression by porcine epidemic diarrhea virus NSP1. *Journal of Virology*, 95(5), e01896-20.
- Shin, D., Mukherjee, R., Grewe, D., Bojkova, D., Baek, K., Bhattacharya, A., ... & Dikic, I. (2020). Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity. *Nature*, 587(7835), 657-662.
- Shu, Y., & McCauley, J. (2017). GISAID: Global initiative on sharing all influenza data—from vision to reality. *Eurosurveillance*, 22(13), 30494.
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., ... & Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 7(1), 539.
- Sigrist, C. J., De Castro, E., Cerutti, L., Cucho, B. A., Hulo, N., Bridge, A., ... & Xenarios, I. (2013). New and continuing developments at PROSITE. *Nucleic Acids Research*, 41(D1), D344-D347.
- Siu, Y. L., Teoh, K. T., Lo, J., Chan, C. M., Kien, F., Escriou, N., ... & Nal, B. (2008). The M, E, and N structural proteins of the severe acute respiratory syndrome coronavirus are required for efficient assembly, trafficking, and release of virus-like particles. *Journal of Virology*, 82(22), 11318-11330.
- Snijder, E. J., Decroly, E., & Ziebuhr, J. (2016). The nonstructural proteins directing coronavirus RNA synthesis and processing. *Advances in Virus Research*, 96, 59-126.

Starr, T. N., Greaney, A. J., Dingens, A. S., & Bloom, J. D. (2021). Complete map of SARS-CoV-2 RBD mutations that escape the monoclonal antibody LY-CoV555 and its cocktail with LY-CoV016. *Cell Reports Medicine*, 2(4), 100255.

Starr, T. N., Greaney, A. J., Hilton, S. K., Ellis, D., Crawford, K. H., Dingens, A. S., ... & Bloom, J. D. (2020). Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. *Cell*, 182(5), 1295-1310.e20.

Sutton, G., Fry, E., Carter, L., Sainsbury, S., Walter, T., Nettleship, J., ... & Stuart, D. I. (2004). The NSP9 replicase protein of SARS-coronavirus, structure and functional insights. *Structure*, 12(2), 341-353.

The Gene Ontology Consortium. (2019). The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Research*, 47(D1), D330-D338.

Thoms, M., Buschauer, R., Ameismeier, M., Koepke, L., Denk, T., Hirschenberger, M., ... & Beckmann, R. (2020). Structural basis for translational shutdown and immune evasion by the NSP1 protein of SARS-CoV-2. *Science*, 369(6508), 1249-1255.

Thorne, L. G., Bouhaddou, M., Reuschl, A. K., Zuliani-Alvarez, L., Polacco, B., Pelin, A., ... & Jolly, C. (2021). Evolution of enhanced innate immune evasion by SARS-CoV-2. *Nature*, 600(7889), 487-493.

Trott, O., & Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, 31(2), 455-461.

Twohig, K. A., Nyberg, T., Zaidi, A., Thelwall, S., Sinnathamby, M. A., Aliabadi, S., ... & Dabrera, G. (2021). Hospital admission and emergency care attendance risk for SARS-CoV-2 delta (B.1.617.2) compared with alpha (B.1.1.7) variants of concern: a cohort study. *The Lancet Infectious Diseases*, 22(1), 35-42.

Ujike, M., Huang, C., Shirato, K., Makino, S., & Taguchi, F. (2016). The contribution of the cytoplasmic retrieval signal of severe acute respiratory syndrome coronavirus to intracellular accumulation of S proteins and incorporation of S protein into virus-like particles. *Journal of General Virology*, 97(8), 1853-1864.

V'kovski, P., Kratzel, A., Steiner, S., Stalder, H., & Thiel, V. (2021). Coronavirus biology and replication: implications for SARS-CoV-2. *Nature Reviews Microbiology*, 19(3), 155-170.

Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., & Berendsen, H. J. (2005). GROMACS: fast, flexible, and free. *Journal of Computational Chemistry*, 26(16), 1701-1718.

Vogels, C. B., Breban, M. I., Ott, I. M., Alpert, T., Petrone, M. E., Watkins, A. E., ... & Grubaugh, N. D. (2021). Multiplex qPCR discriminates variants of concern to enhance global surveillance of SARS-CoV-2. *PLoS Biology*, 19(5), e3001236.

V'kovski, P., Kratzel, A., Steiner, S., Stalder, H., & Thiel, V. (2021). Coronavirus biology and replication: implications for SARS-CoV-2. *Nature Reviews Microbiology*, 19(3), 155-170.

Walia, R. R., Xue, L. C., Wilkins, K., El-Manzalawy, Y., Dobbs, D., & Honavar, V. (2014). RNABindRPlus: a predictor that combines machine learning and sequence homology-based methods to improve the reliability of predicted RNA-binding residues in proteins. *PLoS One*, 9(5), e97725.

Walls, A. C., Park, Y. J., Tortorici, M. A., Wall, A., McGuire, A. T., & Veerler, D. (2020). Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell*, 181(2), 281-292.

Wang, K., Chen, W., Zhang, Z., Deng, Y., Lian, J. Q., Du, P., ... & Sun, B. (2020). CD147-spike protein is a novel route for SARS-CoV-2 infection to host cells. *Signal Transduction and Targeted Therapy*, 5(1), 1-10.

- Wang, L., Huang, C., Yang, M. Q., & Yang, J. Y. (2010). BindN+ for accurate prediction of DNA and RNA-binding residues from protein sequence features. *BMC Systems Biology*, 4(1), 1-9.
- Wang, P., Nair, M. S., Liu, L., Iketani, S., Luo, Y., Guo, Y., ... & Ho, D. D. (2021). Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. *Nature*, 593(7857), 130-135.
- Wang, R., Chen, J., Gao, K., Hozumi, Y., Yin, C., & Wei, G. W. (2020). Analysis of SARS-CoV-2 mutations in the United States suggests presence of four substrains and novel variants. *Communications Biology*, 4(1), 1-14.
- Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ (2009) Jalview Version 2 - A multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25 1189-1191. (doi:10.1093/bioinformatics/btp033).
- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., ... & Schwede, T. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research*, 46(W1), W296-W303.
- Webb, B., & Sali, A. (2016). Comparative protein structure modeling using MODELLER. *Current Protocols in Bioinformatics*, 54(1), 5-6.
- Wolff, G., Limpens, R. W., Zevenhoven-Dobbe, J. C., Laugks, U., Zheng, S., de Jong, A. W., ... & Bárcena, M. (2020). A molecular pore spans the double membrane of the coronavirus replication organelle. *Science*, 369(6509), 1395-1398.
- World Health Organization. (2023). WHO Coronavirus (COVID-19) Dashboard. Retrieved from <https://covid19.who.int/>
- Wu, C., Chen, X., Cai, Y., Xia, J., Zhou, X., Xu, S., ... & Song, Y. (2020). Risk factors associated with acute respiratory distress syndrome and death in patients with coronavirus disease 2019 pneumonia in Wuhan, China. *JAMA Internal Medicine*, 180(7), 934-943.

- Wu, F., Zhao, S., Yu, B., Chen, Y. M., Wang, W., Song, Z. G., ... & Zhang, Y. Z. (2020). A new coronavirus associated with human respiratory disease in China. *Nature*, 579(7798), 265-269.
- Wu, Z., & McGoogan, J. M. (2020). Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. *JAMA*, 323(13), 1239-1242.
- Xia, H., Cao, Z., Xie, X., Zhang, X., Chen, J. Y. C., Wang, H., ... & Liu, X. (2020). Evasion of type I interferon by SARS-CoV-2. *Cell Reports*, 33(1), 108234.
- Yang, D., & Leibowitz, J. L. (2015). The structure and functions of coronavirus genomic 3' and 5' ends. *Virus Research*, 206, 120-133.
- Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., & Zhang, Y. (2015). The I-TASSER Suite: protein structure and function prediction. *Nature Methods*, 12(1), 7-8.
- Yoshimoto, F. K. (2020). The proteins of severe acute respiratory syndrome coronavirus-2 (SARS CoV-2 or n-COV19), the cause of COVID-19. *The Protein Journal*, 39(3), 198-216.
- Zhao, K., Ke, Z., Hu, H., Liu, Y., Li, A., Hua, R., Guo, F., Xiao, J., Zhang, Y., Duan, L., Yan, X.-F., Gao, Y.-G., Liu, B., Xia, Y., & Li, Y. (2021). Structural Basis and Function of the N Terminus of SARS-CoV-2 Nonstructural Protein 1. *Microbiology Spectrum*, 9(1). <https://doi.org/10.1128/Spectrum.00169-21>.
- Zhao, Z., Li, H., Wu, X., Zhong, Y., Zhang, K., Zhang, Y. P., ... & Fu, Y. X. (2004). Moderate mutation rate in the SARS coronavirus genome and its implications. *BMC Evolutionary Biology*, 4(1), 1-9.
- Zhou, P., Yang, X. L., Wang, X. G., Hu, B., Zhang, L., Zhang, W., ... & Shi, Z. L. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, 579(7798), 270-273.

- Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., ... & Tan, W. (2020). A novel coronavirus from patients with pneumonia in China, 2019. *New England Journal of Medicine*, 382(8), 727-733.
- Zhu, W., Chen, C. Z., Gorshkov, K., Xu, M., Lo, D. C., & Zheng, W. (2021). RNA-dependent RNA polymerase as a target for COVID-19 drug discovery. *SLAS Discovery*, 25(10), 1141-1151.
- Ziebuhr, J., Snijder, E. J., & Gorbalenya, A. E. (2000). Virus-encoded proteinases and proteolytic processing in the Nidovirales. *Journal of General Virology*, 81(4), 853-879.
- Zou, L., Moch, C., Graille, M., & Chapat, C. (2022). The SARS-CoV-2 protein NSP2 impairs the silencing capacity of the human 4EHP-GIGYF2 complex. *IScience*, 25(7), 104646. <https://doi.org/10.1016/j.isci.2022.104646>.
- Zost, S. J., Gilchuk, P., Case, J. B., Binshtein, E., Chen, R. E., Nkolola, J. P., ... & Crowe, J. E. (2020). Potently neutralizing and protective human antibodies against SARS-CoV-2. *Nature*, 584(7821), 443-449.