T.R.

GEBZE TECHNICAL UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

INVESTIGATION OF SUBSTRATE RECOGNITION OF SARS-COV-2 MAIN PROTEASE ENZYME BY MOLECULAR DYNAMICS AND FREE ENERGY METHODS

HİLAL SENA TAŞÇI

A THESIS SUBMITTED FOR THE DEGREE OF MASTER OF SCIENCE MOLECULAR BIOLOGY AND GENETICS DEPARTMENT

GEBZE

2023

T.R.

GEBZE TECHNICAL UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

INVESTIGATION OF SUBSTRATE RECOGNITION OF SARS-COV-2 MAIN PROTEASE ENZYME BY MOLECULAR DYNAMICS AND FREE ENERGY METHODS

HİLAL SENA TAŞÇI

A THESIS SUBMITTED FOR THE DEGREE OF MASTER OF SCIENCE MOLECULAR BIOLOGY AND GENETICS DEPARTMENT

SUPERVISOR

ASSOC. PROF. DR. FATMA AYDINOĞLU

II. SUPERVISOR

ASSOC. PROF. DR. ABDULKADİR KOÇAK

GEBZE

2023

T.C.
GEBZE TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

SARS-COV-2 ANA PROTEAZ ENZİMİNİN SUBSTRAT TANILAMASININ MOLEKÜLER DİNAMİK VE SERBEST ENERJİ METOTLARLA İNCELENMESİ

HİLAL SENA TAŞÇI
YÜKSEK LİSANS TEZİ
MOLEKÜLER BİYOLOJİ VE GENETİK ANABİLİM DALI

DANIŞMANI
DOÇ. DR. FATMA AYDINOĞLU
II. DANIŞMANI
DOÇ. DR. ABDULKADİR KOÇAK

GEBZE
2023

GEBZE TEKNİK ÜNİVERSİTESİ	**YÜKSEK LİSANS JÜRİ ONAY FORMU**

GTÜ Fen Bilimleri Enstitüsü Yönetim Kurulu'nun 19/07/2023 tarih ve 2023/40 sayılı kararıyla oluşturulan jüri tarafından 03/08/2023 tarihinde tez savunma sınavı yapılan Hilal Sena TAŞÇI'nın tez çalışması Moleküler Biyoloji ve Genetik Anabilim Dalı'nda YÜKSEK LİSANS tezi olarak kabul edilmiştir.

JÜRİ

ÜYE

(TEZ DANIŞMANI) : DOÇ.DR. FATMA AYDINOĞLU

ÜYE : DR. ÖĞR.ÜYESİ ONUR SERÇİNOĞLU

ÜYE : DR. ÖĞR.ÜYESİ BERNA DOĞAN

ONAY

Gebze Teknik Üniversitesi Fen Bilimleri Enstitüsü Yönetim Kurulu'nun

……/……../…… tarih ve ……/…….. sayılı kararı.

İMZA/MÜHÜR

SUMMARY

Researchers have been diligently investigating antiviral treatments for SARS-CoV-2 due to its significant global impact, with over 380 million infections and 5.7 million deaths in the last three years. The emergence of the highly contagious omicron variant, capable of evading antibodies, underscores the urgency of developing effective therapies. A pivotal target is the SARS-CoV-2 main protease (M^{pro} or $3CL^{pro}$), responsible for cleaving viral polyproteins into essential structural and non-structural proteins during infection. Understanding the interaction between M^{pro} and its substrates is critical, especially considering potential therapeutic resistance driven by mutations. Molecular Dynamics simulations and free energy techniques, MMP(G)BSA, were employed to investigate the SARS-CoV-2 M^{pro}-substrate interaction. This study investigates the impact of residue mutations within the substrate recognition region on binding free energy. Key findings revealed the dominance of amino acids with small aliphatic side chains, such as alanine, valine, and glycine, in M^{pro}-peptide interactions. Notably, the glutamine residue at position P1 exhibited substantial changes in its interaction with M^{pro}, differing between substrates and products. Intriguingly, the study proposed that the arginine amino acid at positions P3-P5, along with P4', could enhance the interaction's strength. This research is the first to shed light on these critical aspects of M^{pro}-peptide recognition, offering valuable mechanistic insights. Targeting the identified regions and amino acids presents a promising avenue for disrupting M^{pro}'s function experimentally. In sum, this study contributes significantly to the ongoing quest for effective antiviral medications against SARS-CoV-2, crucial for managing the global health impact of the virus.

Keywords: Binding Free Energy, Molecular Dynamics, Protein-Protein Interaction, Substrate Recognition, SARS-Cov-2 M^{pro}.

ÖZET

Araştırmacılar, son üç yılda SARS-CoV-2'nin önemli küresel etkisi nedeniyle antiviral tedavilere özenle araştırma yapmaktadır; son üç yılda 380 milyondan fazla enfeksiyon ve 5.7 milyon ölüm yaşanmıştır. Yüksek derecede bulaşıcı olan omicron varyantının ortaya çıkması, antikorları atlayabilme yeteneğine sahip olması, etkili tedavilerin geliştirilmesinin aciliyetini vurgular niteliktedir. Ana hedef, SARS-CoV-2'nin ana proteazı (M^{pro} veya $3CL^{pro}$) olup, enfeksiyon sırasında viral poliproteinleri temel yapısal ve yapısal olmayan proteinlere ayıran sorumlu enzimdir. M^{pro} ile substratları arasındaki etkileşimin anlaşılması, özellikle mutasyonlar tarafından tetiklenebilecek potansiyel terapötik direncin göz önünde bulundurulması bakımından kritiktir. Moleküler Dinamik simulasyonları ve serbest enerji teknikleri olan MMP(G)BSA, SARS-CoV-2 M^{pro}-substrat etkileşimini araştırmak için kullanılmıştır. Çalışma, substrat tanıma bölgesindeki kalıntı mutasyonlarının bağlanma serbest enerjisi üzerindeki etkisine odaklanmıştır. Temel bulgular, M^{pro}-peptit etkileşimlerinde alanin, valin ve glicin gibi küçük alifatik yan zincirli amino asitlerin baskın olduğunu göstermiştir. Özellikle, P1 konumundaki glutamin kalıntısının, substratlar ve ürünler arasında farklılık gösteren büyük değişiklikler sergilediği belirlenmiştir. Çalışma, aynı zamanda P3-P5 konumlarındaki arginin amino asidinin ve P4'teki amino asidin etkileşimin gücünü artırabileceğini öne sürmüştür. Bu araştırma, M^{pro}-peptit tanımanın bu kritik yönlerini ilk kez aydınlatan ve değerli mekanistik anlayışlar sunan bir çalışmadır. Belirlenen bölgeleri ve amino asitleri deneysel olarak hedeflemek, M^{pro}'nun işlevini bozmak için umut veren bir yol sunmaktadır. Özetle, bu çalışma, SARS-CoV-2'ye karşı etkili antiviral ilaçların süregiden arayışına büyük katkı sağlamakta olup, virüsün küresel sağlık etkisini yönetmek açısından hayati önem taşımaktadır.

Anahtar Kelimeler: **Bağlanma Serbest Enerjisi, Moleküler Dinamikler, Protein-Protein Etkileşimi, Substrat Tanıma, SARS-Cov-2 M^{pro}.**

ACKNOWLEDGEMENTS

I would like to express my gratitude to many individuals who have provided their support in completing this thesis.

First and foremost, I would like to sincerely thank my advisors, Assoc. Prof. Fatma AYDINOĞLU and Assoc. Prof. Abdulkadir KOÇAK. Assoc. Prof. Fatma AYDINOĞLU has significantly contributed to the development of my work with her wisdom, guidance, and valuable ideas throughout this process. Assoc. Prof. Abdulkadir KOÇAK has been a companion with whom I have conducted joint research, exchanged ideas, and found solutions. I am grateful for their combined efforts and support.

Additionally, I would like to thank all my colleagues at Modellab who have assisted me with data analysis and techniques, providing their wisdom, guidance, and valuable ideas throughout the process. Their critiques and suggestions have helped me identify and improve the shortcomings in my thesis.

I would also like to extend my gratitude to my family, my fiancée and loved ones who have provided me with moral support throughout this thesis. It is their belief, love, and patience that have given me great strength during this process.

Without the help and support of each and every one of you, this study would not have been possible. I am grateful to all of you.

I also thank TUBITAK ULAKBIM because all calculations that are reported in this thesis were performed at TUBITAK ULAKBIM, High Performance and Grid Computing Center (TRUBA resources). This work was supported by Scientific and Technological Research Council of Turkey-TÜBİTAK (project number: 222Z163).

TABLE of CONTENTS

LIST of ABBREVIATIONS and ACRONYMS

Abbreviations And Acronyms		**Explanations**
Δ	:	Delta
ΔΔG	:	Relative Binding Free Energy
ΔG	:	Gibbs free energy change
λ	:	Number of state
BFE	:	Binding Free Energy
FE	:	Free Energy
L	:	Ligand
MC	:	Monte Carlo
M^{pro},$3Cl^{pro}$	:	SARS-CoV-2 main protease
MD	:	Molecular Dynamics
MM	:	Molecular Mechanics
MM-GBSA	:	Molecular Mechanics Generalized Born Surface Area
MM-PBSA	:	Molecular Mechanics Poisson Boltzmann Surface Area
NPT	:	Isobaric-isothermal ensemble
NVT	:	Canonical ensemble
P	:	Protein
PDB	:	Protein Data Bank
PL	:	Protein-ligand complex
SARS-CoV-2	:	Severe acute respiratory syndrome coronavirus
SASA	:	Solvent Accessible Surface Area
SAV	:	Solvent Accessible Volume
TMPRSS2	:	Cellular Transmembrane Serine Protease 2

LIST of FIGURES

LIST of TABLES

1. INTRODUCTION

As of February 2022, the 2019 Covid illness (Coronavirus) pandemic has infected 396 million people and caused the death of more than 5.7 million [1]. It has become perhaps of the most wrecking irresistible sickness in mankind's set of experiences. The responsible factor of Coronavirus (CoV) is a positive-sense, single-strand RNA (+ssRNA) infection having a place with the Betacoronavirus variety in the Coronaviridae group of the Orthocoronavirinae subfamily. Covids have a long history in the human population. HCoV-229E and HCoV-OC43 were first identified in 1966-1967 and shown to induce mild respiratory tract illnesses similar to the common cold virus. Starting around 2003, a few infections with serious pathologies, for example, extreme intense respiratory disorder Covid (SARS-CoV), Middle East respiratory condition Covid (MERS-CoV), and SARS-CoV-2, have arisen, stressing the need to comprehend and defeat these infections. CoV has an outstandingly enormous +ssRNA genome of roughly 30 kb, requiring remarkable systems to guarantee proficient combination of viral RNA. Upon entering the host cell, the sequential replication and transcription of CoV are associated with the generation of full-length complementary RNA (cRNA) or subgenomic RNA (sgRNA) templates and the copying of cRNA or sgRNA into positive-sense genomic RNA (gRNA) or subgenomic mRNA [2]. Both CoV gRNA and mRNA are covered at the 5' end with a cap structure ($^{7Me}GpppA_{2'OMe}$) and polyadenylated at the 3' end. Notwithstanding, cRNA has polyuridylic corrosive adornment at its 5' end rather than the cap structure. The 5' cap (1) structure on CoV gRNA/mRNA is a construction where the initial 5' nucleotide of the recently started pre-gRNA/mRNA is connected to the N7-methyl-guanine segment, and the ribose 2'- O position of the principal translated nucleotide is further methylated. This design is critical to advance interpretation commencement, safeguard mRNA, and help the infection in sidestepping host resistant acknowledgment. The concurrent covering of gRNA/mRNA is accomplished through four continuous activities [3]. First, a 5'-diphosphate end (ppA) is created when an RNA 5'-triphosphatase (RTPase) removes the -phosphate from the newly initiated pre-gRNA/mRNA's 5'-triphosphate end (pppA) [4], [5]. Second, the cap core (GpppA) is formed when a guanylyltransferase (GTase) moves

a guanosine monophosphate to the remaining 5'-diphosphate end [3]. Thirdly, the cap (0) structure (^{7Me}GpppA) is created when an N7-methyltransferase (N7-MTase) methylates the first guanine at the N7 position [6]. A 2'-O methyltransferase, 2'-O-MTase, methylates the ribose 2'-O position of the first transcribed nucleotide in the final stage to complete cap (0) and transform it into cap (1) (^{7Me}GpppA$_{2'\text{-}OMe}$) [7]–[9]. The genome of SARS-CoV-2 encodes 4 primary proteins, 16 non-underlying proteins (nsp1-nsp16), and a few extra proteins [10]. Within the host cell, non-structural proteins, also known as replicases, are crucial for a variety of aspects of the virus's life cycle. Non-structural proteins are translated into the replication-transcription complex (RTC) upon entry into host cells. This complex oversees several processes, including the transcription and replication of the viral genome, the molecular sieve-like formation of double-membrane vesicles (DMVs), and the control of protein translation on host ribosomes. The development of our knowledge of the structures and functions of coronavirus-encoded non-structural proteins has accelerated since the SARS outbreak. Single-particle cryo-electron microscopy (cryo-EM) technology has made it possible to observe the dynamics of CoV within host cells, which has led to significant advancements in the research on the assembly of CoV RTC.

1.1. Non-Structural Proteins

Figure 1.1 shows the non-structural proteins encoded by SARS-CoV-2. The functions of each nsp will be discussed briefly.

nsp1: In SARS-CoV-2 infection, the N-terminal cleavage product of the CoV genome's replicase polyprotein, protein nsp1, acts as a virulence factor. It has an N-terminal domain and a C-terminal domain that adopt a helix-turn-helix (HTH) structure when folded [11]–[16]. SARS-CoV-2 or SARS-CoV nsp1's C-terminal domain prevents mRNA from entering the host ribosome and diverting antiviral signaling pathways intended to suppress the host's innate immunity [14]–[18]. Additionally, the N-terminal domain binds to the viral genome's 5'-untranslated region (5'-UTR) with high affinity (0.31 M) [19]. In addition, nsp1 facilitates the accelerated cellular 5'-3' exoribonuclease 1-mediated decay of host mRNA by cleaving the 5'-UTRs of host mRNA [20],[21]. It is possible that nsp1's high-affinity binding to the 5'-UTR of the viral genome helps it avoid

mRNA decay [22]. SARS-CoV-2 nsp1's N-terminal domain has been shown to form a stable complex with the host's DNA polymerase, according to recent research [23]. Although the full-length nsp1 was utilized for the structural investigation in this study, only the N-terminal domain could be discerned in the experimental cryo-EM density. This observation suggests that the N- and C-terminal domains might concurrently perform distinct functions.

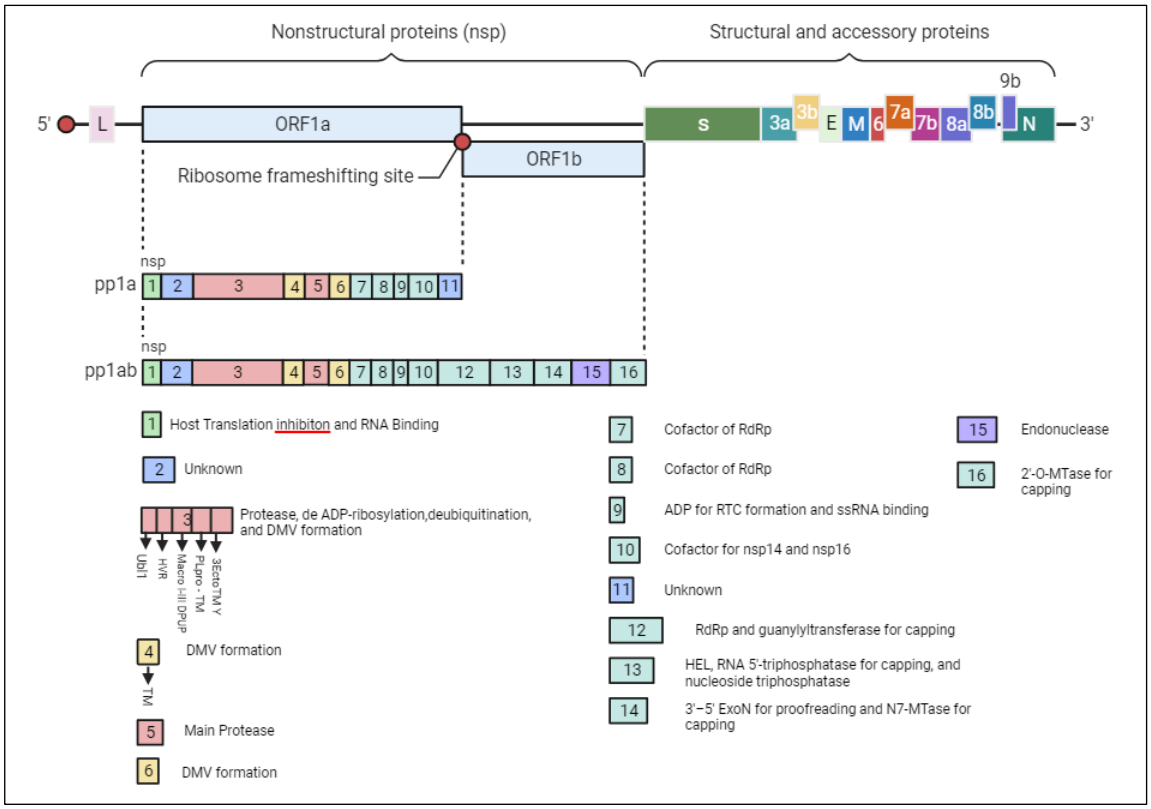

Figure 1.1: Nonstructural proteins encoded by SARS-CoV-2. Domains in nsp1–nsp16 with known functions are indicated. Abbreviations: 2' -O-MTase, 2'-O-methyltransferase; 3Ecto, ectodomain; Adp, adaptor; DMV, double-membrane vesicle; DPUP, domain preceding Ubl2 and PLpro; ExoN, exoribonuclease; HEL, helicase; HVR, hypervariable region; Macro, macrodomain; Mpro, main protease; N7-MTase, N7-methyltransferasensp, nonstructural protein; PLpro, papain-like protease domain; RdRp, RNA-dependent RNA polymerase; RTC, replication– transcription complex; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; ssRNA, single-stranded RNA; TM, transmembrane region; Ubl, ubiquitin-like domain; Y, Y1 and CoV-Y.

nsp2: nsp2 is one of the CoV non-structural proteins whose primary sequence and biological function are among the most diverse. It is connected to viral replication, protein

expression, localization, and functions that interfere with host cells' normal activities [24], [25]. However, the precise role that nsp2 plays throughout the CoV life cycle is still poorly understood. The multidomain structure of nsp2 includes a zinc finger-rich domain (ZRD) with three zinc fingers at the N-terminus, a middle region with a structure resembling an armadillo repeat, and a highly adaptable C-terminal domain with mostly loops and coils. The structure-function relationship of nsp2 requires further examination.

nsp3: The largest of the three membrane-associated proteins that the CoV genome encodes is nsp3. It typically embraces eight prominent domains: ubiquitin-like domain 1 (Ubl1), Glu-rich acidic region (also known as hyper-variable region or HVR), macrodomains I-III (Macro I-III), a region preceding the papain-like protease domain containing Ubl2 (DPUP), PLpro, an exodomain located between two transmembrane regions (TM1 and TM2) (also known as zinc-finger or ZF domain), and two Y domains (Y1 and CoV-Y) [26]–[31]. Although the individual structures of many CoV nsp3 fragments have been characterized, the structure of the full-length protein is still unknown. The papain-like protease (PL^{pro}) not only cleaves pp1a and pp1ab polyproteins to release the viral proteins nsp1, nsp2, and nsp3, but it also removes host ubiquitin and ubiquitin-like interferon-stimulated gene 15 (ISG15) from signaling proteins to suppress innate immune responses [32]. The classic catalytic triad (Cys111-His272-Asp286) in PL^{pro}'s catalytic region preferably cleaves the tetrapeptide motif (LXGG) in adjacent viral proteins (nsp1-nsp2) and the C-terminal tails of cellular ubiquitin and ISG15. The blockade loop 2 (BL2), a flexible -hairpin loop, regulates substrate access to the catalytic site, which is nearly 15 away from the catalytic region. There have been reports of over thirty potent PL^{pro} inhibitors. In light of their components of activity, anti PL^{pro} inhibitors can be classified into three groups: class (i) covalent inhibitors that bind to the catalytic cysteine via a C-S thioether bond; (ii) non-covalent inhibitors of class that prevent PL^{pro} substrates from entering the catalytic site; (iii) non-covalent inhibitors that bind to allosteric pockets. Covalent inhibitors like VIR250 and VIR251 are peptidomimetics in class (i) that inhibit PL^{pro} peptidase activity by mimicking the LXGG tetrapeptide motif [33]. As peptide substrates with only two glycine residues at the P1 and P2 positions can accommodate the substrate binding pockets, the "non-canonical" two glycine residues in the LXGG motif present a challenge for the development of potent peptidomimetic

inhibitors [34]. Non-covalent inhibitors in class (ii) block substrate access to the catalytic site by filling the BL2 pocket [34]. Due to sequence and structural differences in the BL2 pocket, the majority of these inhibitors typically only inhibit SARS-CoV and SARS-CoV-2, but not MERS-CoV. In class (iii), non-covalent inhibitors like HE9 target an allosteric pocket 30 Å away from the catalytic site to prevent ISG15 and ubiquitin from binding to PL^{pro}. There are also zinc-depleting drugs like disulfiram that block the allosteric Zn2+ binding pockets of PLpro. [35]. Human immune responses are altered, and viral protein maturation is prevented by PL^{pro} inhibition [36]. However, designing selective PLpro inhibitors can be challenging, as PLpro structurally resembles a large group of human deubiquitinase enzymes (DUBs) and DUB-like protease family that recognize ubiquitin or ubiquitin-like proteins [37]. Deubiquitinase enzymes and deubiquitinase-like proteases are being investigated as therapeutic targets for other human diseases, but inhibitors for these proteins have not been approved yet. [38].

nsp4 and nsp6: In addition to nsp3, CoV encoding two membrane-spanning proteins: nsp4, nsp6 [39], [40]. The combined efforts of the nsp3 and nsp5 proteases result in the release of the 56 kDa protein known as nsp4 from the polyprotein. nsp4 is anticipated to have four transmembrane spaces: an N-terminal transmembrane domain, a luminal domain, three additional transmembrane regions that are closely spaced apart, and a small C-terminal domain that faces the cytosol and is approximately one hundred amino acids long [39], [40]. Although the structure of nsp4's C-terminal domain has been figured out by crystallography [41], [42], little is known about the protein's overall structure. nsp6, another membrane-associated protein, is an autophagy-activating integral membrane protein [40], [43]. There are no structural details available for nsp6. nsp4 and nsp6 can induce double-membrane vesicles (DMVs) as replication organelles along with nsp3 [30] and form a hexagonal crown-shaped porous structure that covers the DMVs as an opening facing the cytosol [44]. It has been hypothesized that this porous structure eases the transfer of RNA to the cytosol by interacting with RTCs within DMVs.

nsp5: Due to its characteristic chymotrypsin-like structure, nsp5, also known as the 3C-like protease (3CLpro), plays a pivotal role. [45]–[49]. Nsp5 acts as the main protease (Mpro) by cleaving pp1a and pp1ab to form individual proteins, specifically nsp4 to nsp16 [26], [50]–[52]. A characteristic cleavage motif P4-P2' (small)-X-(L/I/V/F/M)-Q(S/A/G),

where X denotes any amino acid and indicates the cleavage site, summarizes the catalytic region of nsp5 [52]. Its essential function and absence of a homolog in humans make nsp5 a promising target for antiviral development against SARS-CoV-2 and other CoVs [45], [53]–[61]. Additionally, nsp5 forms a homodimer where one protomer's N-terminal domain (N-finger) shapes the S1 pocket and oxyanion hole of the other protomer, necessitating dimerization for catalytic activity [45], [46]. The SARS-CoV-2 main protease (Mpro), also referred to as the 3C-like protease, is a crucial cysteine protease responsible for releasing viral proteins from the pp1a and pp1ab polyproteins, specifically nsp4 to nsp16. Inhibiting Mpro-mediated proteolytic cleavages disrupts the maturation of important viral enzymes like nsp12 and nsp13, thus halting viral replication. Mpro's catalytic dyad, consisting of a nucleophilic cysteine (Cys145) at position 145 and a nearby histidine residue (His41) at position 41, exists in each subunit of the homodimer and catalyzes the covalent carbon-sulfur bond formation between the P1 glutamine and substrate's main chain carbonyl group [62], [63].

Similar to the rational design of inhibitors for other proteases, Mpro's catalytic dyad interacts comprehensively with atoms in the S1', S1, S2, S3, and S4 subregions of Mpro [64], [65]. Mechanistically, Mpro inhibitors are categorized into four groups: Class (i) comprises peptidomimetic inhibitors covalently binding to the catalytic pocket via an electrophilic warhead such as nitrile, ketone, α-ketoamide, or aldehyde acceptor. Class (ii) involves non-peptidomimetic inhibitors that block the catalytic pocket via covalent interactions. Class (iii) includes non-covalent inhibitors that occupy Mpro's substrate binding pocket without forming covalent bonds. Class (iv) centers on non-covalent inhibitors preventing Mpro dimer formation at allosteric sites.

In class (I), nirmatrelvir, a peptidomimetic with a nitrile warhead, reversibly inhibits Cys145 by forming a covalent bond. Nirmatrelvir effectively reduces viral loads in mice and hamsters [66] by inhibiting seven human coronaviruses (HCoVs) in cell culture [61] and various SARS-CoV-2 variants [67].

Non-covalent inhibitors synthesized in Class (ii) disrupt the catalytic pocket's accessibility through non-covalent interactions, such as penicillin derivatives [68], ebselen derivatives [69], ester derivatives [70], spiro-cyclic derivatives [71], and myricetin derivatives [72].

Class (iii) features ensitrelvir and other non-covalent inhibitors occupying Mpro's substrate binding pocket through hydrogen bonds, hydrophobic contacts, and stacking interactions [73].

Allosteric inhibitors in Class (iv), including NB1A2, colloidal bismuth subcitrate, and x1187, target Mpro's dimer interface to prevent dimer formation.

nsp7 and nsp8: nsp7 and nsp8 structure are steadily complicated and they act as cofactors for nsp12 polymerase, together framing the replication-transcription complex (RTCs). The nsp7/nsp8 complex exhibits low fidelity de novo RNA synthesis and extension abilities on ssRNA templates, suggesting that nsp8 serves as an RNA primer for the synthesis of short oligonucleotide primers for subsequent extension by the nsp12 polymerase [74], [75]. Biochemical studies have shown that nsp8 functions as a second polymerase in CoVs. The structure of the nsp8 protein consists of two parts, whereas the structure of the nsp7 protein folds into a tight helical bundle. In the C-terminal region, a central -sheet surrounded by helices, and in the N-terminal region, a long helical structure. The postulated double-stranded RNA (dsRNA) binding channel and the SARS-CoV nsp7/nsp8 complex's structure were first identified as a hexadecamer ring structure [76]. In this hexadecamer complex, the N-terminal helix of nsp8 takes on two distinct configurations [76]. In a similar vein, it has been demonstrated that the N-terminal helix of nsp8 stabilizes the template-product duplex that is produced by the nsp12 polymerase, either as a component of the nsp7 complex or in RTCs [77]–[80].

nsp9: The protein nsp9 is involved in CoV replication and is a single-stranded nucleic acid binding protein [81], [82]. It binds long oligonucleotides with moderate binding strength and shares topological characteristics with oligonucleotide and oligosaccharide binding (OB-folding) proteins [81], [82]. Crystallography studies indicate that the interaction between the C-terminal -helices and the GXXXG motif will result in nsp9's transformation into a biological dimer [81]–[86]. The spread of the virus is significantly slowed down by mutations at the nsp9 dimerization interface. In late examinations, monomeric nsp9 was found to tie firmly with nsp12 nidovirus RNA-subordinate RNA polymerase (RdRp)- related nucleotidyltransferase (NiRAN), which constricted NiRAN GTase action, trailed by Cap (0). It has been seen that nsp9 is successfully drawn in by nsp12 to produce RTC [87]. Intriguingly, the fact that the

interfaces for nsp9 dimerization and interaction with other Cap (0)-RTC components largely overlap raises the possibility that the homodimeric form of nsp9 may switch constitutively prior to its assembly into RTCs.

nsp10: The ZF protein nsp10 has a molecular weight of approximately 15 kDa. Two ZFs with the sequence motifs C-(X)2-C-(X)5-H-(X)6-C and C-(X)2-C-(X)7-C-(X)-C are present [88]. Initially reported to possess a global dodecameric architecture in a crystallographic study, it has later been demonstrated that nsp14 and nsp16 function as monomers, acting as stimulants for their enzymatic activities [89], [90].

nsp12: nsp12 is the most important nonstructural protein for virus replication and transcription. It is an RNA-subordinate RNA polymerase (RdRp) that uses nsp7 and nsp8 as cofactors to catalyze the synthesis of viral RNA [91], [92]. There are three regions in SARS-CoV-2 nsp12: the N-terminal nidovirus RdRp-associated nucleotidyl transferase domain (NiRAN) district, a connection point locale in the center area, and the C-terminal right-hand polymerase (Pol) area. This Pol district embraces the preserved design of the viral polymerase family [93] and comprises of the fingers, palm, and thumb subdomains. In the focal channel of nsp12 Pol, the A-G themes structure the synergist community for RNA union. It was first discovered that the N-terminal NiRAN region of nidovirus RdRp possesses G/U-specific nucleotidyltransferase activity [94]. SARS-CoV-2 nsp12 NiRAN has as of late been displayed to work as a GTase that catalyzes the subsequent coat activity to shape the cap center design GpppA from the ppA forerunner [95]. It is likewise guessed that posttranslational change of nsp9 by nsp12 NiRAN assumes a part in RTC development. nsp12 RdRp is a profoundly moderated holoenzyme encoded by nsp12 that is engaged with viral RNA replication and record. In the viral replication-transcriptional complex, mature nsp12 collaborates with other nonstructural proteins (nsp7 to nsp10, nsp13 to nsp16) following M^{pro}-mediated proteolytic processing. This complex catalyzes format opening, RNA combination, RNA correction, and RNA covering. A range of antiviral compounds can be developed to inhibit various stages of RNA synthesis by focusing on the main components of the replication-transcription complex. These stages include RNA synthesis inhibition (nsp12 RdRp inhibitors), template unwinding (nsp13 helicase inhibitors), RNA proofreading (nsp14 exonuclease inhibitors), uridine excision (nsp15 endoribonuclease inhibitors), and RNA capping inhibition (nsp9 inhibitors, nsp12

NiRAN inhibitors, nsp13 ATPase inhibitors, nsp14 guanine-N7-methyltransferase inhibitors, and nsp16 2'-O-methyltransferase inhibitors), all aiming to prevent RNA synthesis. For viral RNA capping and other functions, NiRAN and C-terminal RdRp domain of nsp12 are present. The NiRAN domain is a guanyltransferase that adds a guanosine 5'-triphosphate to the 5'-end of viral RNA during the capping process. The NiRAN dynamic site can be obstructed by nucleoside analogs, for example, remdesivir triphosphate and bemnifosbuvir triphosphate. The NiRAN domain shares significant structural similarities with numerous human kinases, including insulin receptor kinase, spleen tyrosine kinase, and O-mannose kinase, making it difficult to develop NiRAN inhibitors with high specificity and low toxicity. A significant drug target is the RdRp active site of SARS-CoV-2. As nucleoside and non-nucleoside analogues, RdRp inhibitors, like HIV reverse transcriptase inhibitors, have distinct mechanisms of action.

nsp13: The protein nsp13 is a superfamily 1 (SF1) helicase with various enzymatic capabilities, including opening double-strand RNA, nucleoside triphosphatase, and RTPase movement [95], [96]. Nsp13 preprocesses high-order conformation in the ssRNA template and unfolds template–product pairs in order to direct RNA synthesis [97, 98]. Dissimilar to most SF1 helicases and +ssRNA infection SF2 helicases, recomposite apo CoV helicases unpair in the 5'- 3' bearing [97]–[100]. In addition, CoV nsp13 uses RTPase activity to facilitate initial coated action [3]–[9], [99]. SARS-CoV-2 nsp13 has a commonplace collapsing of the SF1 helicase and contains a 1B locale connected to ZBD by means of the N-terminal zinc-restricting district (ZBD), two head Rec-like districts (1A and 2A), and a stem district [97], [98]. Districts 1A and 2A handle the reactant site for ATP hydrolysis, while locales 1A, 2A and 1B structure an RNA restricting section to embrace ssRNA. Recent research has revealed that the 1B region detaches from its initial location, undergoes remarkable conformational change, and may serve as a driving force for a potential back-slip. In contrast to the 5'-3' feature discovered in biochemical experiments [97]–[100], ssRNA binds in the 3'-5' direction in the RNA binding groove.

nsp14: nsp14 is a bifunctional enzyme with a C-terminal N7-MTase domain that catalyzes the third coated action to produce cap (0) and a proposed N-terminal exonuclease (ExoN) domain to maintain replication accuracy. The C-terminal N7-MTase has an abnormal MTase collapsing containing the S-adenosylmethionine (SAM)

restricting theme [90], [101], [102] and embeds SAM in the space adjoining the GpppA restricting site. According to the proposed 3'–5' ExoN activity, the N-terminal ExoN can remove misincorporated nucleotides from the 3' end of the nascent product RNA to correct replication and transcription errors [101]–[104] and perform a trans-transitive retrospective. It has a mechanism for validation [87]. In order for nsp14 to perform its full function, it must be linked to nsp10 [90] . However, nsp10 only binds to nsp14 ExoN and boosts the activity of that protein [7], [105], but it has no effect on nsp14 N7-MTase activity [90].

nsp15: Among coronaviruses, Coronavirus nsp15 is a conserved uridine-specific endoribonuclease (endoU). There is a C-terminal endoU domain as well as an N-terminal domain and a middle domain whose function is unclear in the nsp15 folding [106]–[108]. Albeit the job of its oligomerization isn't completely perceived, nsp15 is dynamic just in a hexametric structure. This protein is primarily uridine-specific in its cleavage of the 3' end of RNA substrates [109], [110]. nsp15 is a particular uridine endoribonuclease that cleaves 5'-polyuridine sequences in negative-sense viral RNA to prevent host immune responses from being activated. A potential drug binding site exists in the Mn2+-dependent endoribonuclease active site of nsp15, which cleaves viral RNA substrates and produces 2′,3′-cyclic phosphodiester and 5′-hydroxyl termini. Due to the absence of close human homologues, nsp15 endoribonuclease might be an appealing antiviral target in theory.

nsp16: The final nonstructural protein in the CoV genome, nsp16 catalyzes the final coat action as a 2'-O-MTase [111], [112]. Cap (0) pockets are formed by loop regions that originate from the central -sheet and house SAM as a substrate and methyl donor. SARS-CoV-2 nsp16, on the other hand, has a distinct adenosine binding pocket about 25 Å from the catalytic pocket, suggesting that it may be involved in the coating's pre-mRNA binding [7], [8], [89], [111]. CoV nsp16's 2'-O-MTase activity can be enhanced when it forms complexes with nsp10. However, it has been observed that nsp10 does not induce any significant conformational changes in either nsp10 or nsp16.

The complexes of these nsps with M^{pro} is the focus of the current thesis. Several C- and N- terminal fragments (peptides) of these nsps have been co-crystalized with M^{pro}.

The sequential difference of these peptides provide insight into understanding the M^{pro}'s substrate recognition.

1.2. Replication-Transcription Complexes

For sequential replication and transcription, a collection of protein machinery known as RTCs (Replication-Transcription Complex) is formed when individual nonstructural proteins organize themselves inside host cells. The primary endeavor to comprehend the profoundly requested association of CoV nonstructural proteins came from the crystallography investigation of the SARS-CoV nsp7/nsp8 complex [76]. Due to the complex's large molecular size and the dynamic process of RTC formation, the investigation of RTC assembly and function came to a halt until the structure of the SARS-CoV nsp12/nsp7/nsp8 complex was determined using single-particle cryo-EM technology in 2019 [91]. With the strong support provided by single-particle cryo-EM, our understanding of CoV RTC assembly and function has significantly advanced, particularly based on studies of SARS-CoV-2 RTCs. Six major stages make up the CoV replication and transcription cycle in host cells, according to our current understanding.

Prior to RNA synthesis, nsp12 interacts with its coenzymes nsp7 and nsp8 to form the central RTC (C-RTC). The C-RTC serves as a central component that encompasses all states of the RTC. Both SARS-CoV and SARS-CoV-2 C-RTCs consist of one nsp12, one nsp7, and two nsp8 molecules each [78], [91]. An individual nsp8 (nsp8-1) collaborates with the nsp12 connection point and finger subdomains, while another nsp8 (nsp8-2) and nsp7 cooperate with the nsp12 finger and head subdomains. The N-terminal helices of both nsp8 are not visible in the apo form of C-RTC. However, when template-product duplex is present, the nsp8 N-terminal helices manifest as long helices and function as a crane, separating the pair from the template that extends from the nsp12 Pol catalytic center [79], [80].

Due to the presence of numerous higher-order structures in the RNA genome, it's unlikely that C-RTC initiates RNA synthesis on its own. For this purpose, the involvement of nsp13 in the C-RTC is necessary. Thus, a template is created to guide C-RTC by processing these higher-order structures in advance. The complex formed by nsp13 and

C-RTC is referred to as the elongation RTC (E-RTC) [98]. Due to nsp13's weak binding affinity with C-RTC, a template-product duplex resembling a single-stranded template extending from the 5' end of the template to the nsp12 Pol catalytic center of nsp13 has been designed. This assists in capturing the structure of E-RTC [97], [98]. An unexpected structural feature is the binding of two nsp13 molecules to C-RTC. One of them binds to the 5' end of the template, while the other does not bind to RNA. The helicase activity of E-RTC is enhanced by interactions between the two nsp13 molecules.

It is likely that positively labeled CoV gRNA and mRNA are capped at the 5' end, whereas negatively labeled cRNA or sgRNA are not [113], [114]. Since the 5' end of the product cRNA or sgRNA dissolved by the helicase does not require further modification, it is likely that it is directly released into the medium to synthesize cRNA or sgRNA under the direction of gRNA. However, a four-step cap (1) process at the 5' end is required to synthesize gRNA or mRNA using cRNA or sgRNA as a template. The nsp13 RTPase activity and the nsp12 NiRAN GTase activity facilitate the first two capping processes [3]–[9], [99]. At the 5' end of the product RNA, the RTCs that make up the ppA tip and the GpppA cap core are referred to as Cap (-2)-RTC and Cap (-1)-RTC, respectively. While we don't utilize the terms Cap (- 2) and Cap (- 1) as formal logical articulations to signify ppA and GpppA develops, we might involve these terms here for simplicity of portrayal. There are currently no constructs for Cap (-2)-RTC and Cap (-1)-RTC. However, due to the extremely large distance (more than 120 in linear distance) between the 5' end of the product RNA in the E-RTC and the catalytic center of the nsp12 NiRAN, either (a) another component or components by relocating to the RTC, a pathway is formed to cis-cap the 5' end of the product to the nsp12 NiRAN, or (b) a specific oligomerization state of the RTC, positioning, which reveals an in-transit mechanism for the initial two capping actions.

An important intermediate state of the RTC, known as the Cap (-1)-RTC, occurs with the provision of nsp9 to E-RTC before the third and fourth capping processes performed by nsp14 and nsp16 [3]. While emerging as a dimer in crystallographic structures, nsp9 is accommodated as a single monomer in Cap (-1)-RTC and serves as an adaptor for further recruitment of nsp10/nsp14 or nsp10/nsp16, thus providing a platform for the formation of the co-transcriptional capping complex (CCC) for capping [81], [82], [84]–[86]. Furthermore, there are two proposed mechanisms by which nsp9 binding may

halt RNA elongation and capping. To begin, the nsp9 N-terminus is deeply embedded in the nsp12 NiRAN catalytic center, where it forms stable interactions with a cation and a GDP [3]. GpppA formation and nsp12 NiRAN GTase activity are clearly diminished when Nsp9 is present [3]. Second, the orientation of nsp13-2 binding to the template in E-RTC is significantly altered when nsp9 is bound [3]. Nsp13-2 blocks upstream duplex movement by rigidly sliding its ZBD into the second minor groove of the template-product duplex when bound to Nsp9. In this specific circumstance, nsp9 not just gives a stage to the development of CCC, however is probably going to practically suspend replication and record after the second covering process. As examined in the nsp9 segment, this protein is a solitary abandoned nucleic corrosive restricting protein and offers topological likeness with OB-collapsing proteins to tie long oligonucleotides [81], [82].

Cap (0)-RTC is created when nsp9 recruits the nsp10/nsp14 complex to Cap (-1)-RTC. Cap (0)- RTC exists in both arrangement and cryo-EM caught state as monomeric and dimeric structures [87]. N7-CCC is formed when nsp9 and nsp12 NiRAN interact with nsp10 but only with nsp14 in monomeric Cap (0)-RTC [mCap(0)-RTC]. In the N7-CCC, several highly positively charged regions have been identified; these are likely to serve as the path through which precursor gRNA and mRNA are transferred from the nsp12 NiRAN catalytic center to nsp14 N7-MTaz. Distinct parts in mCap (0)- RTC (i.e., nsp12 Pol and nsp13) are called EC (stretching complex) and EC and N7-CCC structure a hand weight molded design of mCap (0)- RTC. Most importantly, the presence of the back-correction mechanism in trans is confirmed by the dimeric Cap (0)-RTC [dCap(0)-RTC] [87]. The catalytic center of a Cap (0)-RTC protomer's nsp14 ExoN is located approximately 35 Å from the 3' end of the product RNA, in opposition to the catalytic center of a nsp12 Pol protomer. When RTC finds that a nucleotide is not properly incorporated into product RNA, nsp13-2 1B undergoes significant conformational changes to drag template RNA in the opposite direction of RNA elongation [87]. This allows the 3' end of the wrongly incorporated nucleotide to be reinserted into the nsp14 ExoN to correct the error [87]. After the correction of a mismatched nucleotide, nsp13-2 returns to its regular position for helicase function. This significant conformational change in nsp13-2 1B is not observed in any other SF1 or SF2 helicase.

The nsp10/nsp16 complex with 2'-O-MTase activity is recruited during the decisive step of cap synesthesia to complete the capping process. The name of this complex is Cap (1)-RTC. nsp14 and nsp16 are less likely to be co-recruited into RTCs by nsp10 because their footprints on nsp10 overlap [105]. In addition, if we superpose Cap (0)-RTC structures [87] with nsp10/nsp16 structures [111], [112] the position of nsp16 significantly conflicts with the interprotomer interface in dCap (0)-RTC, rendering dCap (0)-It may adversely affect the correction in RTC.

1.3. Other Non-Structural Proteins with A Related RTC

CoVs typically replicate and transcribe their genomes at a location associated with modified host membranes, which are referred to as DMVs (double-membrane vesicles). These membranes are the predominant component of the transformed endoplasmic reticulum (ER) membranes of the viral replication organelle [115]–[117]. DMVs may provide a microenvironment to shield viral RNA from host cell innate immunity by packaging the dsRNA replication intermediate [115]–[117]. RTC is associated with the luminal side of the nsp3/nsp4/nsp6 pores in subtomogram analysis of DMVs induced by CoVs [44], suggesting that mRNA can be exported to the cytosol immediately following expression in the RTC. In addition, the N-terminus domain of nsp1 has a high affinity for binding to the 5'-UTR [19].

1.4. Structural Proteins

SARS-CoV-2 is a positive single-abandoned RNA-containing envelope infection going from 80 to 220 nm in measurement, having a place with the β-CoV class of human Covids [118]. A fragile lipid envelope, four structural proteins, and genomic RNA make up the entire SARS-CoV-2 fragment. The membrane (M), nucleocapsid (N), envelope (E), and spike (S) proteins are the four structural proteins. In virus assembly, the M protein is crucial. Viruses and host factors can assemble at the cell membrane to form daughter virus particles thanks to their presence [119]. In viral transcription and splicing, the N protein and genomic RNA form an important complex. There are three parts to the N protein: N-

terminal, C-terminal, and disordered central domains (RNA binding domain, CTD, and NTD) [120]. The E protein is a small, complete membrane protein that plays a role in the assembly and pathogenicity of SARS-CoV-2 throughout its life cycle [121]. The S protein assumes a critical part in the attack of SARS-CoV-2 into cells and is available as trimers on the outer layer of the infection film [122]. It comprises of S1 and S2 subunits, and the last option is the most moderated district of the spike protein [123]. The N protein interacts with the viral RNA to form the nucleus of the virus particle, while the E protein and M protein are located sequentially on the surface of the virus membrane and form the virus envelope with the S protein [118]. During viral replication, a polypeptide 1ab is translated from the ORF1ab in the viral genome and subsequently cleaved into 16 non-structural proteins by protease. The proteins of SARS-CoV-2 are highly glycosylated 8uy g[124]. There are 22 N-glycosylation sites per protomer in the SARS-CoV-2 S protein [125]. When developing vaccines and antibodies, glycosylation of the S protein must be considered because glycosylation is necessary for virus invasion [124].

1.5. The SARS-CoV-2 Life Cycle

SARS-CoV-2 binds to the cell surface via the S protein to Angiotensin-converting enzyme 2 (ACE2), a functional SARS-CoV receptor, allowing it to enter the cell via membraneization and endocytosis [123]. SARS-CoV-2's invasion processes are aided by proteins from host cells, such as the serine protease TMPRSS2 [126] and the HDL scavenger receptor type B (SR-B1) [127]. Transmembrane serine protease 4 (TMPRSS4) is a protein generally fundamentally connected with ACE2 [128]. SARS-CoV-2 has an exceptional Furin cleavage site not recently found in other Covids, and this is expected for infection section into cells lacking cathepsin protease [129]. Furin can activate the S1 and S2 subunits of the SARS-CoV-2 spike protein by cleaving the S1/S2 region of the protein [118]. During the process of packaging the virus, the truncation occurs [130]. The virus particle uses the host ribosome to translate the viral polyproteins pp1a and pp1ab after entering the host cell [131]. Viral proteases 3CLpro [132] and PL^{pro} [133] then cut the polyproteins into different dynamic proteins. A nonstructural protein-based replication/transcriptional complex dominates viral replication [134]. The virus-

independent RNA polymerase construct (RdRp), for instance, is made up of nsp7, nsp8, and nsp12 [79]. Interpretation of nsp14 can capture protein union and restrain the natural resistant reaction, while its relationship with nsp10 improves this impact [135]. The N protein binds to the genome after transcription produces four structural proteins. The remaining three proteins S, E, and M integrate into the endoplasm before being transported to the endoplasmic reticulum-Golgi intermediate region (ERGIC) for further processing, such as Furin-mediated cleavage. Consequently, the daughter virus that is produced is used for the subsequent invasion of the host cell and is released through excitolysis or budding. Protein E can cause pro-inflammatory effects in ERGIC by forming pores that release calcium and activate the NLRP3 inflammasome [136].

1.6. Drugs That Target The SARS-CoV-2 Life Cycle

There are multiple stages in which SARS-CoV-2 enters the human body and self-replicates to release the progeny virus. Drug targets for these specialized processes have been developed to inhibit the virus's viability based on the characteristics of each dynamic stage.

1.6.1. Competitive Binding to ACE2 and the SARS-CoV-2 Spike Protein

The primary entry receptor for some human coronaviruses, including SARS-CoV, SARS-CoV, and HCoV-NL63 [137], is ACE2. Spike RBD recognizes the N-terminal domain of the ACE2 dimer on the cell surface during the first step of virus entry into host cells. After SARS-CoV-2 disease, ACE2 articulation is decreased and guideline of the human renin-angiotensin framework by means of ACE2 is lessened, bringing about aspiratory hypertension, aggravation, and cardiovascular intricacies [138]. By imitating ACE2 with peptide fragments or mini-proteins like APN01, a potential treatment strategy aims to prevent SARS-CoV-2 from entering host cells. The soluble extracellular fragment of wild-type human ACE2 known as APN01 is a miniature protein [139]. APN01, on the other hand, did not significantly outperform the placebo control in a phase II study involving COVID-19 patients in terms of 28-day all-cause mortality or the use of invasive

mechanical ventilation. In preliminary clinical studies, small molecules like SB27041 that target ACE2 have been investigated as potential anti-SARS-CoV-2 agents. SB27041 inhibits ACE2 interactions with the SARS-CoV-2 spike without affecting ACE2 enzymatic functions by targeting the allosteric binding site of ACE2 and causing conformational changes [140].

The SARS-CoV-2 spike is an attractive antiviral target because it is a homotrimeric class I fusion glycoprotein on the virus surface that is necessary for viral entry. N- and O-lycosy bonds are used by host proteases to cleave the spike protein to the receptor-binding subunit S1 and envelope-fusion subunit S2 in most cases. After critical conformational adjustments, the receptor restricting site (RBD) of the SARS-CoV-2 spike ties with ACE2 on the cell surface with high affinity, just multiple times higher than the SARS-CoV spike's ACE2 [141]. The postfusion conformation of the three helix bundles that connect the viral membrane to the plasma membrane of the host is driven by subsequent structural transitions and proteolytic interruptions [142]. Spike-ACE2 interaction or membrane splicing can be inhibited by various agents, including neutralizing antibodies, small molecule inhibitors, and peptide inhibitors.

1.6.2. Cell Proteases

Cellular proteases such as Furin, Transmembrane Protease Serine 2 (TMPRSS2), cathepsin proteases, ADAM10, and ADAM17 cleave and prime the SARS-CoV-2 spike protein for viral entry [143]–[146]. The envelope fusion of SARS-CoV-2 virions with host cells is linked to the proteolytic cleavage of the viral spike and relies on a two-step process: Firstly, cleavage at the S1/S2 junction (684AR↓SV687) by furin releases the receptor-binding subunit S1 and the envelope fusion subunit S2 [146]. Secondly, a proteolytic cleavage is performed at the S2′ site (814KR↓SF817) within the envelope fusion subunit S2 to release the fusion peptide of the spike that holds the cell membrane [147]. During rapid cell-envelope fusion, TMPRSS2 or TMPRSS13 are responsible for cleavage in the S2′ region of SARS-CoV-2 S protein, while endosomal cathepsin proteases (primarily cathepsin B and L) are responsible for slow endosomal internalization [148].

These structural proteins are beyond the scope of the current thesis.

1.7. Rising SARS-CoV-2 Variants

SARS-CoV-2 is constantly evolving, resulting in a greater risk to global public health and a faster spread and higher infectious efficacy. WHO classifies variants as variants of interest (VOI) and variants of concern (VOC) to facilitate prevention or medical countermeasures and better assess the consequences of various variants [138]. There are at present two VOIs (Kappa and Mu) and five VOCs (Alpha, Beta, Gamma, Delta, and Omicron). On November 24, 2021, another SARS-CoV-2 variation B.1.1.529, named Omicron, was found in South Africa. In the past, the SARS-CoV-2 Delta variant had emerged as the predominant epidemic strain in numerous nations [149]. A new level of focus and vigilance has emerged because of the omicron's emergence. Mutations in Omicron are concentrated in the S protein, and there is a tendency to accumulate mutations of the virus that are beneficial for immune evasion. [139], [140]. Omicron's infectivity, according to one model, is approximately ten times that of the original virus or twice that of the Delta variant. Omicron can significantly reduce the effectiveness of the FDA-approved monoclonal antibody (mAb) from Eli Lilly [143]. New concerns regarding antiviral medications and vaccines for the COVID-19 pandemic are raised by major variants of SARS-CoV-2 that have been reported in a variety of locations across the globe. The virulence and spread of SARS-CoV-2 are determined by critical mutation sites in the genome, which offer novel drug design concepts for major emerging variants.

The D614G mutation in the S protein will not significantly affect the neutralizing properties of antibodies against SARS-CoV-2, and the vaccines that are currently in development are still effective against strain D614G [144]. The neutralizing power of various potent mAbs that alter the receptor binding motif on the RBD will be diminished if E484K, the S protein's receptor binding site (RBD), is mutated [145]. Addition of the E484K change into other variation strains will likewise diminish balance of the immunization and antibodies [146]. The neutralizing effect of neutralizing nanospheres (Nbs) is little affected by the N501Y mutation in British variant strain B.1.1.7 [147]. However, the mutant strain frequently contains additional significant amino acid mutations that affect the S protein's ability to bind to ACE2. The L18F mutation in the NTD region may prevent the S2L28 monoclonal antibody from binding to NTD [150] and

may have a positive correlation with mortality [148]. The K417N change near the ACE2 restricting site marginally hinders interactions with ACE2 [146] yet can advance effective avoidance of variations from antibodies by disposing of the inserted interface salt extension among RBD and the killing counter acting agent CB6 [151]. The L452R mutation found in the hydrophobic plates of the RBD [152] builds the spike protein's limiting affinity for ACE2 and disables Capture restricting [153] and sidesteps the infection from the monoclonal immunizer LY-CoV555 [154]. P681R is a mutation that occurs close to the Furin cleavage site [155]. It increases the full-length spike's cleavage of S1 and S2 to facilitate infection from the cell surface [156].

1.7.1. Non-Structural Protein Mutations

Among the nsp of SARS-CoV-2, important drug targets include virus-associated enzymes like RdRp and 3C [157]. These mutations may make the virus more resistant to drugs that are related [158], [159]. Studies have shown that the mutation C14408T increases the viral mutation rate, while the mutation C15324T reduces its impact [160]. There may be a correlation between the severity of SARS-CoV-2 symptoms and mutations in nsp. One study found a significant correlation between asymptomatic COVID-19 and mutations in the nsp6 coding region [161], which may make it challenging to begin screening COVID-19 patients. Mutations can alter the structures and functions of various nsps [162]. The G1691C mutation in Nsp3 decreases the protein's flexibility, while the V121D mutation in nsp1 may have a disruptive effect [54]. When developing COVID-19 medications, vaccines, or treatments, mutations in these nsps should be considered.

1.7.2. Viewpoints In Drug Plan For SARS-CoV-2 Variations

There are more than 100 monoclonal antibodies (mAbs) utilized in the treatment of human illnesses, a few of which are utilized for viral diseases, for example, palivizumab for respiratory syncytial infection and ansuvimab for Ebola infection [163], [164]. Several mAb and mAb cocktails against SARS-CoV-2 have been approved for emergency use or emergency use authorization (EUA) so far. In outpatients with mild to moderate COVID-

19, the majority of mAbs and their cocktails are approved as prompt treatment options [165]. The majority of mAbs are no longer recommended due to the continued emergence of SARS-CoV-2 variants, and their efficacy must be continuously evaluated [166]. With amino acid modifications in the fragment crystallized (Fc) region, potent wild type mAbs can then be engineered to produce mAbs with longer half-lives and enhanced effector functions [167]. IgG1 mAbs make up all commercially available anti-SARS-CoV-2 antibodies. Finally, as a natural source of polyclonal antibodies against SARS-CoV-2, high-titer convalescent plasma derived from COVID-19 survivors can be used for passive immunotherapy [168], [169].

There are two primary ways that anti-spike antibodies reduce SARS-CoV-2 infection. The first is to eliminate the pathogen by blocking viral entry with antibodies that bind specifically to spike RBD [170]. In COVID-19 recovery plasma, approximately 90% of neutralizing antibody titers are made up of antibodies that bind to RBD [171]. Antibodies use the opsonization pathway, complement-dependent cytotoxicity, and/or antibody-dependent cellular phagocytosis and cytotoxicity to eliminate SARS-CoV-2 virions or infected cells as the second mechanism [172]. Currently available antibodies typically inhibit the spike–ACE2 interaction by targeting the RBD. Antibodies binding to the RBD can be divided into two classes based on their target epitopes. Firstly, the highly variable nature of the SARS-CoV-2 spike protein allows for the emergence of drug-resistant mutations [166]. Many existing anti-spike antibodies are weakly effective or ineffective against concerning Omicron variants like BA.1, BA.1.1, BA.2, BA.4, and BA.5, as these variants exhibit over 30 amino acid changes, including at least 15 within the RBD of the spike protein. To prevent neutralization escape, it's crucial to develop effective antibodies and combinations targeting highly conserved, non-overlapping epitopes within or outside the spike RBD. Secondly, antibody-based therapies may have clinical benefits for specific patients in the early stages of the disease or those with undetectable anti-SARS-CoV-2 antibodies [173].

In this way, the revelation of a wide range antiviral medications ought to be engaged and created. Most of the mutation sites that are being investigated at the moment are found in the S protein, which is a crucial step in the virus's infiltration of the human body [174]. However, the mutation rate is rapid, and a variant of SARS-CoV-2 typically carries a

combination of multiple mutation sites, resulting in a multiple mutant strain [175]. The effect of impacts from mixes of various transformation locales on viral disease exercises is of extraordinary concern. In conclusion, drug targets for highly conserved sites in the SARS-CoV-2 genome may warrant increased focus on maintaining antiviral activities against variants.

1.8. SARS-CoV-2 M^{pro} Action Mechanism

SARS-CoV-2 M^{pro}, like other proteases, is responsible for the cleavage of peptide bonds. Therefore, different proteins' cleavage sites are recognized and cut by binding to the active site of the main protease. SARS-CoV-2 Main protease (M^{pro} or $3CL^{pro}$) is a cysteine protease with 306 amino acids.

It has been observed that this protease binds to many proteins with high specificity. As a result, the enzyme can easily develop resistance against drug molecules designed for it, even though mutations that may not be in the active site. This protease binds to many different proteins with high specificity. In other words, this enzyme can attach to specific proteins distinctly. However, as a result of this situation, the enzyme can easily develop resistance against designed drug molecules. Drugs usually attempt to hinder the function of these types of enzymes by targeting them. However, if this enzyme undergoes changes that lead to resistance against drugs, the drugs can become ineffective. This situation implies the development of resistance against drugs that are designed to hinder the effective functioning of the enzyme. The particularly emphasized point is that this resistance can develop not only in the specific region known as the "active site" of the enzyme but also through mutations outside this region. In other words, drugs can develop resistance not only against a specific region of the enzyme but also against changes in its overall structure. In conclusion, it can be inferred that a specific protease can bind to various proteins with high specificity, and this characteristic can result in the development of resistance against drugs.

Studies have reported that it cleaves the amino acid sequence of A-X-L-Q-(A/S) (where X can be any amino acid) after glutamine (Q). As substrates, nsps with this recognition sequence bind to the catalytic region (with the catalytic dyad of His41 and

Cys145) and cleaved (Table 1.1). The glutamine residue (Q) in the P1 position of the substrate is cleaved by M^{pro} using the protease Cys145-His41 dyad, where the cysteine thiol serves as the nucleophile in the proteolytic reaction. However, in this sequence of recognition, only glutamine (Q) at P1, leucine (L) at P2 and serine/alanine (A/S) at P1' are specific residues and preserved in all substrates with a few exceptions. The amino acid at P3 is totally non-specific while the P4 region is mostly alanine or valine. There are also nsps with threonine (T) or proline (P) in the P4 region.

Table 1.1: The amino acids of the SARS-CoV-2 M^{pro} and substrate in the cleavage mechanism.

Enzyme M^{pro}	Pocket	S4	S3	S2	S1	catalytic dyad	S1'
	Amino acid(s)	L167, Q192, M165, T190, A191, P168, E166	M165, E166, N142, Q189	H41, M49, P52, Y54, H164, M165, D187, R188, Q189	F140, N142, S144, C145, H163, M165, H172	H41, C145	T25, T26, H41, N142, G143, A145
Cleaved substrate	Recognition site	P4	P3	P2	P1	peptide bond	P1'
	Recognition sequence	A/V	X	L	Q	cleaved	A/S

In a recent study, Shaqra et al. [176] demonstrated that the preserved 3-dimensional structure of the substrate binding region is responsible for recognition and specificity [9]. The intermolecular interactions between the enzyme and substrate in this region play a crucial role in forming the preserved 3-dimensional structure and thus substrate recognition. Therefore, identifying the most dominant substrate residues in the recognition mechanism is of great importance. This will also provide insights into the resistance of SARS-CoV-2 M^{pro} mutations to drugs without compromising substrate recognition.

Recently, complex structures of M^{pro} with some substrates and the corresponding cleaved fragment products of substrates (M^{pro}-fragment complex structures) have been reported [9]. Figure 1.2 illustrates the important regions of the substrate that form complexes with the main protease.

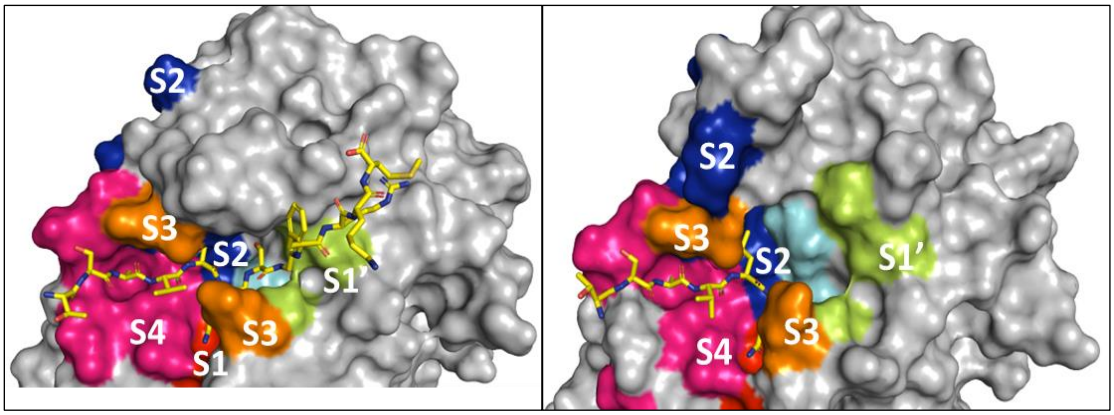

Figure 1.2: SARS-CoV-2 M^{pro} – substrate (peptide) (first picture 7T70 and second one is 7MB4. 7MB4 is cleaved one of 7T70) complex and recognition regions. Cleavage occurs in the P1-P1' region. Colored regions on the surface representation of M^{pro} corresponds to S1-6 interaction regions as shown on Table 1.1.

1.9. Mechanism for Drug Discovery and Protein-Peptide Interactions

Numerous cellular processes are governed by protein-protein interactions (PPI). Many diseases in humans can be traced back to the disruption or disorganization of these intricate interactions. As a result, inhibitors that target specific protein-protein interactions are the subject of intense research. Hence, the creation of new drug candidates places complex protein-protein interactions at the center of their development. The discovery of numerous proteins and interaction interfaces as potential drug targets has been facilitated by the emergence of "omic" technologies like genomics, transcriptomics, and proteomics, which have greatly accelerated our comprehension of protein-protein interaction networks.

A peptide is usually based on the sequence of the binding region between two proteins. Linear sequences can often arise from a loop within a structural domain or from a disordered region between protein terminals or specific areas. To accomplish the ideal viability, the designed peptide should target and tie to the fitting site, following vehicle to the right cell compartment. A short protein sequence that reaches, binds, and modifies the function of the target protein for a specified duration and efficacy would be an excellent

therapeutic agent. These linear sequences can be hard to find because of their short length and propensity to get comfortable cluttered districts in proteins.

Protein and vesicle transport, gene expression, DNA repair, cytoskeleton control, and signal transduction cascades with targeted protein degradation are all regulated by dynamic protein interactions [177]–[179]. When designing peptide inhibitors in response to a protein-protein interaction, peptides derived from the binding site of an inhibitory protein frequently serve as the starting point. X-ray crystallography or NMR techniques may be used to obtain coordinate and distance information about the binding interface. Alanine scanning mutagenesis [180] chemical modification, mass spectrometry, and phage display [181] are additional experimental strategies for locating interface residues. Peptide binding does not alter the conformation of the peptide binding site, which is typically a large, shallow pocket on the protein surface. Additionally, interactions with "hot spot" residues and hydrogen bonds with the peptide backbone contribute enthalpically to protein-peptide recognition. Aromatic residues with hydrophobic residues like Leu and Ile dominate the protein-peptide interface. More hydrophobic residues than the protein surface and structural motifs found in protein folding have been shown to characterize the protein-peptide interface, which shares a structure with the protein core [182], [183]. Conserved protein recognition domains and short linear peptide motifs with less than ten amino acids have been found to engage in numerous protein interactions [184]–[187]. Typically, members of a particular family of protein domains recognize a consensus motif, but they may also recognize various variants of this motif and possess distinct binding preferences [187]–[193]. Peptides can cooperate with globular protein spaces in a wide assortment of ways. Peptide interactions include binding to slits in extended beta or proline type II helix conformations, adopting a helical conformation, or binding to a protein domain forming an additional beta structure. Obstructing intracellular PPIs has been incredibly challenging for bigger or level restricting locales, for example, antibodies that can't cross the cell film. As of late, the more modest size of the peptides and the conformational equilibrium of firmness and adaptability have made them promising applicants with the possibility to target troublesome restricting connection points with good restricting proclivity and explicitness. Resolving and characterizing peptide-protein recognition mechanisms are important for developing peptide-based

strategies that intervene in endogenous protein interactions or enhance the binding affinity and specificity of existing approaches. Studies have uncovered which substrate buildups are most prevailing in the acknowledgment system of SARS-CoV-2 primary protease. As a result, SARS-CoV-2 has made a significant scientific contribution to the design of drugs that can target the primary protease. This will likewise give data about the opposition of transformations in the SARS-CoV-2 principal protease to existing or to be created therapeutics without debilitating substrate acknowledgment.

Based on the newly crystallized structures of peptides with the M^{pro} of SARS-CoV-2, this thesis work reveals the substrate recognition mechanism of M^{pro} by utilizing molecular dynamics and free energy calculation methods to determine M^{pro}-substrate binding energies and how the binding energy changes with mutations in the peptides. Classic molecular dynamics simulations were conducted using the AMBER software in the study, followed by the calculation of end-state free binding energies. By elucidating the sequence that best identifies peptides by assessing the rates of decreasing or increasing binding energies of the structures, an attempt has been made to contribute to drug development efforts for the M^{pro} of SARS-CoV-2.

2. THEORY: FREE ENERGY CALCULATIONS

2.1. MM-P(G)BSA (Molecular Mechanics Poisson-Boltzmann Surface Area)

It is acknowledged that high-level quantum mechanical (QM) techniques offer the most comprehensive and accurate descriptions of the structures, dynamics, and functions of molecules. However, due to their efficiency and reasonable accuracy, classical approaches have been still used commonly for many biochemical systems that are frequently overly complex or for biochemical processes that take a long time (at least nanoseconds of time scales). For molecular recognition models like protein-ligand binding interactions, Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PBSA) has emerged as an effective and dependable approach. Most of the time, there aren't many computational approaches for prediction of binding free energy of large and complex biomolecular recognition. Low efficiency and slow convergence are the main reasons for this.

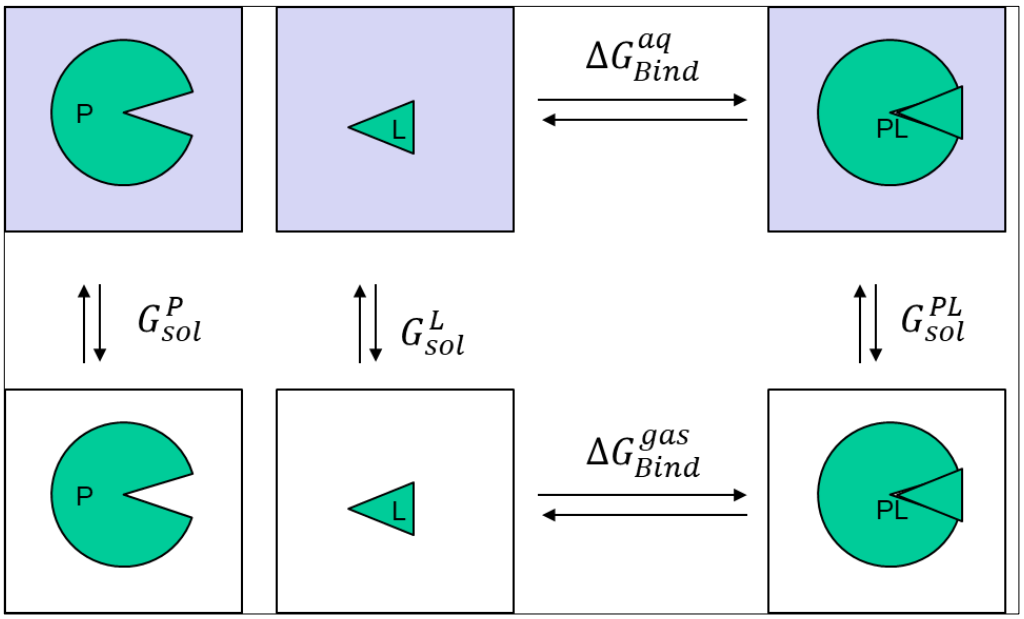

Figure 2.1: Thermodynamic cycle used in MM-PBSA calculations.

The MM-PBSA method's use as an effective and reasonable strategy for free energy simulations is made possible by several of its fundamental approaches. A continuum solvent model approximates the solution's contribution to the free energy using the PBSA model. The binding of a ligand (or a peptide) to a protein in aqueous is decomposed into the collection of several energy terms. Figure 2.1: Thermodynamic cycle used in MM-PBSA calculations. shows the thermodynamic cycle and energy components.

The MM-PBSA method is frequently utilized in large-scale biomolecular recognition as well as the calculation of the binding free energies (ΔG_{bind}) of small molecules to large biomolecular receptors. In an aqueous solvent, the approximate binding free energy ($\Delta G_{bind,aq}$) of the bound ligand-receptor complex is as follows:

$$\Delta G_{bind}^{aq} = \Delta H - T\Delta S \approx \Delta E_{MM}^{gas} + \Delta G_{solv} - T\Delta S \quad (2.1)$$

$$\Delta E_{MM}^{gas} = \Delta E_{cov} + \Delta E_{ele} + \Delta E_{vdW} \quad (2.2)$$

$$\Delta E_{cov} = \Delta E_{bond} + \Delta E_{angle} + \Delta E_{torsion} \quad (2.3)$$

$$\Delta G_{solv} = \Delta G_{pol} + \Delta G_{np} \quad (2.4)$$

Here, the changes in molecular mechanical energy in the gas phase, free energy in the solution, and conformational entropy change upon binding are represented by ΔE_{MM} , $\Delta G_{bind,solv}$, $-T\Delta S$, respectively. The ensemble average is used to calculate these changes across a set of sampled conformations. ΔE_{MM} can be further decomposed into bonded (covalent) and non bonded terms. The ΔE_{cov} arises from covalent bonds with (ΔE_{bond}), point terms (ΔE_{angle}), and twist terms ($\Delta E_{torsion}$). However, in the protein-ligand complex and structures containing only ligands, these terms balance each other out for intermolecular non-covalent interactions; the ligand can be another protein, a substrate, a peptide, or an inhibitor. Non-covalent terms are calculated by the sum of van der Waals (ΔE_{vdW}) and electrostatic (ΔE_{ele}) terms. The polar (ΔG_{pol}) and nonpolar (ΔG_{np}) contributions to the solvation free energy variation ($\Delta G_{bind,solv}$) are typically separated. The normal mode method typically uses a portion of selected snapshots to approximate

the entropy term, which is the most challenging term to calculate. The MM-PBSA method typically begins with either a single trajectory molecular dynamics (MD) simulation of the complex or three separate MD simulations of the complex, receptor, and ligand in the multi-trajectory approach to calculate the average of all ensembles for the calculation of binding affinity. The average values and uncertainties of various quantities of interest are then calculated using the snapshots, retrieved from the trajectory of the MD simulations. Before performing any calculations, MD simulations are frequently run on an explicit solvent model to obtain the most accurate snapshots. It is essential to collect many distinct snapshots or conformations within a reasonable time period for use in subsequent statistical analysis [194].

2.1.1. Polar solvation term

The initial computation of the polar solvation term (ΔG_{pol}) in equation (2.4) involved the numerical solution of the Poisson–Boltzmann equation (PBE) using a finite difference (FD) method. When considering a biomolecular system without any mobile ions, the Poisson equation can be expressed as:

$$\nabla \cdot \varepsilon(\boldsymbol{r})\nabla\varphi(\mathbf{r}) = -4\pi\rho(\boldsymbol{r}) \tag{2.5}$$

where $\varepsilon(\boldsymbol{r})$ is the variation in dielectric constant with position, $\varphi(\mathbf{r})$ is the electric potential, $\rho(\boldsymbol{r})$ is the charge density of the solution. Nevertheless, in most biomolecular systems, the existence of salt in the solution necessitates solving the following equation to obtain the electrostatic potential $\varphi(\mathbf{r})$:

In this equation, $\lambda(\boldsymbol{r})$ represents the ion-exclusion function, which is predefined.

$$\nabla \cdot \varepsilon(\boldsymbol{r})\nabla\varphi(\mathbf{r}) + 4\pi\lambda(\boldsymbol{r})\sum_{i=1}^{N} z_i.e.c_i.e^{\left[\frac{-z_i.e.\varphi(\mathbf{r})}{k_BT}\right]} = -4\pi\rho(\boldsymbol{r}) \tag{2.6}$$

Its value is 0 within both the Stern layer and the interior of the molecule, while it takes on a value of 1 outside the Stern layer. Additionally, z_i denotes the charge of ion type i, c_i represents the bulk number density of that ion type far from the solute at a given temperature T, and e corresponds to the charge of an electron. This PBE equation is non-linear and needs to be solved using numerical approaches. However, when electric field and ionic strength are both small (i.e., $\frac{-z_i.e.\varphi(\mathbf{r})}{k_BT} \to 0$), the equation can be linearized, which is easier to compute. The linear form is given by:

$$\nabla \cdot \varepsilon(\boldsymbol{r})\nabla\varphi(\mathbf{r}) - \varepsilon_{sol}\kappa^2\varphi(\mathbf{r}) = -4\pi\rho(\boldsymbol{r}) \tag{2.7}$$

where $\kappa^2 = \frac{8\pi e^2 I}{\varepsilon_{sol}k_BT}$ is the Debye-Hückel parameter. ε_{sol} and I are the solvent dielectric constant and the ionic strength of the solution, respectively. Even the linear equation is difficult to solve analytically, and accurate numerical methods need to be used. The major difficulty is the continuum distribution of the electric field over the 3D space and the field has to be mapped onto atoms. In the finite difference (FD) method, the 3D space covered by atoms are replaced in a mesh (with small grids) and atomic charges are mapped to these grid points. Then, solvent (exterior) and solute (interior) dielectric regions are mapped into the edges of these FD grids. In AMBER package, PBSA module is used to solve linear or nonlinear form of PBE using FD method and several alternative numerical methods such as geometric grid, conjugate gradient (CG), incomplete Cholesky conjugate gradient (ICCG), and successive over-relaxation (SOR), of which the details are beyond the scope of the current thesis.

As an approximation to Poisson-Boltzmann (PB) term, a faster and efficient model, Generalized Born (GB) has been developed. In this model, atoms are represented as charged spheres with much lower dielectric constant than the environment (solvent). The charges on atoms are screened by their local environment. Since the surrounding atoms will have less dielectric constant than the continuum dielectric constant of the solvent, the atoms that are buried (i.e., surrounded by other atoms) will have less electrostatics screening. This process is called de-screening. De-screening is used to calculate the Born radius of each atom. Large de-screening (small screening) means strong electric field, so

the large Born radius. A small Born radius represents a weak electric field as if the point charge in the bulk solvent environment. The GB equation is given by:

$$\Delta G_{GB} = -\left(\frac{1}{\varepsilon_{in}} - \frac{1}{\varepsilon_{sol}}\right)\sum_{i,j}\left[q_i q_j r_{ij}^2 + \alpha_{ij}^2 . \left(e^{\frac{r_{ij}^2}{4\alpha_{ij}^2}}\right)^{-1/2}\right] \quad (2.8)$$

where ε_{in} and ε_{sol} are interior (solute) and solvent dielectric constants, respectively. In the case of protein, ligand, or protein-ligand complex $\varepsilon_{in} = 1$ and for aqueous media $\varepsilon_{sol} = 80$ (at 300 K). Similarly, q_i and q_j are partial atomic charges and α_{ij} is the de-screened Generalized Born radius.

There are two main shortcomings of the model. First, the formula totally depends on the GB radius and several methods are available to calculate GB radius. Using different models can yield totally different numerical results for the polar solvation term. Second, this model as with the PB model has the assumption of protein atoms to have much smaller interior dielectric constants and their effects are set to a value of 1 throughout the simulation or independent of protein type. On the other hand, the amino acids with acidic and basic groups will have an impact on the interior dielectric constant. For highly charged systems, this value must be larger than 1 (2, 4, 6, etc.). Therefore, several attempts have been made to test the performance of the use of different values of ε_{in} to accurately model the polar solvation energy term in the binding free energy calculations in MM-PBSA or MM-GBSA. Indeed, in recent studies variable interior dielectric constants [195]–[197] are used in the simulations in which for charged residues such as Arg and Glu ε_{in} is set to 4-8 while for neutral but highly polar residues such as Ser and Thr, it is set to 2-4. The rest of the residues (neutral and non-polar) still use the value of 1-2. In the most recent work, Akkus et al. has shown that using bivalent ε_{in} is generally sufficient for ligands mostly buried in neutral amino acids [198].

However, in the case of protein-peptide or protein-protein interactions, the use of single value of ε_{in} should be sufficient.

2.1.2. Non-Polar Solvation Term

The non-polar contribution to solvation free energy (ΔG_{np}) represents the non-electrostatic van der Waals interactions between the solute and the solvent and stems from two different sources as 1) the cavity formation in the binding region, which corresponds to attractive terms of the Lennard-Jones (LJ) potential in the explicit solvent model and 2) dispersion, which corresponds the repulsive terms of the LJ potential in the explicit solvent model [199]–[205]. The former is the function of solvent accessible surface area (SASA) or solvent accessible volume (SAV) and can be given with an empirical formula:

$$\Delta G_{np}^{SA} = \gamma^{*} SA + b \tag{2.9}$$

$$\Delta G_{np}^{CD} = \Delta G_{disp} + \Delta G_{cav} = \Delta G_{disp} + \gamma^{*} SA + b \tag{2.10}$$

where ΔG_{cav} , ΔG_{disp} are the cavity formation and dispersion to the nonpolar solvation free energy, respectively. The classical approximation is the name given to this equation. Most of the time, all solvent molecules have the same correction term b and surface tension parameter γ. Because they operate at different scales with solvent size, van der Waals (dispersion) free energy and vacancy formation free energy are modeled separately in more recent approaches [204]. The volume enclosed by the solvent accessible surface (SAV) is one way to relate it to the free energy of void formation. The dispersion term (ΔG_{disp}) refers to the energy associated with the interactions that contribute to the overall binding energy. This term plays a crucial role in understanding the molecular interactions involved in a system. There are two methods commonly employed to calculate this term: solvent accessible volume integration and solvent accessible surface integration. In the context of these methods, the atom probe radius and solvent probe radius are key parameters that impact the determination of scale factors used in the calculations. The atom probe radius defines the spatial range around an atom that is considered when calculating interactions, while the solvent probe radius relates to the effective size of the solvent molecules. Therefore, when evaluating the dispersion term, the choices made regarding these radii can significantly affect the calculated values and the accuracy of the

results. It's important to carefully consider these parameters to ensure a comprehensive and precise assessment of the dispersion energy contribution within the system.

2.1.3. Entropy Term

The calculations of entropy changes from the initial state to the final state of the system are quite costly and challenging. The entropy change in the protein-peptide binding process consists of translational, rotational, vibrational, and conformational entropy components. Translational and rotational entropy changes contribute very small values and are considered constant. However, vibrational, and conformational entropy changes are important thermodynamic quantities that need to be considered. The following methods are commonly used for entropy calculations:

Normal Mode Analysis-Based Approach: By performing normal mode analysis on trajectories obtained from a classical molecular dynamics (MD) simulation, all vibrational modes can be determined, and the entropic term can be expressed as a combination of these modes. Performing normal mode analysis in classical MD simulations is computationally expensive, so instead of using all trajectory frames (thousands of frames), only a small sample (hundreds) is typically used. After producing vibrational energy levels of a system by performing normal mode analysis, statistical thermodynamics can be utilized to find the entropy.

IE (Interaction Entropy) Approach: Zhang et al. [199] demonstrated a method called Interaction Entropy (IE) that can calculate the entropy change in peptide-protein binding at a much lower cost.

C2 Approach: The exponential average of the free energy given in the IE method can be written as a Taylor expansion. Using the second order cumulant approximation (C2), the entropy change can be calculated from the square of the standard deviation (variance), which can be easily obtained from the fluctuation of interaction energy throughout the MD simulation.

Conformational Entropy: For this entropy change, separate MD simulations will be performed for the protein's apo-form (without the peptide) and holo-form (with the peptide), as well as for the peptide in water. Covariance matrices will be constructed using

the atomic coordinates from each simulation trajectory, and principal component analysis (PCA) will be performed. Eigenvalues and eigenvectors will be extracted, and entropies will be calculated using a quasi-harmonic approximation. Such calculations are automated in both GROMACS and AMBER software.

As with previous studies, we have completely ignored the entropic contribution in the free energy calculations.

3. COMPUTATIONAL METHODS

3.1. System Preparation

For binding studies, crystal structures of 15 protein-peptide complexes of the SARS-CoV-2 Mpro main protease were obtained from the Protein Data Bank (PDB). Pymol was utilized for visual analysis of the PDB files, and all water molecules were removed. Ions, cofactors, and other components that were not particularly important for the planned simulation were also deleted. The chains used for the proteins shown in Table 3.1. And some amino acids in the crystal structures had missing atoms. These are shown in the Table 7.1. Pymol-Mutagenesis was employed to replace these missing atoms, as the system cannot function without them.

Table 3.1: SARS-CoV2-M^{pro}-C145A crystallization in a complex with substrate (non-cleaved) and product (cleaved) residues of nonstructural proteins (nsp) (Single-letter amino acid codes of cleavage site sequences. While bold letters indicate fully resolved residues, the blue ones represent truncated side chains in the co-crystal structures. Underlined N-terminal sequences correspond to product complexes as per independently determined co-crystal structures.

Substrates	PDB ID	Substrate residues	Product residues	PDB ID
nsp4-nsp5	7T70	**TSAVLQ|SGFRKM**	TSAVLQ|	7MB4
nsp5-nsp6	7T8M	SGVTFQ|SAVKRT	SGVTFQ|	7MB5
nsp6-nsp7	**	KVATVQ|SKMSDV	KVATVQ|	7MB6
nsp7-nsp8	7T8R	**NRATLQ|AI**ASEF	NRATLQ|	7MB7
nsp8-nsp9	7T9Y	**SAVKLQ|NNEL**SP	SAVKLQ|	7MB8
nsp9-nsp10	7TA4	**ATVRLQ|AGNAT**E	**	**
nsp10-nsp11	7TA7	REPMLQ|SADAQS	REPMLQ|	7MB9
nsp12-nsp13	7TB2	PHTVLQ|AVGACV	**	**
nsp13-nsp14	7TBT	**NVATLQ|AENV**TG	**	**
nsp14-nsp15	**	TFTRLQ|SLENVA	**	**
nsp15-nsp16	7TC4	**FYPKLQ|SSQ**AWQ	**	**

3.2. MD Simulations

All simulations were performed using AMBER22 software along with AmberTools20 packages [206]. The ff99SB Amber force field was used in the topologies of proteins and peptides. A truncated octahedron box was centered on the protein-peptide complex. Using TIP3P model-type water, each system was solvated at a cell margin distance of 10 Å for each size. The required amount of Na^+ or Cl^- ions were added to the system to balance the charges of the systems. Molecular dynamics simulations were performed in 250,000,000 steps (500 ns). The number of volume change attempts performed as part of the Monte Carlo barostat was set to 100 (mcbarint=100). All H atom containing bonds were constrained to classical harmonic potential using the SHAKE algorithm. The Verlet cut off method was used to minimize energy up to a force of 100 $kJ.mol^{-1}$. nm^{-1}. For long-range electrostatic and van der Waals interactions, the Particle Mesh Ewald method (PME) was used to calculate a nonbonded cutoff of 8 Å in all cases. The update frequency for the neighbor list is set to dt=0.002 ps. Methods for energy minimization and equilibration were used in three stages. Up to 10000 cycles of Steepest Descent were used in each minimization step. Using Langevin Dynamics, each system was stabilized in three steps following minimization. The NVT ensemble went through the first step, which took 2 nsec. The interaction between protein and peptide, along with structural relaxation, was characterized by categorizing them into two temperature clusters at 310 K. In the subsequent stages, under NPT conditions, the simulations were performed by adapting the structures to a reference strain of 2.0 picoseconds at 1 atm pressure. This was followed by applying the Berendsen barostat technique for 1 nanosecond using the Parrinello-Rahman isotropic tension coupling approach. Upon achieving equilibrium, a molecular dynamics (MD) simulation was conducted under NPT conditions for a duration of 500 nanoseconds, and this process was replicated three times.

3.3. Free Energy Calculations

The AmberTools package [206] was employed to compute MM-PBSA and MM-GBSA by utilizing trajectories generated from MD simulations through the cpptraj module

and MMPBSA.py script. The trajectory was in the NetCDF format. was used to carry out these calculations. The single trajectory approach was used in the MM-PBSA calculations. Only the protein + peptide complex part from the trajectories was extracted and default parameters of exterior dielectric constant of water with $\varepsilon = 80$ and interior solute with $\varepsilon = 2$ were used. For the nonpolar contribution to the solvation free energies, one-term approach (SASA only) was used, which defined by solvent accessible surface area (SASA) with empirical parameters of b=-0.5692 kcal/mol and γ=0.0378 kcal/ (mol. A^2). SASA calculations were performed using igb=5 and ipb=2. Additionally, per-residue decomposition was employed, where 1-4 electrostatic interactions (1-4 EEL) were incorporated into the electrostatic energy (EEL) term, and 1-4 Van der Waals interactions (1-4 VDW) were included in the Van der Waals (VDW) potential terms.

4. RESULT AND DISCUSSION

4.1. RMSD and RMSF Analysis

We have explored the protein-peptide complex stability over the course of MD simulation by analyzing root mean square deviation (RMSD). We have compared these RMSD values from two different perspectives. First, we have checked whether there is any sequence specific difference in the RMSD values of cleaved or non-cleaved structures Figure 4.1. Second, we have checked whether there is any stability gain or loss upon cleavage (Figure 4.2). We have also compared the RMSD of peptides only with respect to crystal structures along with alpha-carbon atoms of Mpro (Figure 7.16-17).

Overall RMSD of almost all protein-peptide complexes (whether cleaved or not) reaches a plateau confirming sufficient sampling of the simulations. In comparison of the products (cleaved peptides) bound M^{pro} complexes, the RMSD values are steady except for NSP4 which is slightly instable with RMSD of 5 Å as opposed to the rest ranging between 2.5 Å and 4 Å. On the other hand, the comparison of substrates (non-cleaved peptides) bound M^{pro} complexes shows that some of the NSPs form more stable complexes with Mpro over the other substrates. Protein RMSD values get too high (up to 8 Å) in the case of NSP12-NSP13. This substrate contains proline and histidine as the P6 and P5, respectively. These residues have cyclic groups on side chains on contrary to other substrates, which are all aliphatic. In the case of NSP5-NSP6, the RMSD values is the smallest and steadiest among other substrate (non-cleaved) or apo- structures, which indicates much stable complex formation. This substrate has PHE in the P2 region, which is unusual when compared to other substrates, with leucine as listed in the recognition sequence. This variation in the recognition sequence of P2 gives extra stability to protein structure, which might affect M^{pro}'s specificity for this substrate.

There is also a general trend among the products (cleaved) and substrates (non-cleaved). When a substrate is bound to M^{pro} the RMSD are relatively lower, which might indicate that the complex becomes instable upon the cleavage process. This tendency

might also occur due to the extended interaction of non-cleaved peptides (up to 12 amino acids), which are longer than the substrates (with 6 amino acids)

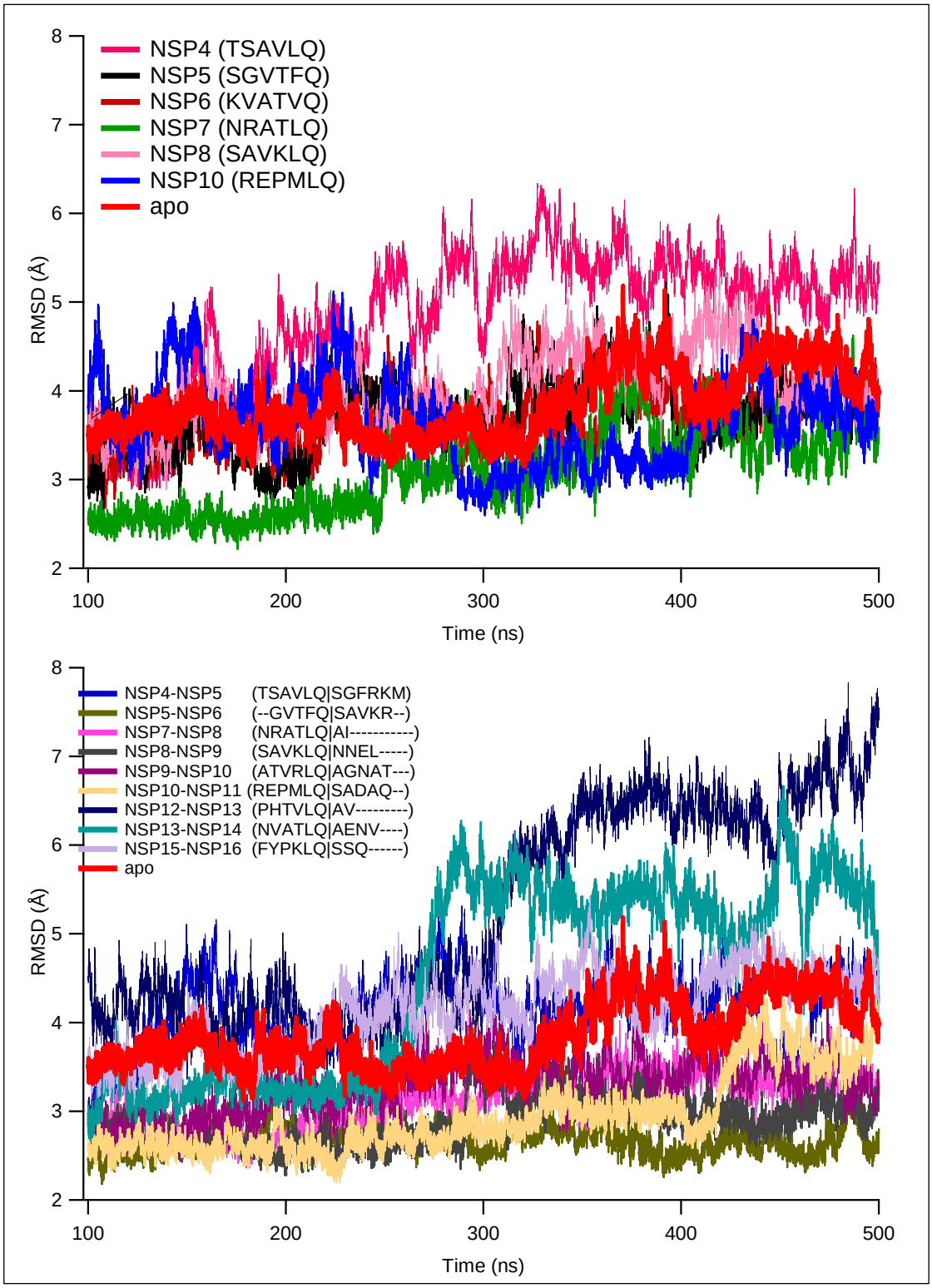

Figure 4.1 RMSD of protein-substrate complexes before and after cleavage. The values are the averages of three replica MD simulations.

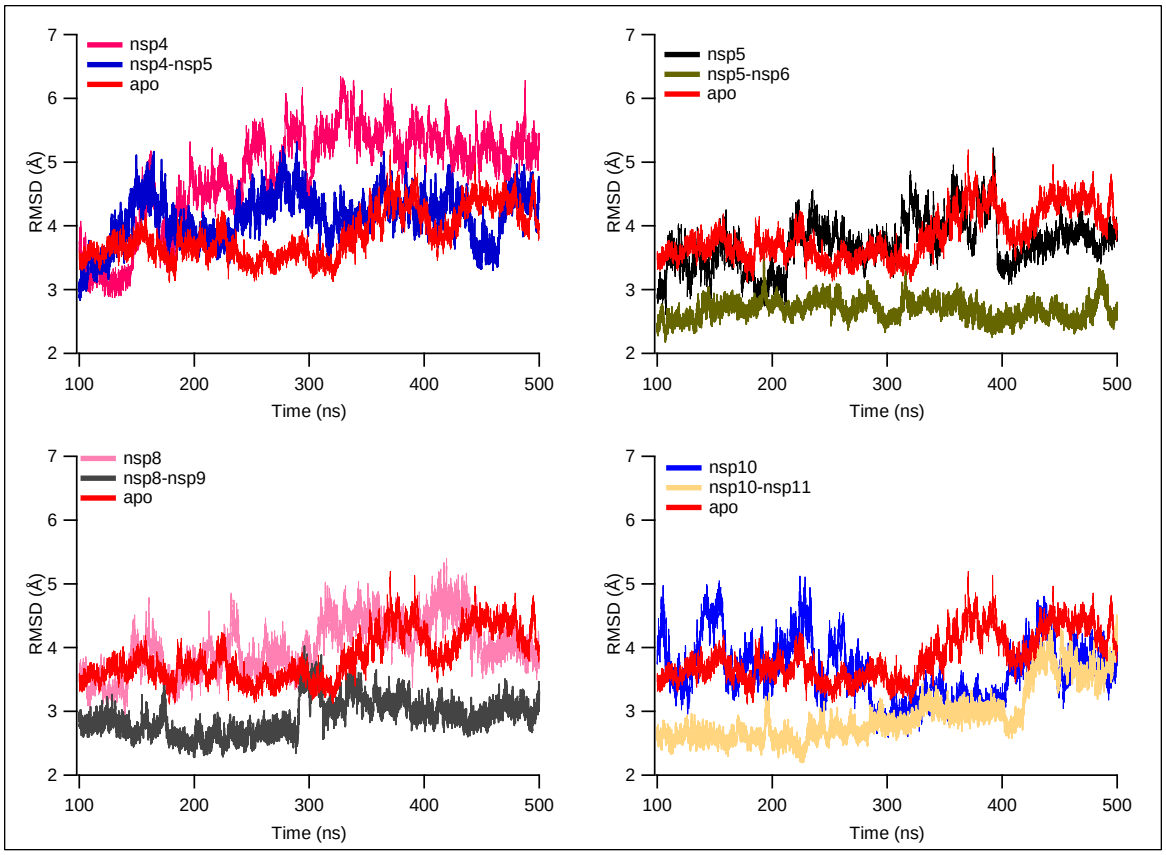

Figure 4.2 RMSD of protein-substrate complexes according to the different NSPs bound and cleaved. Apo state: In the RMSD graph for the APO state, we are examining the protein's natural, unliganded condition. This graph focuses on measuring changes in the protein's structure over time. The RMSD values indicate how the fundamental structure of the protein is changing. Cleaved state (nsp4, nsp5, nsp8, nsp10): The RMSD graph under the cleaved peptide state represents the condition where a specific peptide segment is removed from the protein. This graph shows how the protein's structure changes when the peptide is cleaved and helps us analyze the effects of the absence of the cleaved peptide. Non-cleaved state (nsp4-nsp5, nsp5-nsp6, nsp8-nsp9, nsp10, nsp11): The RMSD graph under the non-cleaved peptide state represents the condition where the protein has the same peptide segment, but it is not cleaved. This graph helps us assess the impact of the peptide's integrity on the protein structure. By comparing these four graphs, we can understand the structural differences between the APO state, the cleaved peptide state, and the non-cleaved peptide state. The RMSD values allow us to quantitatively measure these differences, while the curves of the graph illustrate how changes occur over time.

We have also compared the root mean square fluctuations (RMSF). Overall, the fluctuation of SARS-CoV-2 M^{pro} residues increases upon peptide bonding independent from substrate or product (

Figure 4.3). On the other hand, there are several differences between apo- structure and substrate/product bound complexes. In particular, the 38-52 region with the maximum fluctuating residue of 46, cleaved peptides (products) gain additional stability to the RMSF values. The case is the opposite in the non-cleaved (substrate) complexes. Interestingly, this region covers the S2 recognition site. The residues of 187-192, which are also in S2 binding pocket, fluctuate upon peptide binding and it gets even more instable when the peptide is cleaved. The differences among fluctuations in this region is prominent in the case of NSP4 and NSP4-NSP5, which is the N-terminal autocleavage of Mpro. A significant fluctuation in the residues at S1 and S1' regions is not observed upon peptide (substrate or product) binding. The residues T190 and Q192 in S4 region are also affected by the cleavage of NSP5-NSP6 substrate to NSP5 product.

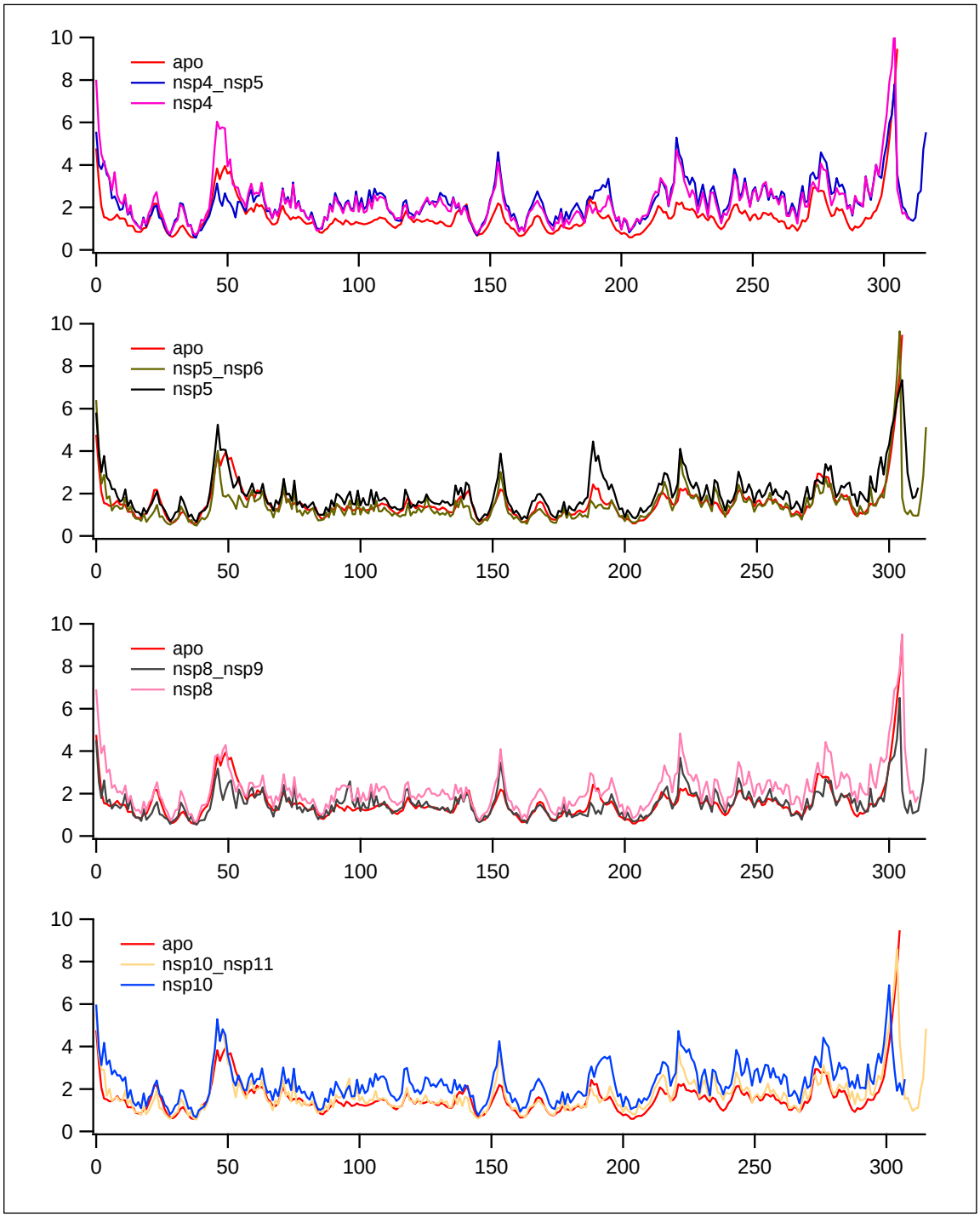

Figure 4.3 RMSF (y-axis) of M^{pro} residues (x-axis) complexed with different substrate (non-cleaved) and product (cleaved) peptides.

4.2. Cluster Analysis

Cluster analysis is a data analysis method based on the similarities of statistical data, enabling the grouping of data points. Its objective is to gather data points with similar attributes into the same group and separate those with different attributes into distinct groups. This allows for understanding the structures within a dataset and uncovering patterns. In the cluster analysis, the complex parameter topology without water and trajectory files from MD outputs were utilized. The similarity threshold, "epsilon," was set to 4.5 Å, resulting in the creation of 10 clusters. A hierarchical agglomerative (bottom-up) algorithm was used in clustering. Clustering finishes when a minimum distance between clusters is greater than 4.5 Å and a total of 10 clusters is reached with average-linkage approach, in which the average distance between members of two clusters is linked. During clustering, atom selection was specified to include C, N, O, CA and CB atoms for calculating Root Mean Square Deviation (RMSD), excluding H atoms.

As a result of the cluster analysis I conducted, the outcome in Table 7.2 was obtained. From this table, the cluster indicates which group it represents; Fraction represents the proportion of total frames to the entire dataset, indicating how much of the dataset the cluster covers; Average Distance denotes the mean distance value, portraying the average distance between all frames within the cluster.

The most populated clusters of all complexes are in total agreement with experimentally refined crystal structures (Figure 7.1-15), which shows the success of the MD simulations. Overall, the interactions that occur in the crystal structures are preserved throughout the simulations. We should note that there are very slight changes in the phenyl rings of Y54 and H41 when the most populated structures are compared to the x-ray structures. In the complexes of nsp4, nsp6 and nsp10, the cleaved peptides remain the orientation of the crystal structure whereas nsp5, nsp7 and nsp8 peptides seem to reorient their terminal residues during the simulations. This might also indicate the instability of these complexes at the terminal regions. Similarly, non-cleaved peptides deviate from the xray structures at a few of the terminal residues.

4.3. H-Bond analysis

We have also performed H-bond analysis on the trajectories of the simulations of M^{pro} -peptide complexes. Table 4.1 shows the overall H-bond formed between protein and peptides. Non-polar contacts were also analyzed.

Table 4.1: H-bonds between protein and cleaved peptides.

NSP	ES Pair	Hbond	NSP	ES Pair	Hbond	NSP	ES Pair	Hbond
nsp4	E166-V309	0.66	nsp4-nsp5	E166-V309	0.64	nsp9-nsp10	E166-R308	0.62
	T190-S307	0.61		T190-A308	0.47		T26-G312	0.59
	G143-Q311	0.46		T24-R315	0.58		T24-A314	0.45
	A145-Q311	0.40		T26-G313	0.54		A145-Q310	0.36
	T190-A308	0.57	nsp5-nsp6	T26-A312	0.72		T190-V307	0.33
nsp5	E166-T310	0.55		E166-T308	0.58	nsp10-nsp11	T26-A313	0.64
	H164-Q312	0.34		T24-K314	0.62		E166-M309	0.61
	G143-Q312	0.36		A145-Q310	0.31		T24-A315	0.54
nsp6	E166-T309	0.58		E166-T308	0.58		H164-Q311	0.32
	T190-A308	0.59		T26-A312	0.72	nsp12-nsp13	E166-V309	0.65
	G143-Q311	0.46		T24-K314	0.62		A145-Q311	0.32
	H164-Q311	0.41		A145-Q310	0.31	nsp13-nsp14	E166-T309	0.66
	A145-Q311	0.38	nsp6-nsp7	E166-T308	0.67		H164-Q311	0.57
nsp7	E166-T306	0.73		H164-Q310	0.50		T26-E313	0.37
	T190-A305	0.65		A145-Q310	0.39		A145-Q311	0.33
	G143-Q308	0.43	nsp7-nsp8	E166-K309	0.59		N314-S46	0.36
	A145-Q308	0.37		T26-N313	0.50	nsp15-nsp16	E166-K309	0.56
	H164-Q308	0.31		T24-L315	0.34		H164-Q311	0.31
nsp8	E166-K310	0.47		A145-Q311	0.31		T26-S313	0.35
nsp10	E166-M306	0.60		T190-V308	0.31		A145-Q311	0.40

By van der Waals interactions, we refer to the distance between heavy atoms not involved in H-bonding and salt bridges. These types of interactions due to their nature are much weaker. Therefore, their contribution to binding free energies should be rather limited. The analysis of van der Waals interaction shows that these types of interactions in NSP4, NSP6 and NSP7 are the most dominant and protected throughout the simulations. In addition, P1 and P3 are the most common residues that play a role in these interactions.

Table 4.2: van der Waals interactions between protein and cleaved peptides between Protein (residues on left) and peptides (residues on left).

nsp4	nsp5	nsp6	nsp7	nsp8	nsp10
H41-Q(P1)	H41-F(P2)	H41-Q(P1)	H41-Q(P1)	H41-L(P2)	F140-Q(P1)
H41-L(P2)	F140-Q(P1)	H41-V(P2)	H41-L(P2)	F140-Q(P1)	L141-Q(P1)
F140-Q(P1)	L141-Q(P1)	F140-Q(P1)	F140-Q(P1)	L141-Q(P1)	G143-Q(P1)
L141-Q(P1)	N142-Q(P1)	L141-Q(P1)	L141-Q(P1)	N142-Q(P1)	S144-Q(P1)
N142-Q(P1)	G143-Q(P1)	N142-Q(P1)	N142-Q(P1)	G143-Q(P1)	A145-Q(P1)
G143-Q(P1)	S144-Q(P1)	G143-Q(P1)	G143-Q(P1)	S144-Q(P1)	H163-Q(P1)
S144-Q(P1)	A145-Q(P1)	S144-Q(P1)	S144-Q(P1)	A145-Q(P1)	H164-Q(P1)
A145-Q(P1)	H163-Q(P1)	A145-Q(P1)	A145-Q(P1)	H163-Q(P1)	H164-L(P2)
H163-Q(P1)	H164-Q(P1)	H163-Q(P1)	H163-Q(P1)	H164-Q(P1)	M165-M(P3)
H164-Q(P1)	H164-F(P2)	H164-Q(P1)	H164-Q(P1)	H164-L(P2)	E166-Q(P1)
H164-L(P2)	M165-T(P3)	H164-V(P2)	H164-L(P2)	M165-K(P3)	E166-P(P4)
M165-A(P4)	M165-F(P2)	M165-T(P3)	M165-T(P3)	M165-V(P3)	E166-M(P3)
M165-V(P3)	E166-Q(P1)	E166-A(P4)	E166-A(P4)	E166-Q(P1)	P168-R(P6)
E166-A(P4)	E166-V(P3)	E166-Q(P1)	E166-Q(P1)	E166-K(P3)	
E166-Q(P1)	E166-T(P3)	E166-T(P3)	E166-T(P3)	T190-V(P3)	
E166-V(P3)		P168-K(P6)	Q189-A(P4)		
P168-S(P5)		T190-A(P4)	Q189-T(P3)		
T190-S(P5)		T190-V(P5)	Q189-L(P2)		
T190-A(P4)		H41-Q(P1)	T190-A(P4)		
			P168-N(P6)		
			Q189-R(P5)		
			T190-R(P5)		

1 | 0.9 - 1 | 0.8 - 0.9 | 0.7 - 0.8 | 0.6 - 0.7

4.4. MM-PBSA and MM-GBSA Results

We have calculated the Binding Free Energy (BFEs) by MM-PBSA and MM-GBSA methods, both of which use single trajectory approach and SASA only one-term model for non-polar interactions. When the different products (all with 6 amino acids) are compared, binding free energies calculated by MM-PBSA one term model (SASA_only approach) range from -42 kcal/mol to -68 kcal/mol. Using MM-GBSA model, the same energies are ranged between -42 kcal/mol and -63 kcal/mol. However, both methods yield similar trends among different peptides (Figure 4.4).

The free energy calculations show a clear distinction between the substrate (non-cleaved) and product (cleaved) complexes of M^{pro}. In line with RMSD and RMSF calculations, the substrate complexes have much greater binding energies than the product complexes. This is partly due to the extended interaction by longer chains of the substrates with respect to products. However, there are cases where the chain length is smaller, but the binding energy is greater than the counterparts (Table 4.3). Among the cleaved structures (products), the BFE change is the largest in NSP7 (-63 and -68 kcal/mol) and smallest in NSP5 (-43 kcal/mol). It should be noted that the P2 region of NSP5 is different than the other products. The PHE instead of a common recognition residue of leucine in this region causes dramatic decrease in BFE. In addition, the P5 and P6 residues, although are not reported as common recognition residues of A(P4)-X(P3)-L(P2)-Q(P1), seem to play an important role in binding. Moreover, the greater binding energies are succeeded by amino acids with small length aliphatic side chains such as valine, serine, and leucine in the P1-P3 regions while longer length aliphatic side chains and basic (positively charged) such as lysine or arginine decreases BFEs. On the other hand, the prolonged residues increase BFE when they are in the P5-P6 region. Thus, main conclusions from the BFE values of products binding to M^{pro} might be that the P5-P6 regions play important roles for the recognition of M^{pro} to the peptides and the closest regions to the cleavage site needs to be small length aliphatic amino acids. For the substrate bound complexes of M^{pro}, it is difficult to obtain a direct conclusive result from the data since the x-ray structures of these do not have the same number of amino acids in their structures (i.e., P' regions).

NSP4-NSP5 complex has 6 residues (P1'-P6') while NSP5-NSP6 and NSP10-NSP11 have only 5 residues (P1'-P5'). Similarly, NSP8-NSP9, NSP9-NSP10 and NSP13-NSP14 have 4 residues (P1'-P4') while NSP7-NSP8 and NSP12-NSP13 have only 2 residues (P1'-P2'). Further decomposition analysis is required for these complexes.

Table 4.3. The binding free energies (kcal/mol) calculated by MM-GBSA and MM-PBSA (SASA only) approaches.

		MM-GBSA				MM-PBSA			
Ligand	**Sequence**	**1**	**2**	**3**	**Avg.**	**1**	**2**	**3**	**Avg.**
nsp4	TSAVLQ	-51	-51	-55	-53	-56	-53	-58	-56
nsp5	SGVTFQ	-43	-49	-37	-43	-41	-51	-37	-43
nsp6	KVATVQ	-50	-46	-50	-49	-57	-51	-55	-54
nsp7	NRATLQ	-63	-68	-57	-63	-68	-73	-62	-68
nsp8	SAVKLQ	-40	-51	-38	-43	-46	-55	-40	-47
nsp10	REPMLQ	-34	-46	-50	-43	-38	-48	-53	-46
nsp4-nsp5	TSAVLQ|SGFRKM	-77	-71	-70	-73	-78	-71	-69	-73
nsp5-nsp6	-GVTFQ|SAVKR	-64	-63	-63	-63	-65	-65	-64	-65
nsp7-nsp8	NRATLQ|AI	-63	-51	-58	-57	-64	-55	-59	-59
nsp8-nsp9	SAVKLQ|NNEL	-67	-60	-65	-64	-70	-63	-69	-67
nsp9-nsp10	ATVRLQ|AGNA	-71	-74	-73	-73	-74	-76	-78	-76
nsp10-nsp11	REPMLQ|SADAQ	-65	-64	-63	-64	-68	-64	-62	-65
nsp12-nsp13	PHTVLQ|AV	-60	-65	-61	-62	-58	-59	-58	-58
nsp13-nsp14	NVATLQ|AENV	-68	-58	-60	-62	-68	-62	-63	-64
nsp15-nsp16	FYPKLQ|SSQ	-68	-59	-61	-63	-67	-59	-60	-62

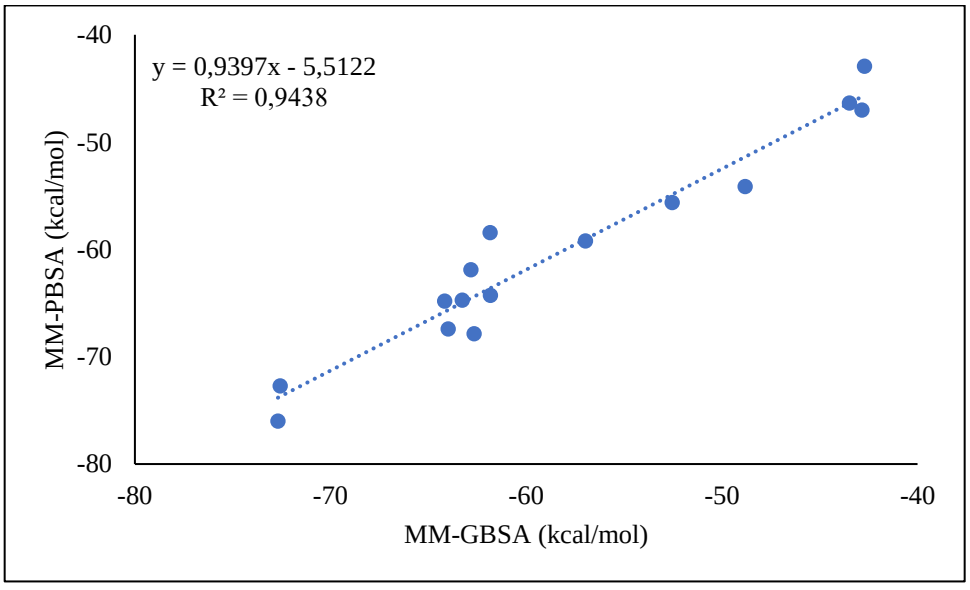

Figure 4.4. Comparison of MM-GBSA and MM-PBSA predicted binding free energies of M^{pro}-substrate/product complexes.

In order to better understand the M^{pro}'s recognition of peptides, we have performed a decomposition analysis (

Figure 4.5). One of the most surprising features of the decomposition analysis is that the GLU at P1 does not have any dominant contribution to the BFE. This residue is the termini in all cleaved peptides. Similarly, P3 residue, which varies from one product to another, does not have a dominant contribution to the BFE. This is in total agreement with the experimental recognition sequence of AXLQ where P3 is the X (i.e., can be any residue). According to the results, P2 and P4 regions are the most important residues for the binding. In addition, the P6 region apparently reduces the binding affinity (in particular for MM-GBSA). Finally, the P5 residue normally does not contribute to BFE except that when this residue is arginine as in the case of nsp7, in which the binding energy dramatically is affected. This might be due to the side chain length or the electrostatic contribution since this residue is positively charged.

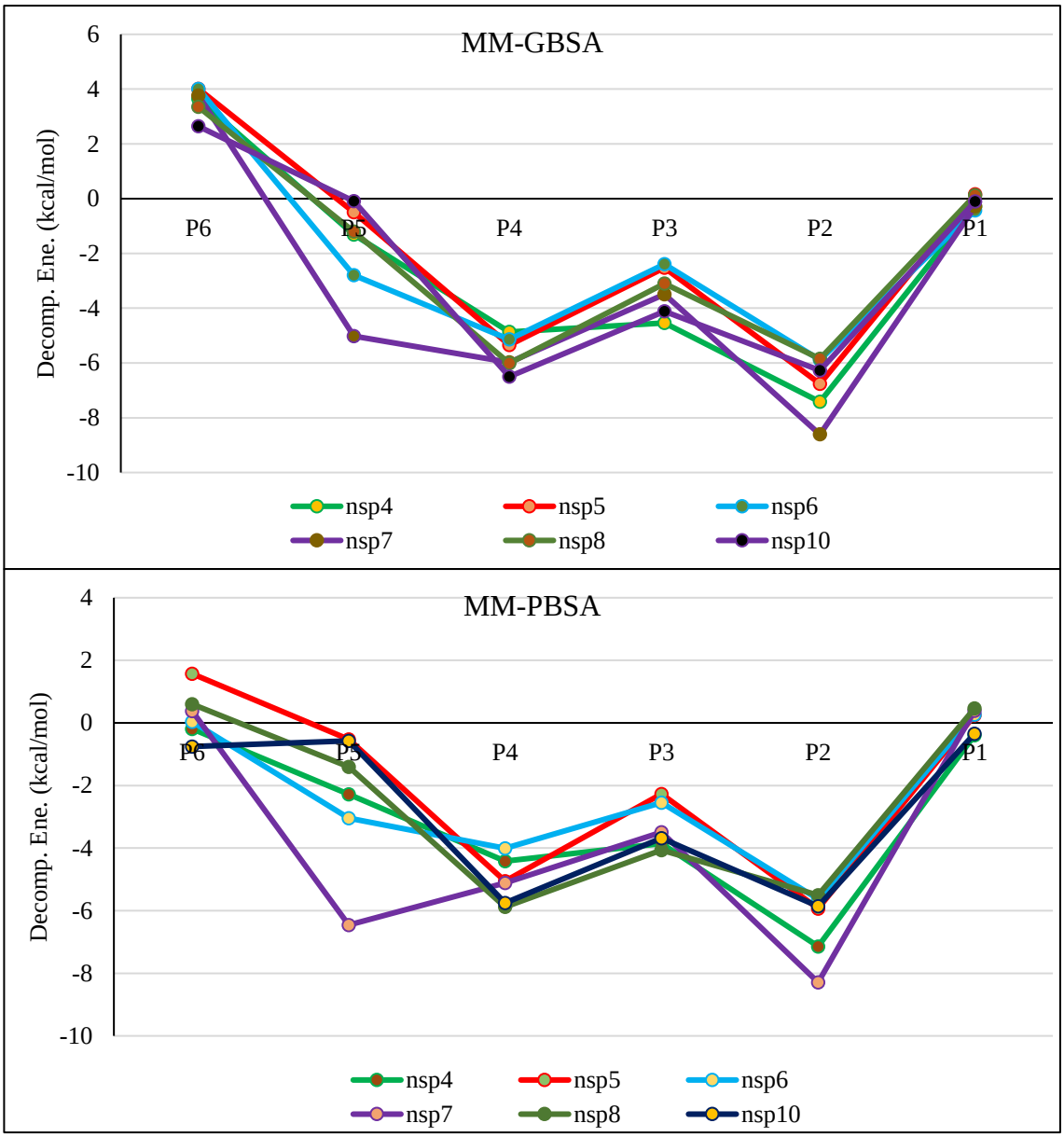

Figure 4.5. Contribution of residues in the cleaved peptides to the BFE between M^{pro} and cleaved peptides.

From the decomposition analysis Figure 4.6, we have observed some common features between substrates (non-cleaved) and products (cleaved) bound to M^{pro}. The terminal residues do not seem to contribute to binding if not weakening the interaction. Considering our MD simulations which were performed without capping the terminal residues, it suggests that the existence of charged groups (-NH3+ or -COO-) does have an adverse effect in binding. In addition, like the product complexes, the existing ARG

residue in the P3 or P5 region favors the binding. Very interestingly, GLU amino acid at the P1 region does not contribute to the binding free energy (BFE) when it is the product (i.e., C-termini) while it contributes to BFE by (-5) - (-7) kcal/mol when it is in the substrate (i.e., non-terminal residue). This explains the regeneration of the enzyme's active site by the mechanism in which the product leaves the active site right after enzymatic turnover and a new substrate to bind the active site. Another interesting result is that the P' regions seem to have relatively much smaller contribution to the binding. P1', although experimentally suggested as alanine or serine as the recognized residue, does have little impact on binding.

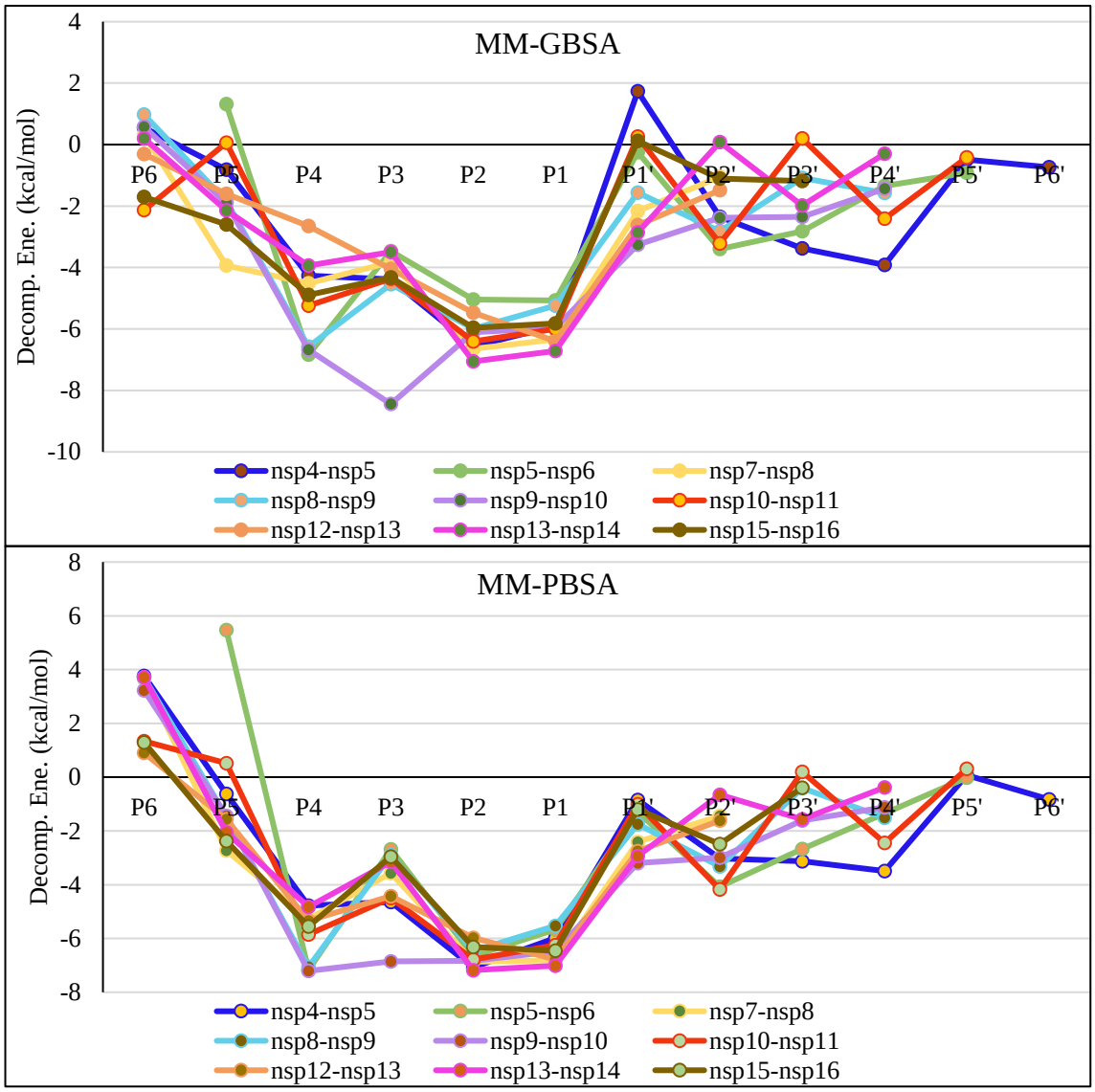

Figure 4.6: Contribution of residues in the cleaved peptides to the BFE between M^{pro} and non-cleaved peptides.

We have also analyzed amino acid dependence of BFE according to their location at each region (

Figure 4.7 and

Figure 4.8). The P1 region has only GLU in all NSP substrate and product complexes. The interaction of this residue in substrate complexes is much larger than the products. LEU is the most common residue in complexes except for three of them in which it is replaced by phenyl alanine or VAL, which do not change the contribution to the binding. The P3 region is the most variable binding site, and as mentioned earlier, having an ARG in this region drastically stabilizes the interaction. Interestingly, when this region is VAL, it has greater interaction. The P4 region can be either VAL or ALA in most cases. Having another CH_3- group seems to affect the contribution, and thus VAL has more contribution to BFE than alanine. Rarely occurring PRO and threonine has similar effects to alanine. The prime (P') regions that correspond to the amino acids that are sequentially occurring after the cleaved peptide have rather smaller contributions to the total BFEs. P1' is either SER or ALA in most of the substrate complexes except in the case of nsp8-nsp9. The data suggests that the contribution is greater when this region is ALA in comparison to SER. Similarly, P2' region is also favored by ALA. ARG also has an impact on BFE in P4' similar to P3 and P5.

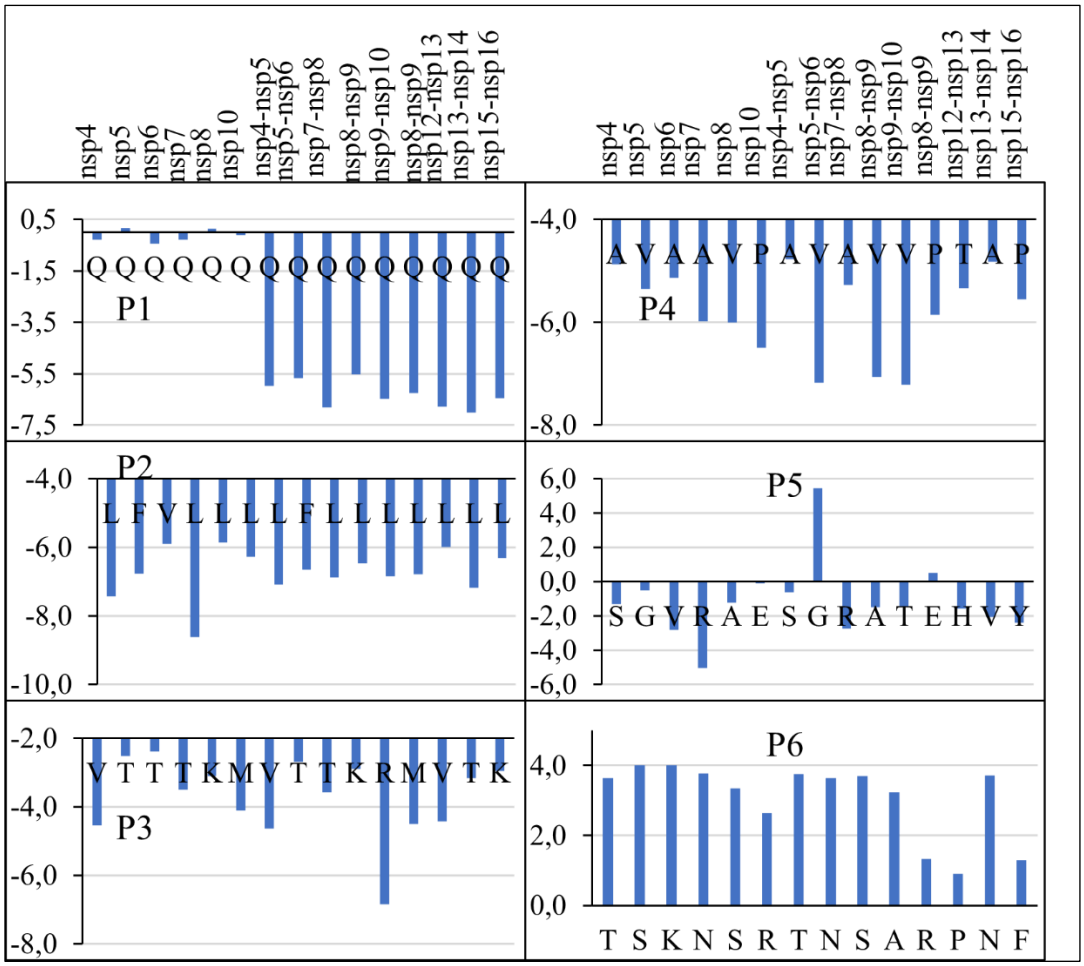

Figure 4.7: The whole protein's (cleaved and non-cleaved) amino acid dependent BFE contribution of regions P1-P6.

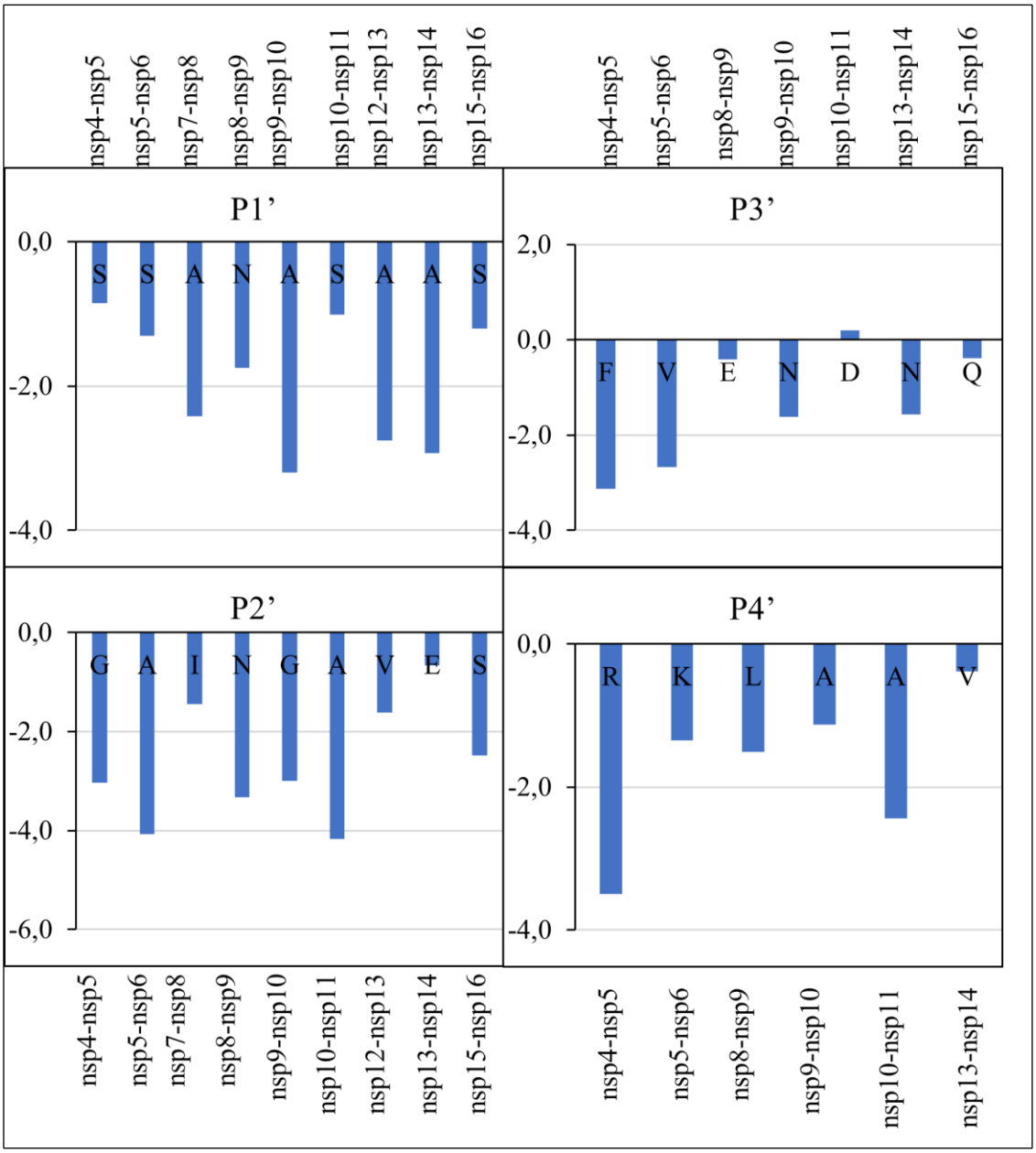

Figure 4.8: The non-cleaved protein's amino acid dependent BFE contribution of regions P1'-P4'.

Table 4.4. The contribution of Mpro residues on binding when complexed with product and substrate peptides. The color map shows the level of interaction with the dark blue as highest and white as no interaction. The numbers indicate energy (in kcal/mol) associated with the interaction.

		Cleaved (product)						Non-cleaved (substrate)								
M^{pro} residue	**Res. Index**	**4**	**5**	**6**	**7**	**8**	**10**	**4 - 5**	**5 - 6**	**7 - 8**	**8 - 9**	**9 - 10**	**10 - 11**	**12 - 13**	**13 - 14**	**15 - 16**
T	**21**							-1.0	-0.8							
G	**23**							-1.5								
T	**24**							-2.6	-2.2		-1.3	-1.4	-2.0			
T	**25**							-2.5	-2.5	-0.6	-2.5	-2.3	-2.6	-0.5	-1.1	-1.8
T	**26**							-3.9	-4.1		-3.4	-3.6	-3.6	-0.6	-0.8	-2.1
L	**27**		-0.6	-0.6	-0.6	-0.5		-1.2	-1.3	-0.9	-1.0	-0.9	-1.2	-1.0	-0.9	-1.2
H	**41**	-3.5	-3.4	-3.6	-4.5	-2.8	-2.0	-1.3	-1.3	-1.2	-2.1	-1.1	-1.3	-1.5	-1.5	-1.4
S	**46**										-0.7				-0.6	-0.7
D	**48**				-1.3					-0.6						
M	**49**		-0.6					-1.9	-1.1	-0.6	-2.0	-1.2	-0.8	-0.8	-2.0	-1.1
F	**140**	-1.4	-1.1	-1.4	-1.4	-1.0	-1.0	-1.0	-0.9	-0.9	-0.7	-0.7	-1.1	-1.0	-1.1	
N	**142**	-2.3	-2.0	-2.5	-2.4	-2.2	-1.8	-0.8	-1.0	-2.4	-1.2	-1.2	-1.0	-3.2	-3.0	-1.6
G	**143**	-3.7	-3.0	-3.8	-3.8	-3.2	-2.8	-1.6	-1.7	-1.6	-1.6	-1.7	-1.7	-1.8	-1.6	-2.0
S	**144**	-2.6	-2.1	-2.7	-2.7	-2.4	-2.3	-0.9	-1.0	-0.9	-1.0	-1.1	-0.9	-1.1	-0.9	-1.2
A	**145**	-3.1	-2.6	-3.1	-3.1	-2.7	-2.6	-2.1	-2.3	-2.1	-1.9	-1.9	-2.2	-2.1	-2.2	-2.1
H	**163**	-2.4	-2.0	-2.4	-2.5	-1.9	-1.6	-2.2	-1.9	-1.8	-1.7	-1.4	-2.3	-1.9	-2.1	-0.9
H	**164**	-1.5	-1.4	-1.7	-1.7	-1.1	-1.0	-1.8	-1.4	-2.0	-1.6	-1.6	-1.7	-1.3	-2.2	-1.8
M	**165**	-3.5	-3.8	-3.2	-3.6	-3.3	-3.4	-3.3	-3.7	-3.5	-3.2	-3.2	-3.4	-3.5	-3.5	-3.5
E	**166**	-3.2	-2.8	-3.1	-3.4	-3.1	-3.0	-3.3	-3.3	-3.5	-3.7	-5.0	-3.6	-3.6	-3.3	-3.4
L	**167**	-0.7	-0.7			-0.5	-0.6	-0.6	-0.6	-0.5		-0.5	-0.7	-1.0		-1.0
P	**168**	-2.1	-1.4	-2.5	-2.5	-1.5	-2.9	-2.1	-1.1	-2.4	-2.1	-2.4	-2.9	-2.4	-2.5	-2.1
Q	**189**	-1.6	-1.3	-1.1	-3.6	-1.0	-1.7	-1.9	-2.6	-2.5	-1.5	-1.8	-1.3	-2.4	-1.4	-1.7
T	**190**	-2.7	-0.7	-2.1	-2.4	-1.4		-2.6	-1.0	-1.8	-2.1	-2.5	-0.6	-2.3	-1.5	-0.9
A	**191**	-1.1	-0.6	-1.1	-1.2	-1.1		-0.9		-1.1	-1.2	-1.3	-0.6	-1.3	-1.0	-1.1
Q	**192**			-0.5										-0.9		-1.5

When the protein residues are investigated (Table 4.4), we see that the cleaved peptides do not have any interaction with the first 139 residues other than the H41, catalytic dyad. Apart from HIS41, there are three distinct regions of M^{pro} for the cleaved peptide complexes: 140-145 (S1), 163-168 (S1-S3) and 189-192 (S3-S4).

On the other hand, non-cleaved peptides have somewhat strong interactions with residues of 23-27 and 49. In particular, THR25 and THR26 (S1') play an important role in the interaction in most of the substrate complexes. The catalytic dyad residue of HIS41 have relatively much smaller interaction with the substrates than the products. The other dyad residue, ALA145 also has less interactions with the substrates than products. We should note that this residue is cysteine normally, but the x-ray structures were refined with ALA. MET49 also plays an important role in binding in a way that upon cleavage the protein starts losing interactions with it. In addition, ASN142, GLY143 and SER144 are stabilizing the product complexes although they do not contribute to the binding as much in the substrates. Moreover, the residues MET165, GLU166, PRO168, GLN189 and THR190 play a role in both substrate and product binding with the former two being more dominant.

5. CONCLUSIONS

In this study, we have thoroughly analyzed the SARS-CoV-2 M^{pro}-peptides interactions where the peptides as the substrates or products of the cleavage are recently co-crystalized. We have used MM-GBSA and MM-PBSA decomposition analysis to understand the mechanistic details of the recognition of peptides. We showed that the amino acids with small aliphatic side chains such as ALA, VAL and glycine are dominant residues in the interactions. We also showed that the GLU residue at P1 has drastic change in the interaction with the M^{pro} protein from substrates to products. For the first time, we suggested that the ARG amino acid at the P3-P5 along with P4' can strengthen the interaction. Experimentally targeting the regions and the amino acids discussed throughout the thesis could have a potential to disrupt the function of M^{pro}.

6. FUTURE DIRECTIONS

Alanine scanning is a valuable method for the scientific analysis of sequences, aiding in the identification of critical residues within biomolecules. In the realm of computational biology, the consideration of conformational entropy changes holds significance, particularly in the context of Binding Free Energy (BFE) calculations. To enhance the accuracy of molecular dynamics (MD) simulations, substituting Cys145 for Ala145 can yield more insightful results. Furthermore, the interaction landscape between whole Non-Structural Proteins (NSPs) and Mpro presents an intriguing avenue for exploration, shedding light on complex formation mechanisms. Notably, investigating Covid-19 variants featuring mutations in Mpro provides a crucial opportunity to comprehend the impact of such alterations on the protein's functionality and potential therapeutic interventions.

7. APPENDICES

Appendix A: Publications within the Scope of Thesis Study

Taşçı H.S., Koçak A., Aydınoğlu F. (2023), "Investigation of Substrate Recognition of Sars-Cov-2 Main Protease Enzyme by Molecular Dynamics and Free Energy Methods", 7th Gebze Technical University Graduate Studies Symposium, Gebze, Kocaeli, Türkiye, 31 May.

Appendix B: Additional Tables and Figures

Table 7.1: Structure of SARS-CoV-2 Mpro: Chain and Missing Atom Information.

PDB ID	Missing Residue Name	Missing Residue Number
7MB4	GLU	47
	LYS	100
7MB5	LYS	5
	ASN	72
	GLN	74
	LYS	100
	GLN	306
7MB6	THR	45
	SER	46
	GLU	47
	LEU	50
	ARG	60
	ARG	76
	LYS	100
	LYS	102
	LYS	137
	ARG	188
	THR	196
	ARG	222
	PHE	223
7MB8	ASP	153
	ASP	155
7MB9	VAL	303
	THR	304
	PHE	305
	GLN	306
	SER	46
	GLU	47
	ARG	76
	LYS	100
	LYS	102
	TYR	154
	ASP	216
	LYS	236
7T70	GLU	47
	TYR	154
	LEU	232
	LYS	236
	LYS	215
7T8M	ARG	215
7TA7	ASN	72
	LYS	100
	LYS	102
	TYR	154
	ASP	197
	ASP	229
	LEU	232
	LYS	236
	ASN	277
	SER	284
	LEU	286
	PHE	294
	SER	301
	VAL	303
	PHE	305
	ARG	256
	GLU	257
	GLN	266
7TB2	GLU	47
	ASN	72

	THR	225
	LEU	227
	ASP	229
	LEU	232
	MET	235
	LYS	236
	ASN	238
	ILE	259
	ASP	263
	LYS	269
	GLU	270
	ASN	274
	MET	276
	ASN	277
	ARG	279
	LEU	286
7MB7	VAL	303
	THR	304
	PHE	305
	GLN	306
	GLU	47
	ASP	48
	ASN	72
	GLN	74
	LYS	100
	ASP	153
	TYR	154
	ASP	155
	LYS	236
	ASN	277
	SER	301

	ASN	72
	ASP	155
	LYS	127
	ARG	128
7T8R	SER	46
	ASN	72
	LYS	100
	VAL	303
	THR	304
7T9Y	GLN	306
	SER	46
	GLU	47
	ASN	51
	ARG	60
	SER	62
	LEU	232
	MET	235
	LYS	236
	SER	118
7TC4	LYS	100
	LYS	102

	LYS	100
	ASP	153
	MET	235
	LYS	236
7TBT	SER	46
	GLN	74
	LYS	100
	TYR	154
	MET	235
	LYS	236
	ASN	277
	GLU	125
	ASN	126
7TA4	PHE	305
	GLN	306
	GLU	47
	ASN	72
	ASP	153
	ARG	222
	LYS	236
	ARG	208

Table 7.2 : Structural Analysis and Cluster Results: Statistical Evaluation of Clustering Performance and Intrinsic Features from Molecular Dynamics Simulation Data.

PDB ID / The most populated # cluster	Fraction	Average Distance(Å)	PDB ID / The most populated # cluster	Fraction	Average Distance(Å)	PDB ID / The most populated # cluster	Fraction	Average Distance(Å)
7MB4 / 0	0.4	2.8	**7MB9 / 0**	0.5	2.9	**7TA4 / 0**	0.3	2.4
7MB4 / 1	0.2	3.1	**7MB9 / 1**	0.2	3.0	**7TA4 / 1**	0.3	2.3
7MB5 / 0	0.3	2.8	**7T70 / 0**	0.3	2.3	**7TA7 / 0**	0.6	2.7
7MB5 / 1	0.2	2.3	**7T70 / 1**	0.2	2.7	**7TA7 / 1**	0.2	2.5
7MB6 / 0	0.5	2.7	**7T8M / 0**	0.4	2.5	**7TB2 / 0**	0.4	3.0
7MB6 / 1	0.2	2.7	**7T8M / 1**	0.3	2.4	**7TB2 / 1**	0.2	2.8
7MB7 /0	0.7	2.8	**7T8R / 0**	0.2	2.4	**7TBT / 0**	0.4	3.0
7MB7 /1	0.1	2.6	**7T8R / 1**	0.2	2.5	**7TBT / 1**	0.2	2.8
7MB8 / 0	0.3	2.5	**7T9Y / 0**	0.3	2.2	**7TC4 / 0**	0.3	2.6
7MB8 / 1	0.2	2.8	**7T9Y / 1**	0.3	2.2	**7TC4 / 1**	0.2	2.4

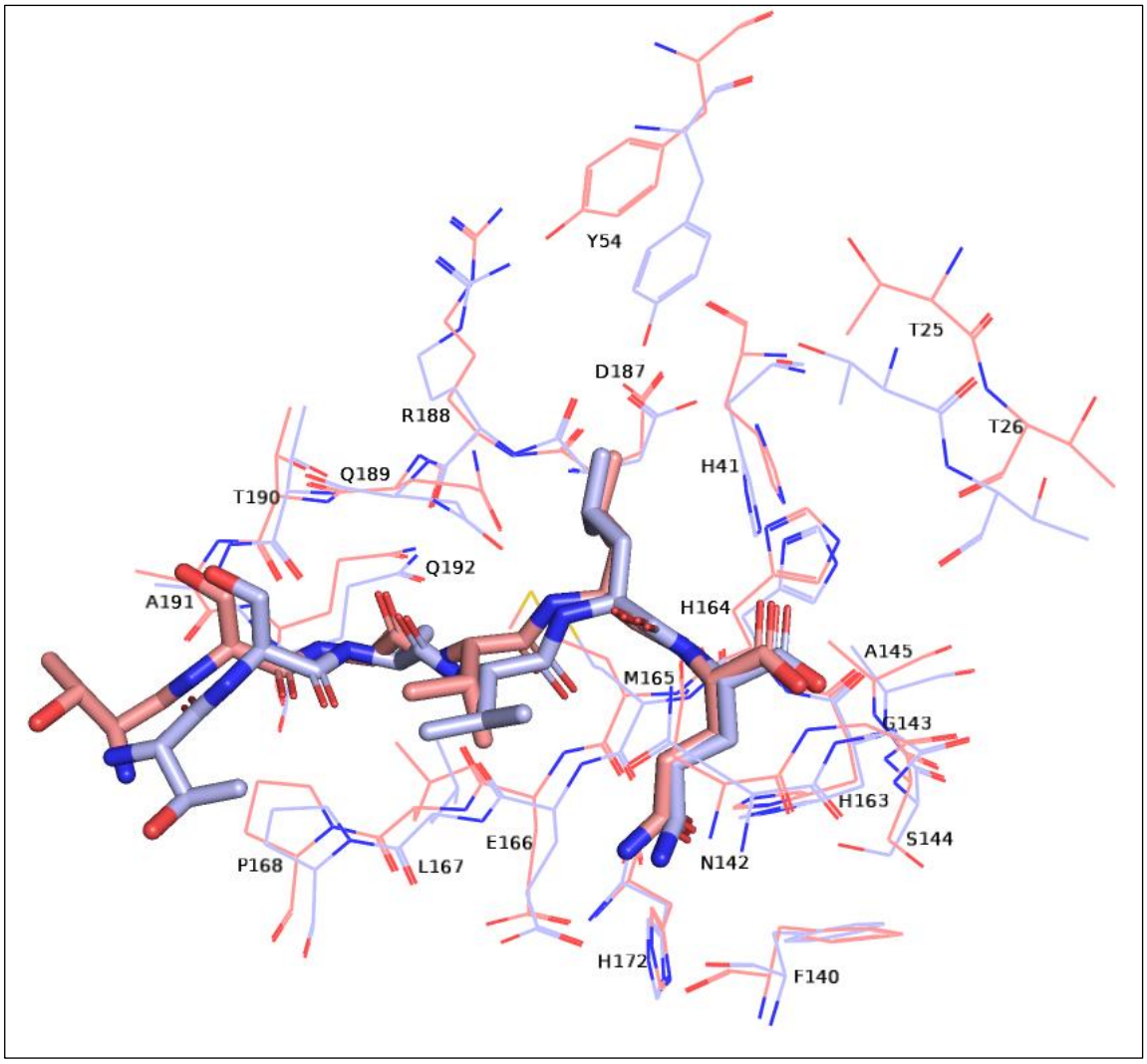

Figure 7.1: The representative structure of the most populated cluster (salmon) retrieved by three replica MD simulations compared with the x-ray structure (light blue) for M^{pro}-nsp4 complex (PDB ID: 7MB4).

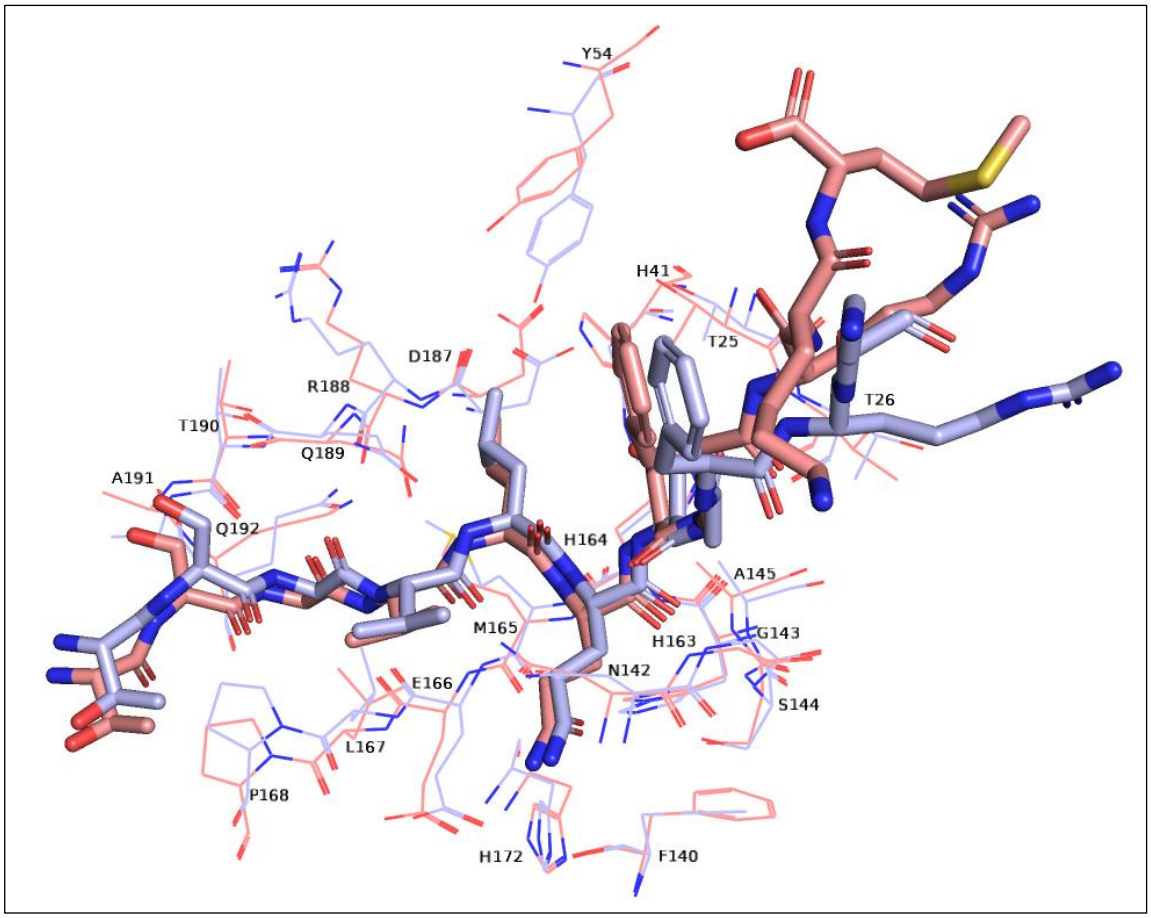

Figure 7.2: The representative structure of the most populated cluster (salmon) retrieved by three replica MD simulations compared with the x-ray structure (light blue) for M^{pro}-nsp4-nsp5 complex (PDB ID: 7T70).

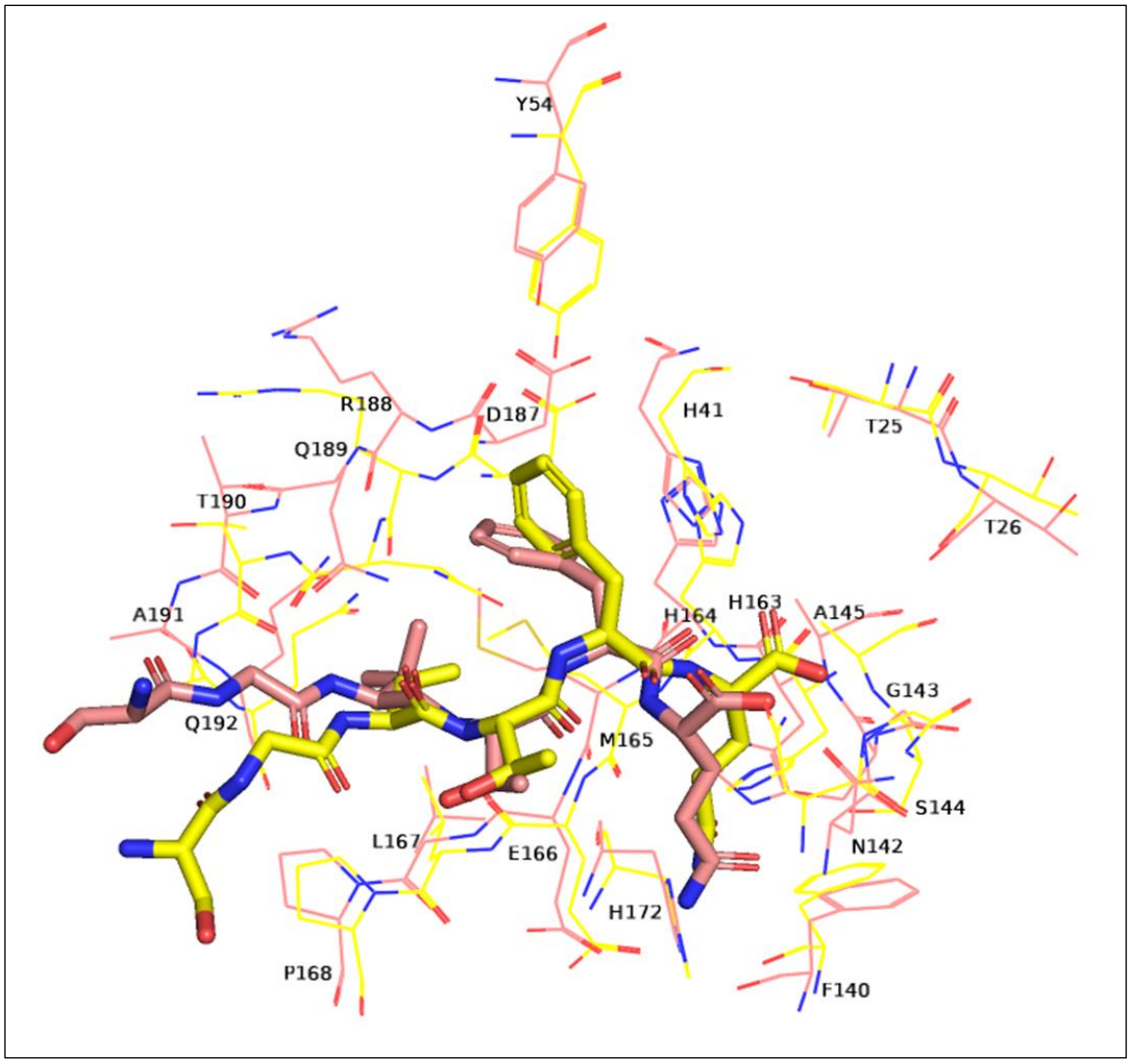

Figure 7.3: The representative structure of the most populated cluster (salmon) retrieved by three replica MD simulations compared with the x-ray structure (yellow) for M^{pro}-nsp5 complex (PDB ID: 7MB5).

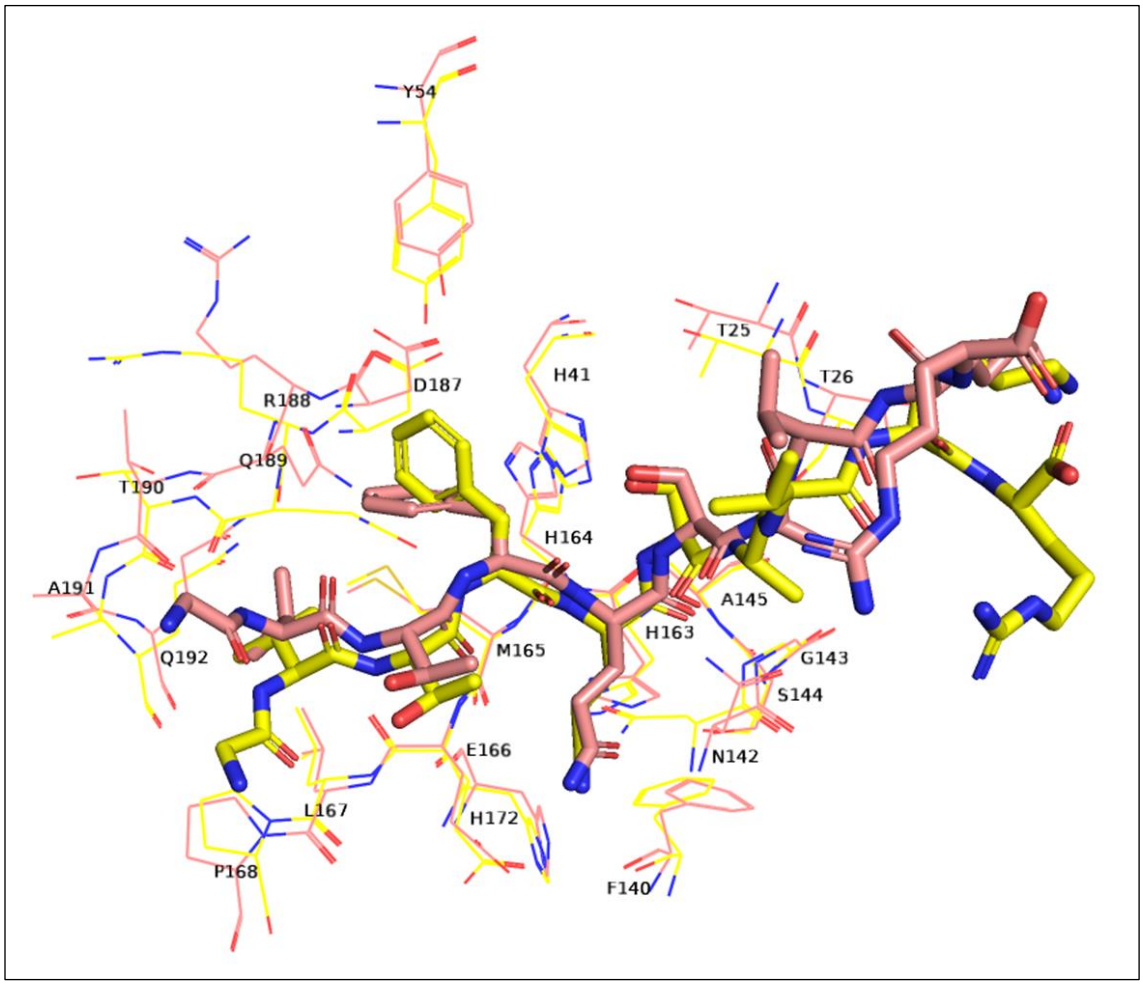

Figure 7.4: The representative structure of the most populated cluster (salmon) retrieved by three replica MD simulations compared with the x-ray structure (yellow) for M^{pro}-nsp5-nsp6 complex (PDB ID: 7T8M).

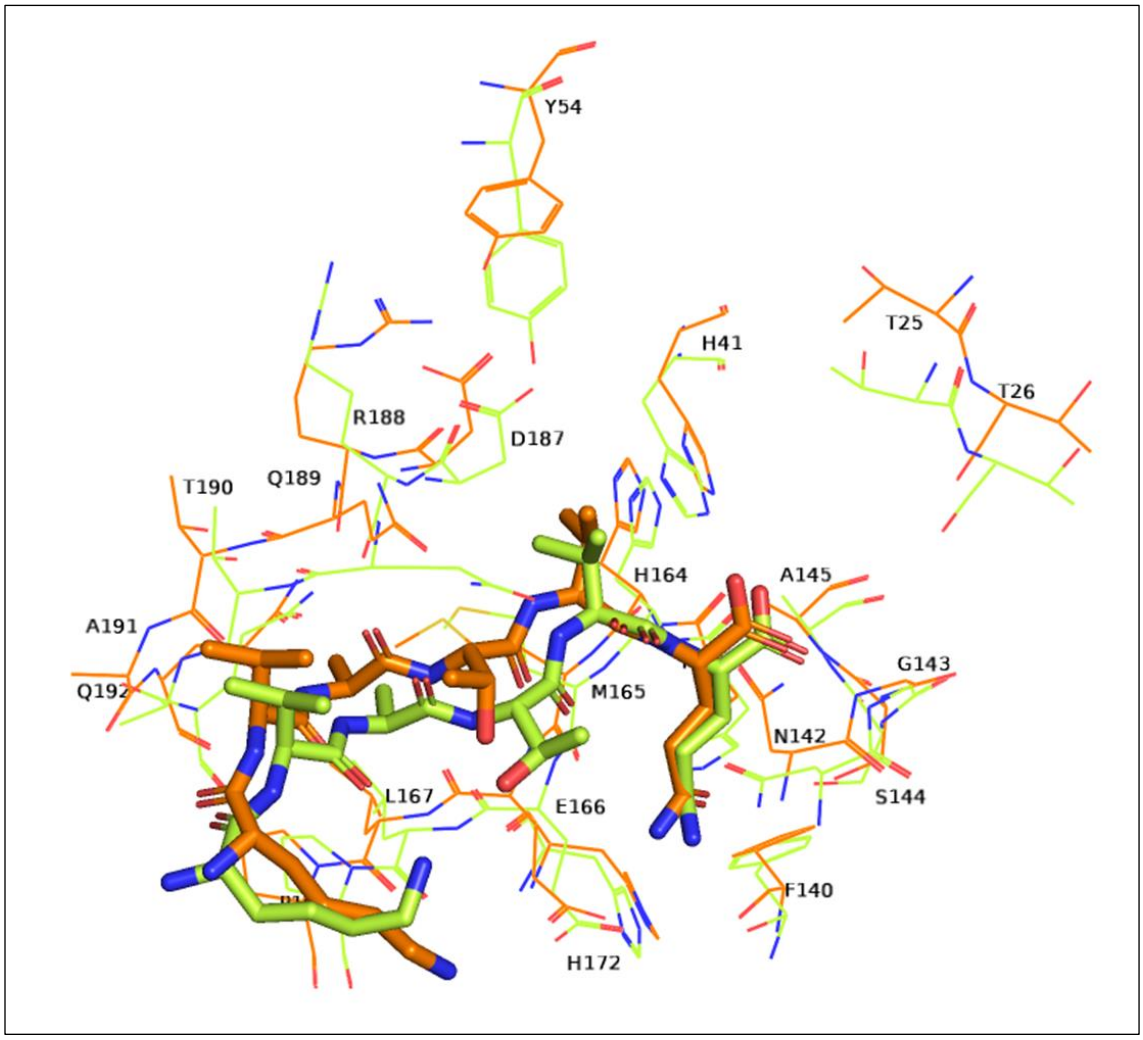

Figure 7.5: The representative structure of the most populated cluster (orange) retrieved by three replica MD simulations compared with the x-ray structure (limon) for M^{pro}-nsp6 complex (PDB ID: 7MB6).

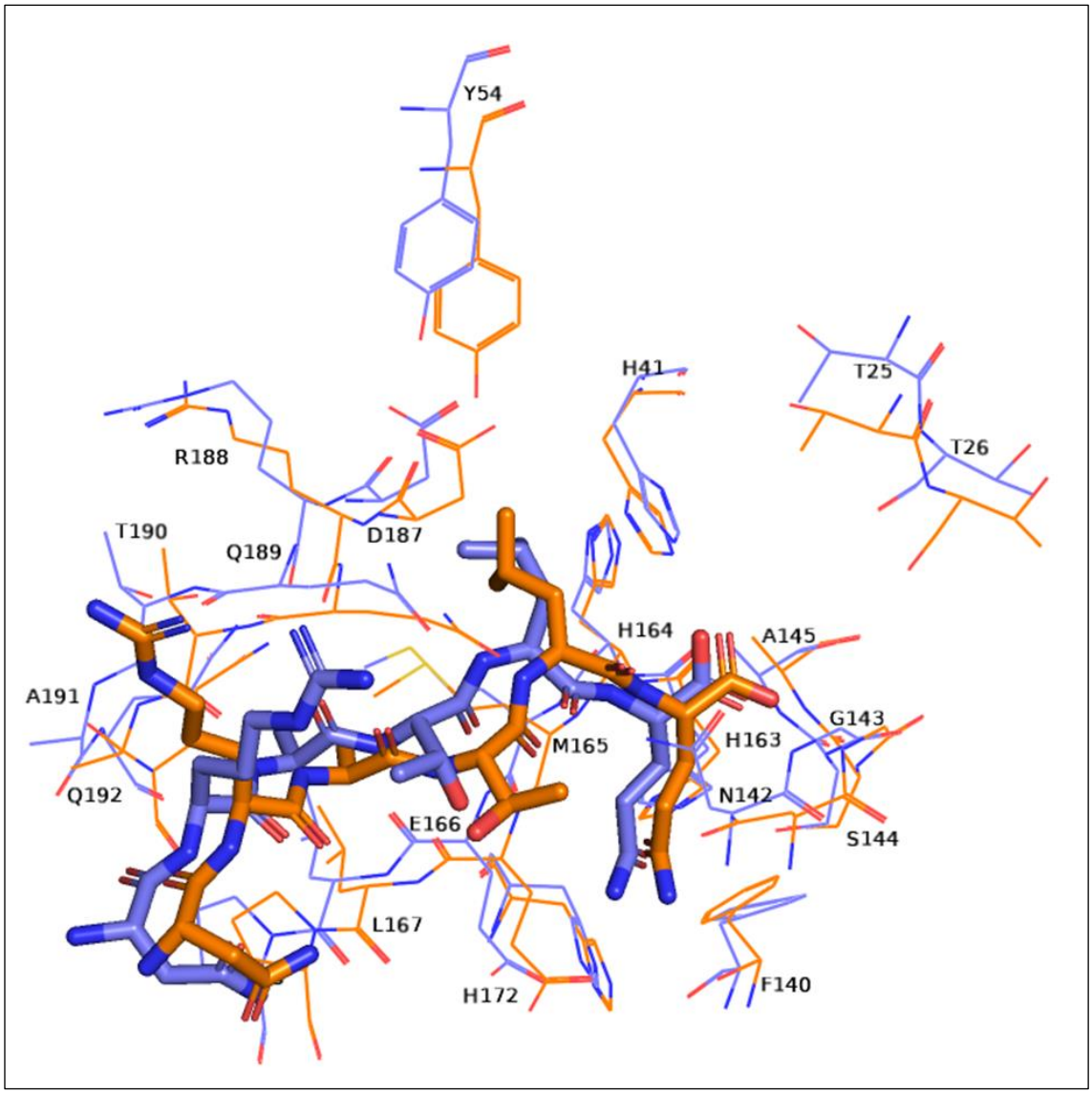

Figure 7.6: The representative structure of the most populated cluster (blue purple) retrieved by three replica MD simulations compared with the x-ray structure (orange) for M^{pro}-nsp7 complex (PDB ID: 7MB7).

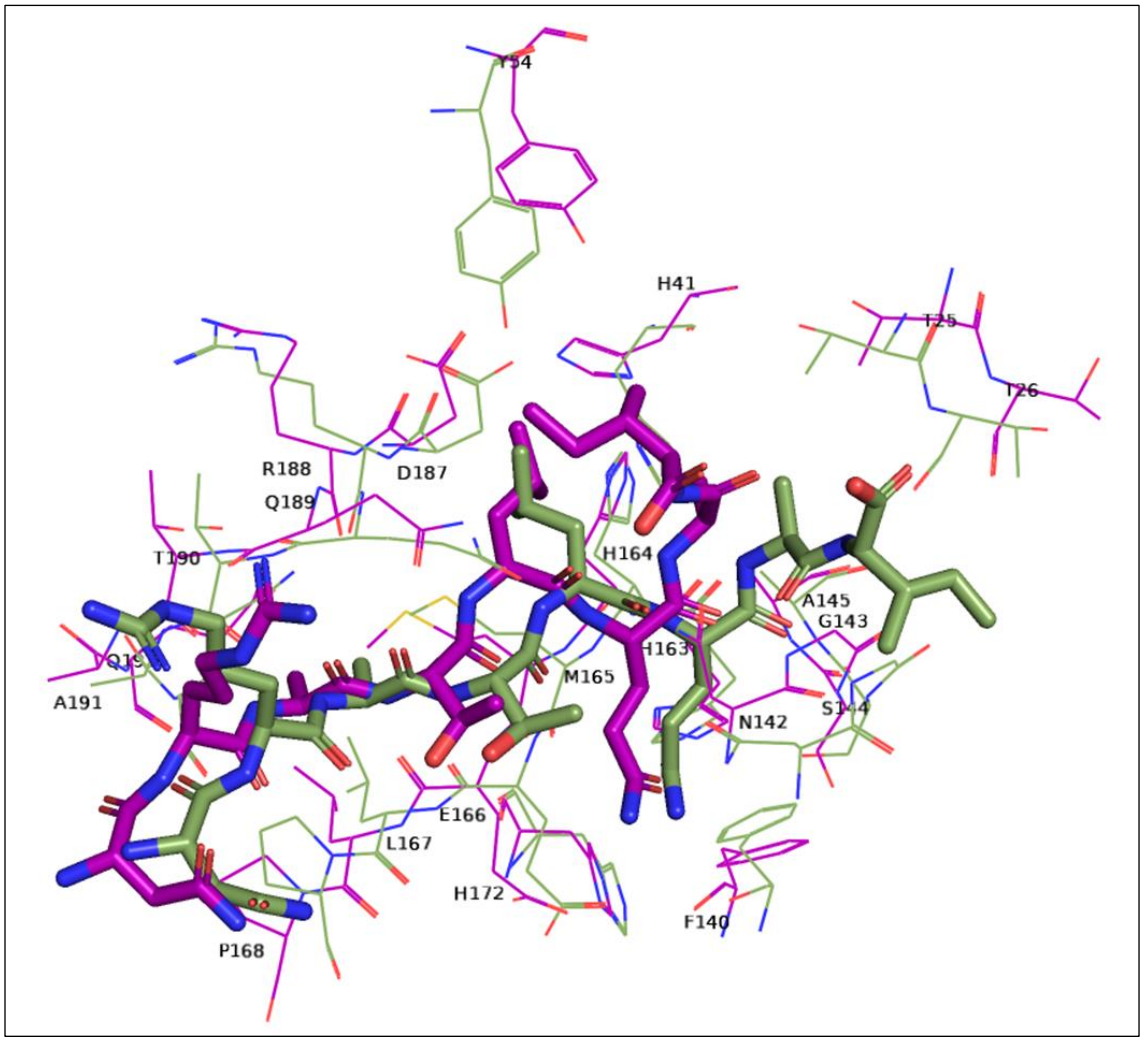

Figure 7.7: The representative structure of the most populated cluster (purple) retrieved by three replica MD simulations compared with the x-ray structure (green) for M^{pro}-nsp7-nsp8 complex (PDB ID: 7T8R).

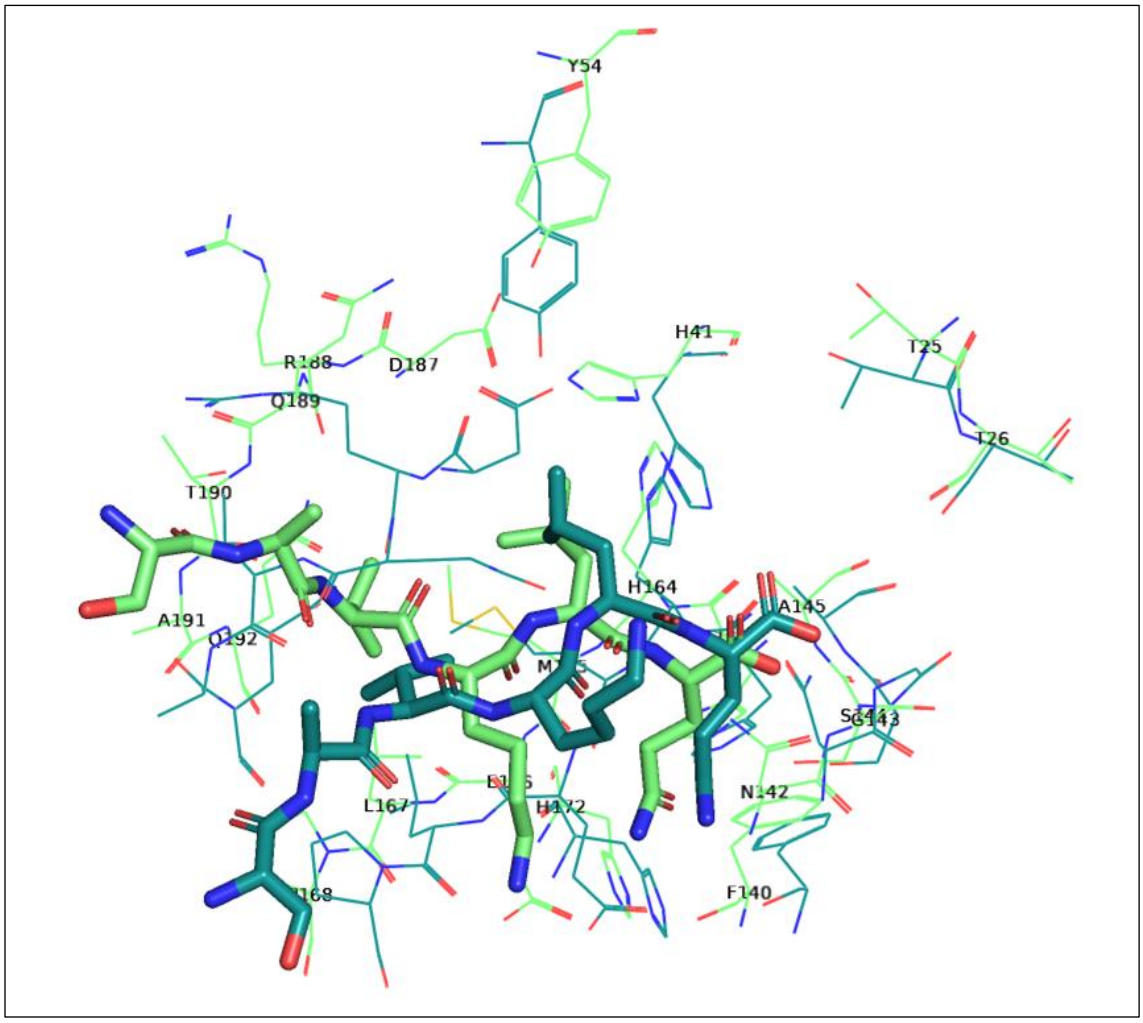

Figure 7.8: The representative structure of the most populated cluster (green) retrieved by three replica MD simulations compared with the x-ray structure (turquoise) for M^{pro}-nsp8 complex (PDB ID: 7MB8).

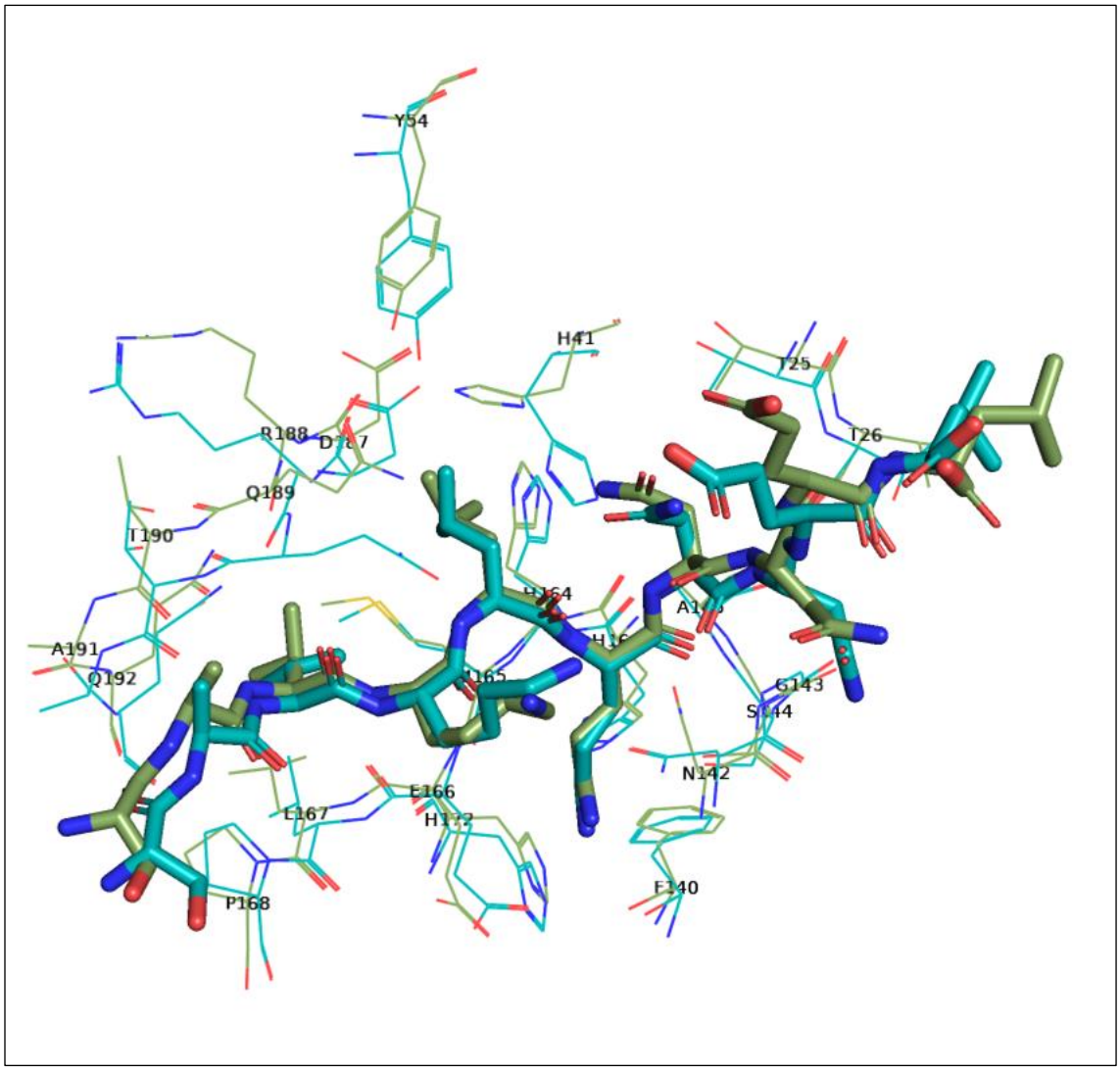

Figure 7.9: The representative structure of the most populated cluster (khaki green) retrieved by three replica MD simulations compared with the x-ray structure (turquoise) for M^{pro}-nsp8-nsp9 complex (PDB ID: 7T9Y).

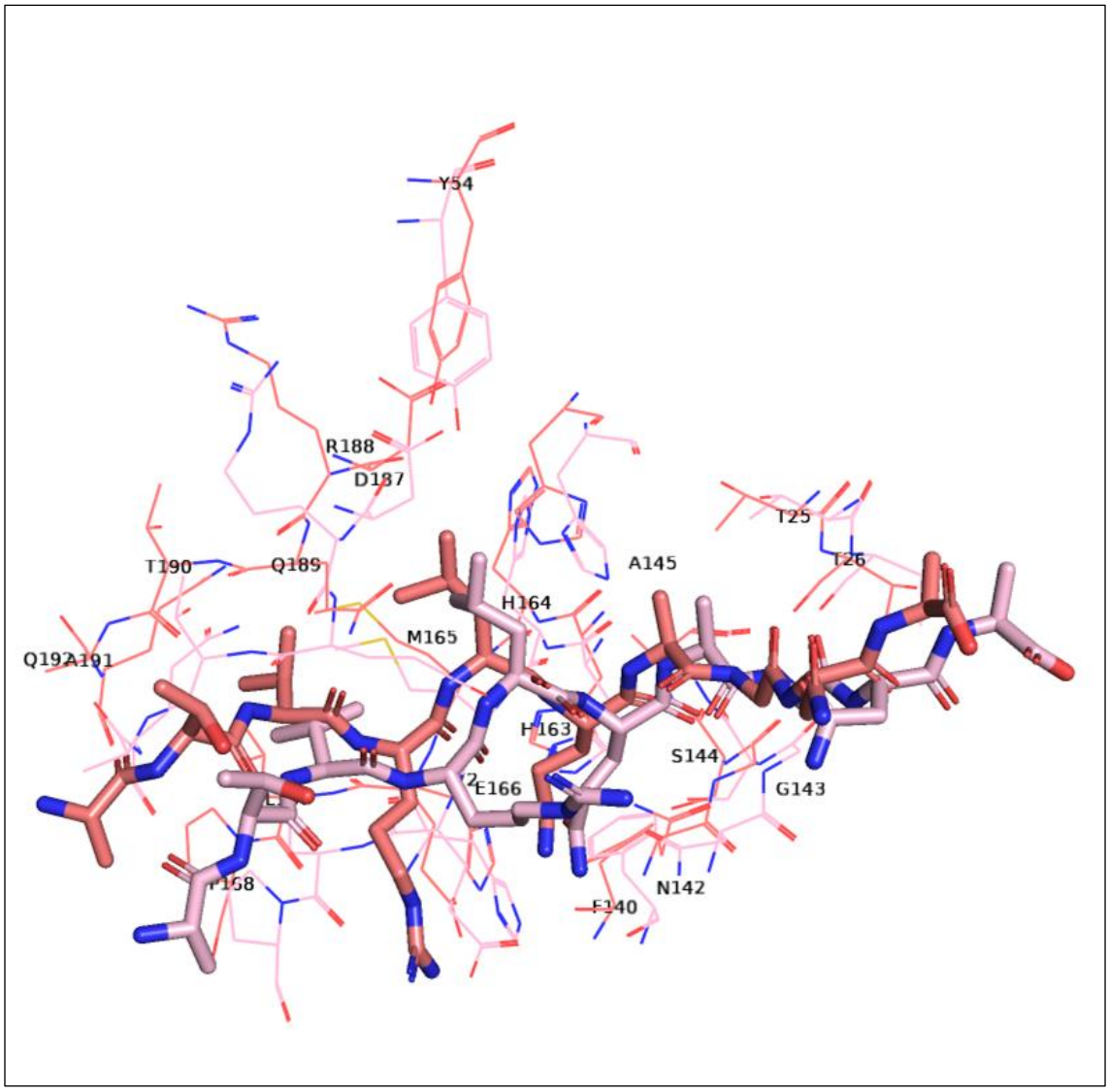

Figure 7.10: The representative structure of the most populated cluster (salmon) retrieved by three replica MD simulations compared with the x-ray structure (pink) for M^{pro}-nsp9-nsp10 complex (PDB ID: 7TA4).

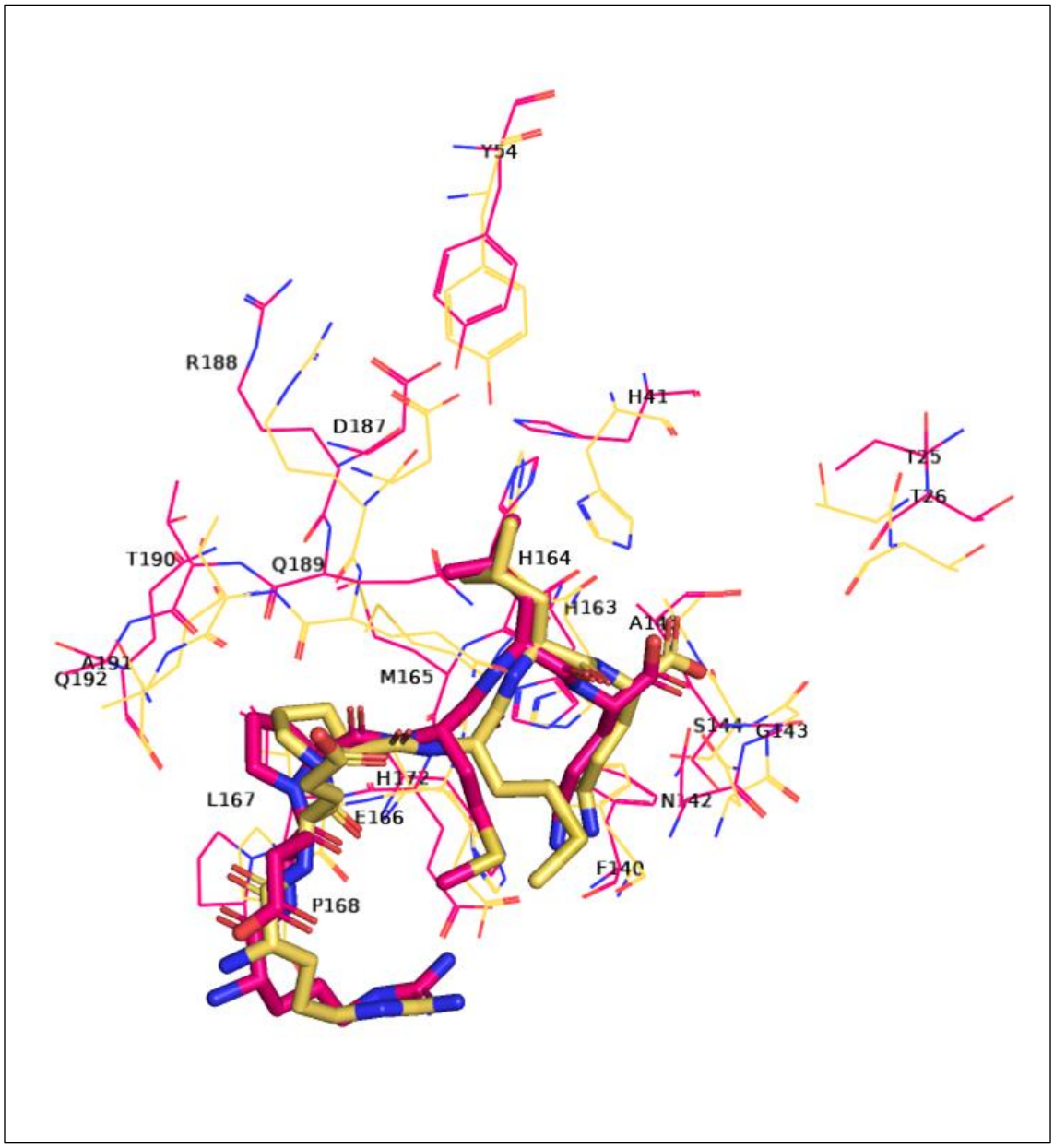

Figure 7.11: The representative structure of the most populated cluster (pink) retrieved by three replica MD simulations compared with the x-ray structure (yellow) for M^{pro}-nsp10 complex (PDB ID: 7MB9).

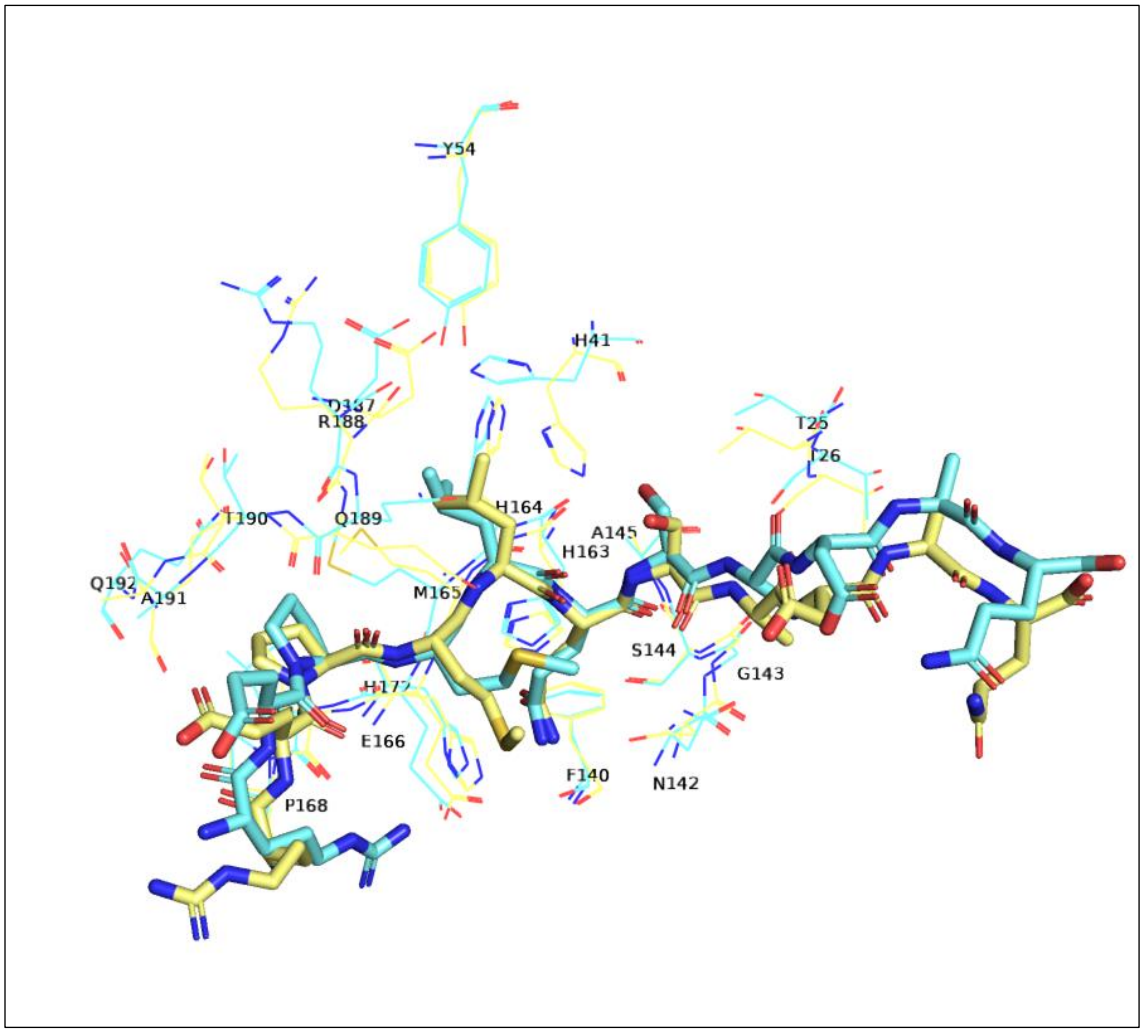

Figure 7.12: The representative structure of the most populated cluster (blue) retrieved by three replica MD simulations compared with the x-ray structure (yellow) for M^{pro}-nsp10-nsp11 complex (PDB ID: 7TA7).

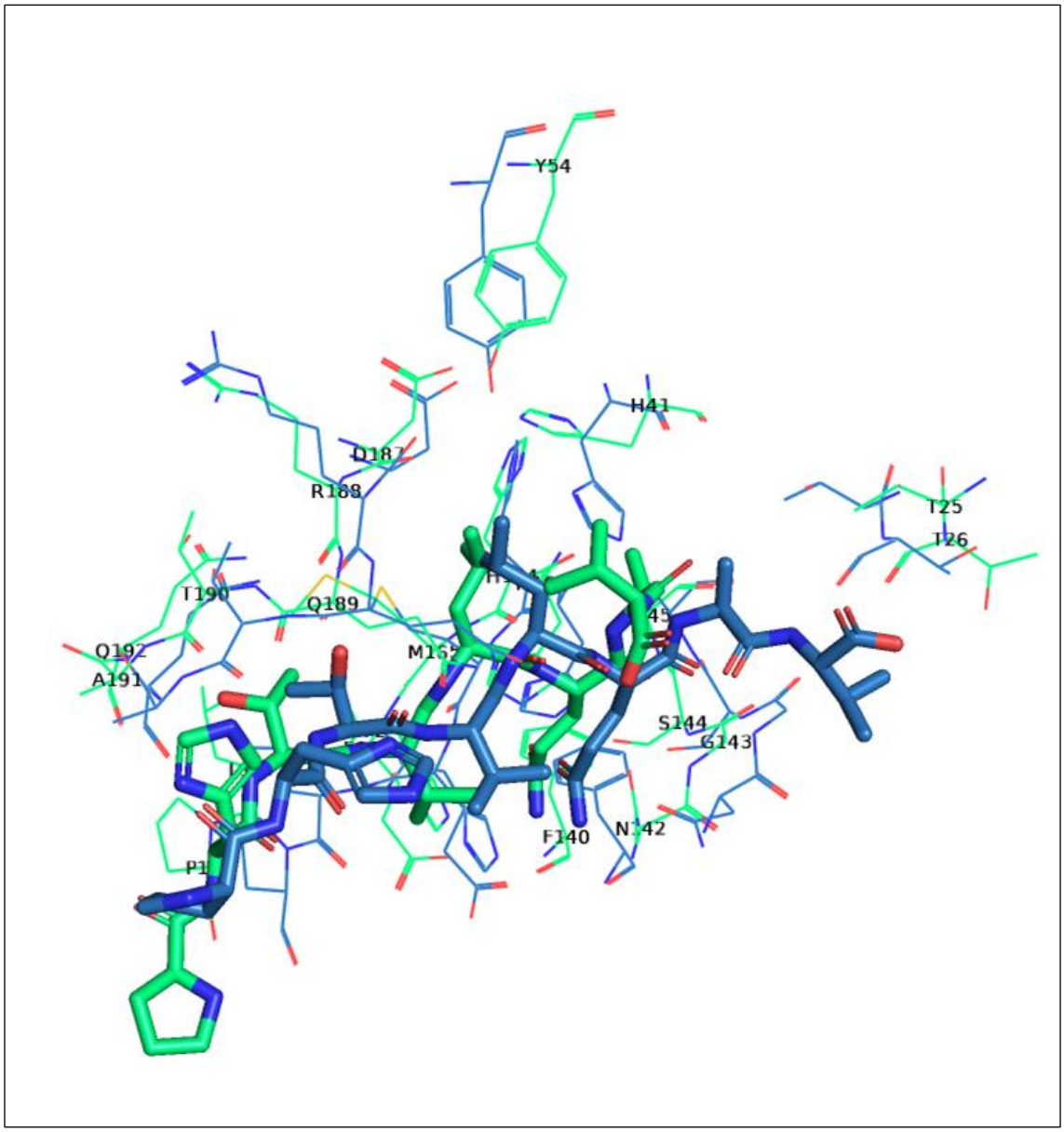

Figure 7.13: The representative structure of the most populated cluster (green) retrieved by three replica MD simulations compared with the x-ray structure (dark blue) for M^{pro}-nsp12-nsp13 complex (PDB ID: 7TB2).

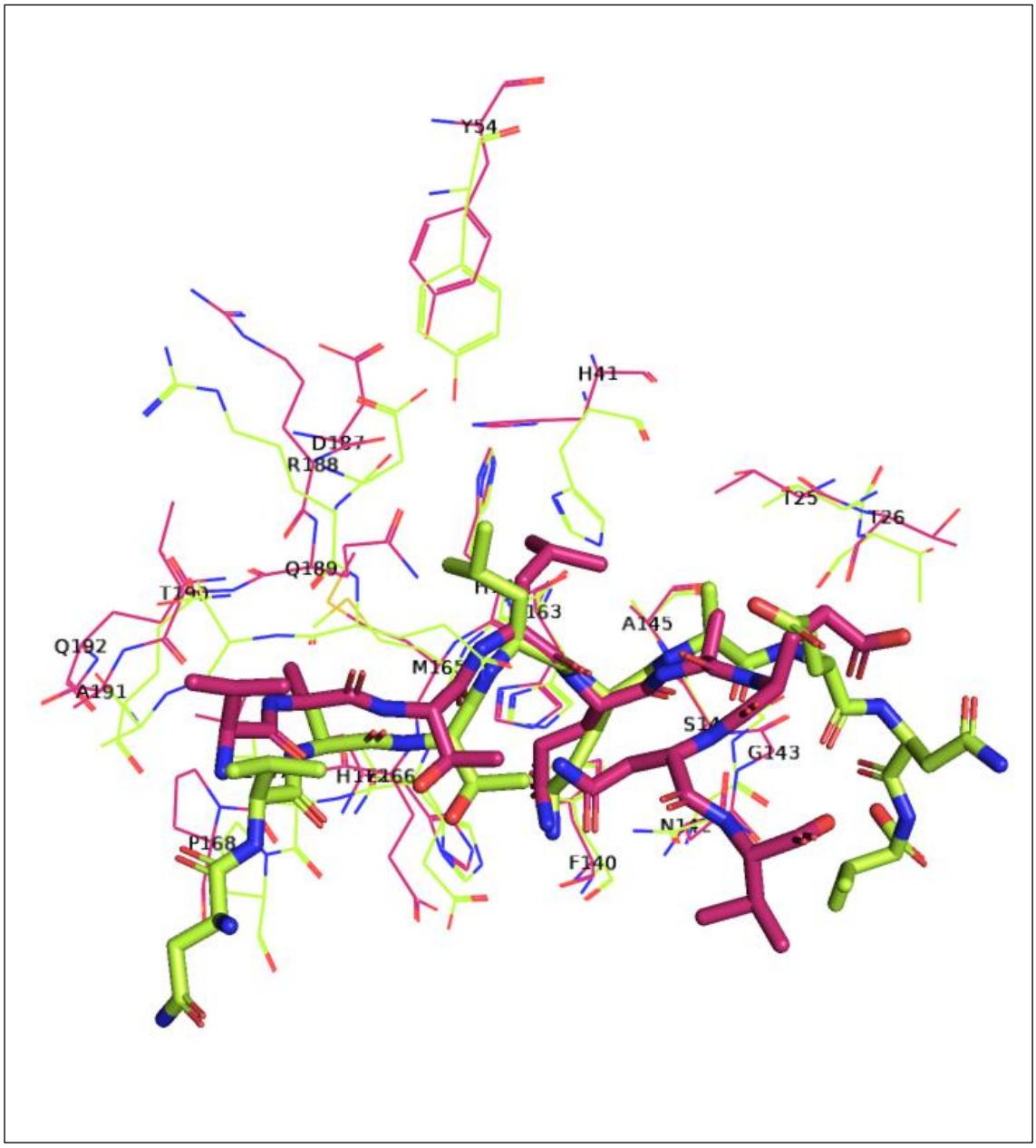

Figure 7.14: The representative structure of the most populated cluster (magenta) retrieved by three replica MD simulations compared with the x-ray structure (lemon) for M^{pro}-nsp13-nsp14 complex (PDB ID: 7TBT).

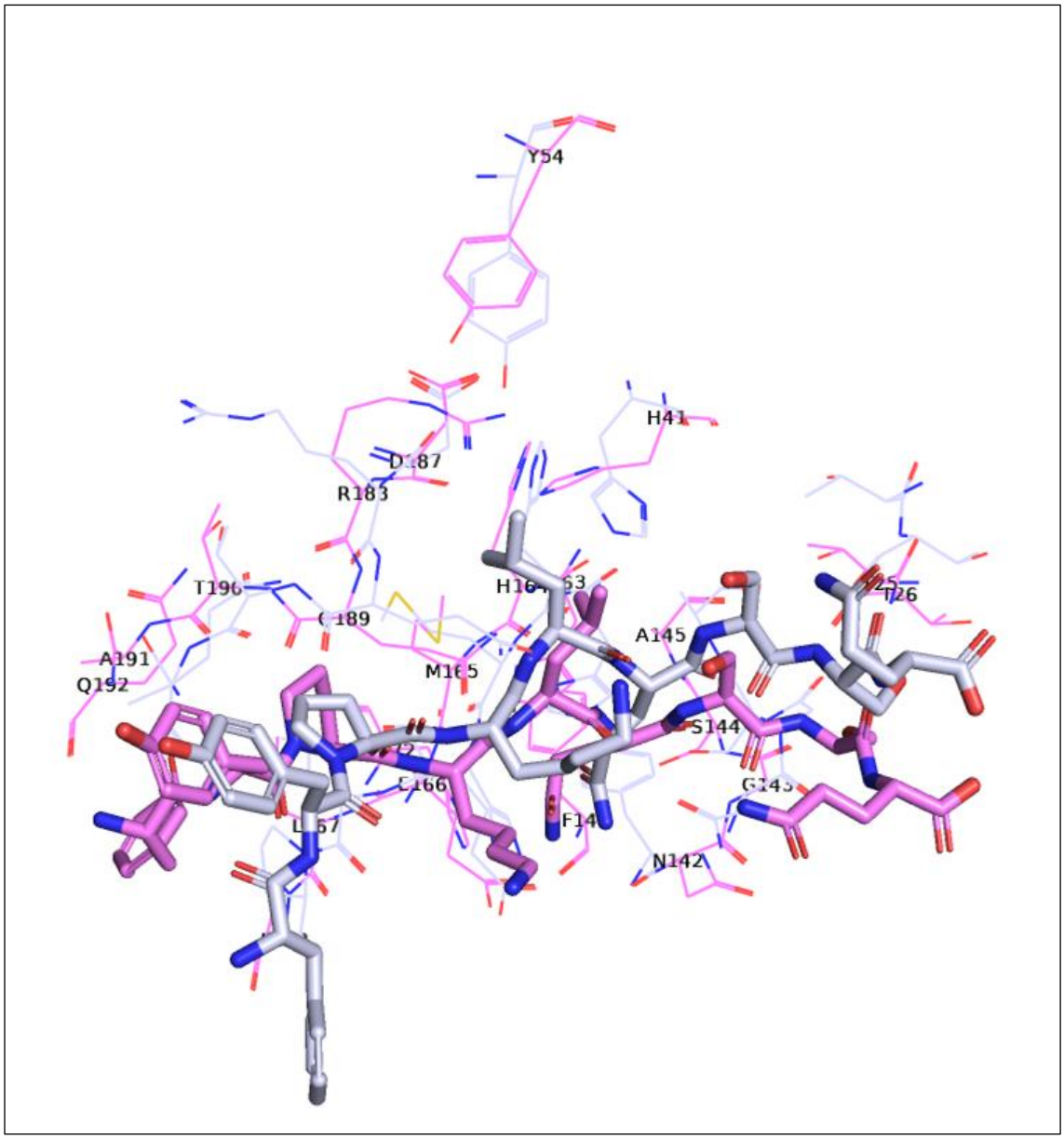

Figure 7.15: The representative structure of the most populated cluster (pink) retrieved by three replica MD simulations compared with the x-ray structure (gray) for M^{pro}-nsp15-nsp16 complex (PDB ID: 7TC4).

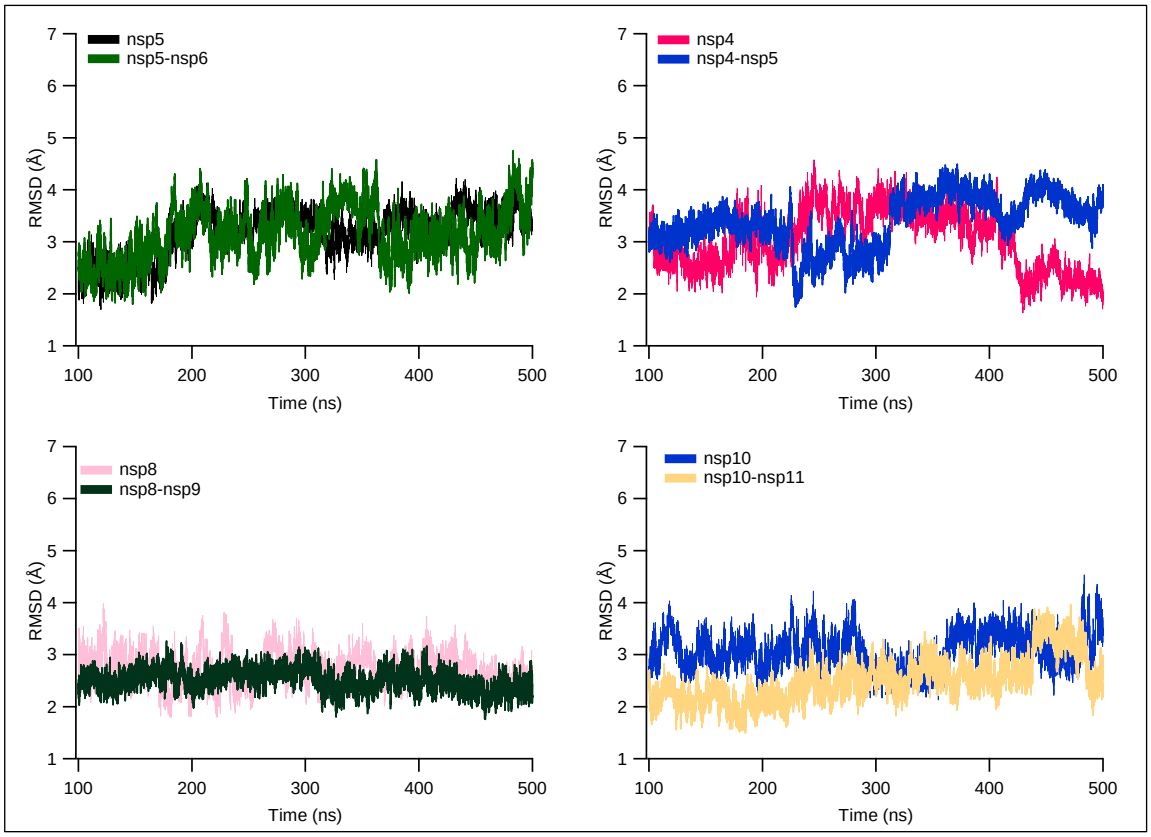

Figure 7.16: RMSD of heavy atoms of peptide fragment in M^{pro}-substrate complexes before and after cleavage. The values are the averages of three replica MD simulations.

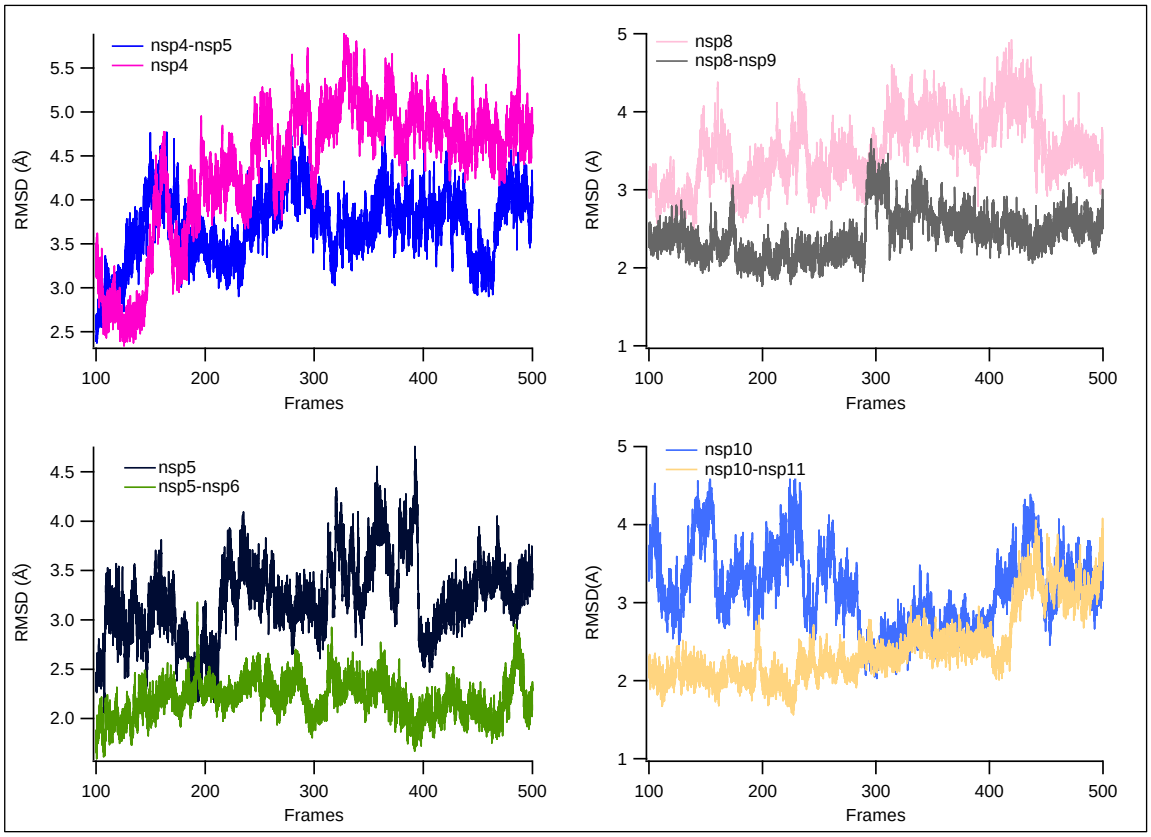

Figure 7.17: RMSD of Cα atoms of peptide fragment in M^{pro}-substrate complexes before and after cleavage. The values are the averages of three replica MD simulations.

REFERENCES

[1] WHO, 'WHO coronavirus disease (COVID-19) dashboard', *WHO*, Jun. 2023. https://covid19.who.int/data (accessed Jun. 26, 2023).

[2] Robson F., Khan K. S., Le T. K., Paris C., Demirbag S., Barfuss P., Rocchi P., Ng W. L., 'Coronavirus RNA Proofreading: Molecular Basis and Therapeutic Targeting', *Molecular Cell*, 79 (5), . Cell Press, 710–727, Sep. 03, 2020. doi: 10.1016/j.molcel.2020.07.027.

[3] Yan L. *et al.*, (2021), 'Cryo-EM Structure of an Extended SARS-CoV-2 Replication and Transcription Complex Reveals an Intermediate State in Cap Synthesis', Cell, 184 (1), 184-193.e10, doi: 10.1016/j.cell.2020.11.016.

[4] Ivanov K. A., Ziebuhr J., (2004), 'Human Coronavirus 229E Nonstructural Protein 13: Characterization of Duplex-Unwinding, Nucleoside Triphosphatase, and RNA 5′-Triphosphatase Activities', Journal of Virology, 78 (14), 7833–7838, doi: 10.1128/jvi.78.14.7833-7838.2004.

[5] Ivanov K. A., Thiel V., Dobbe J. C., van der Meer Y., Snijder E. J., Ziebuhr J., (2004), 'Multiple Enzymatic Activities Associated with Severe Acute Respiratory Syndrome Coronavirus Helicase', Journal of Virology, 78 (11), 5619–5632, doi: 10.1128/jvi.78.11.5619-5632.2004.

[6] Chen Y., Cai H., An Pan J. ', Xiang N., Tien P., Ahola T., Guo D., 'Functional screen reveals SARS coronavirus nonstructural protein nsp14 as a novel cap N7 methyltransferase'. [Online]. Available: www.pnas.org/cgi/content/full/

[7] Bouvet M., Debarnot C., Imbert I., Selisko B., Snijder E. J., Canard B., Decroly E., (2010), 'In vitro reconstitution of sars-coronavirus mRNA cap methylation', PLoS Pathogens, 6 (4), 1–13, doi: 10.1371/journal.ppat.1000863.

[8] Chen Y. *et al.*, (2011), 'Biochemical and structural insights into the mechanisms of sars coronavirus RNA ribose 2′-O-methylation by nsp16/nsp10 protein complex', PLoS Pathogens, 7 (10), , doi: 10.1371/journal.ppat.1002294.

[9] Decroly E., Imbert I., Coutard B., Bouvet M., Selisko B., Alvarez K., Gorbalenya A. E., Snijder E. J., Canard B., (2008), ‘ Coronavirus Nonstructural Protein 16 Is a Cap-0 Binding Enzyme Possessing (Nucleoside-2′ O)-Methyltransferase Activity ’, Journal of Virology, 82 (16), 8071–8084, doi: 10.1128/jvi.00407-08.

[10] Wu A. *et al.*, ‘Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China’, *Cell Host and Microbe*, 27 (3), . Cell Press, 325–328, Mar. 11, 2020. doi: 10.1016/j.chom.2020.02.001.

[11] Almeida M. S., Johnson M. A., Herrmann T., Geralt M., Wüthrich K., (2007), ‘Novel β-Barrel Fold in the Nuclear Magnetic Resonance Structure of the Replicase Nonstructural Protein 1 from the Severe Acute Respiratory Syndrome Coronavirus’, Journal of Virology, 81 (7), 3151–3161, doi: 10.1128/jvi.01939-06.

[12] Semper C., Watanabe N., Savchenko A., (2021), ‘Structural characterization of nonstructural protein 1 from SARS-CoV-2’, iScience, 24 (1), , doi: 10.1016/j.isci.2020.101903.

[13] Clark L. K., Green T. J., Petit C. M., (2021), ‘Structure of Nonstructural Protein 1 from SARS-CoV-2’, Journal of Virology, 95 (4), , doi: 10.1128/jvi.02019-20.

[14] Schubert K. *et al.*, (2020), ‘SARS-CoV-2 Nsp1 binds the ribosomal mRNA channel to inhibit translation’, Nature Structural and Molecular Biology, 27 (10), 959–966, doi: 10.1038/s41594-020-0511-8.

[15] Thoms M. *et al.*, (2020), ‘Structural basis for translational shutdown and immune evasion by the Nsp1 protein of SARS-CoV-2’, Science, 369 (6508), 1249–1256, doi: 10.1126/SCIENCE.ABC8665.

[16] Yuan S. *et al.*, (2020), ‘Nonstructural Protein 1 of SARS-CoV-2 Is a Potent Pathogenicity Factor Redirecting Host Protein Synthesis Machinery toward Viral RNA’, Molecular Cell, 80 (6), 1055-1066.e6, doi: 10.1016/j.molcel.2020.10.034.

[17] Narayanan K., Huang C., Lokugamage K., Kamitani W., Ikegami T., Tseng C.-T. K., Makino S., (2008), ‘Severe Acute Respiratory Syndrome Coronavirus nsp1 Suppresses Host Gene Expression, Including That of Type I Interferon, in Infected Cells’, Journal of Virology, 82 (9), 4471–4479, doi: 10.1128/jvi.02472-07.

[18] Kumar A. *et al.*, (2021), 'SARS-CoV-2 Nonstructural Protein 1 Inhibits the Interferon Response by Causing Depletion of Key Host Signaling Factors', Journal of Virology, 95 (13), , doi: 10.1128/JVI.00266-21.

[19] Vankadari N., Jeyasankar N. N., Lopes W. J., (2020), 'Structure of the SARS-CoV-2 Nsp1/5′-Untranslated Region Complex and Implications for Potential Therapeutic Targets, a Vaccine, and Virulence', Journal of Physical Chemistry Letters, 11 (22), 9659–9668, doi: 10.1021/acs.jpclett.0c02818.

[20] Kamitani W., Narayanan K., Huang C., Lokugamage K., Ikegami T., Ito N., Kubo H., Makino S., 'Severe acute respiratory syndrome coronavirus nsp1 protein suppresses host gene expression by promoting host mRNA degradation'. [Online]. Available: www.pnas.orgcgidoi10.1073pnas.0603144103

[21] Kamitani W., Huang C., Narayanan K., Lokugamage K. G., Makino S., (2009), 'A two-pronged strategy to suppress host protein synthesis by SARS coronavirus Nsp1 protein', Nature Structural and Molecular Biology, 16 (11), 1134–1140, doi: 10.1038/nsmb.1680.

[22] Tanaka T., Kamitani W., DeDiego M. L., Enjuanes L., Matsuura Y., (2012), 'Severe Acute Respiratory Syndrome Coronavirus nsp1 Facilitates Efficient Propagation in Cells through a Specific Translational Shutoff of Host mRNA', Journal of Virology, 86 (20), 11128–11137, doi: 10.1128/jvi.01700-12.

[23] Kilkenny M. L., Veale C. E., Guppy A., Hardwick S. W., Chirgadze D. Y., Rzechorzek N. J., Maman J. D., Pellegrini L., (2022), 'Structural basis for the interaction of SARS-CoV-2 virulence factor nsp1 with DNA polymerase α–primase', Protein Science, 31 (2), 333–344, doi: 10.1002/pro.4220.

[24] Gadlage M. J., Graham R. L., Denison M. R., (2008), 'Murine Coronaviruses Encoding nsp2 at Different Genomic Loci Have Altered Replication, Protein Expression, and Localization', Journal of Virology, 82 (23), 11964–11969, doi: 10.1128/jvi.01126-07.

[25] 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, doi: 10.1128/jvi.00842-09.

[26] Ziebuhr J., Snijder E. J., Gorbalenya A. E., 'Printed in Great Britain Virus-encoded proteinases and proteolytic processing in the Nidovirales'.

[27] Imbert I., Snijder E. J., Dimitrova M., Guillemot J. C., Lécine P., Canard B., (2008), 'The SARS-Coronavirus PLnc domain of nsp3 as a replication/transcription scaffolding protein', Virus Research, 133 (2), 136–148, doi: 10.1016/j.virusres.2007.11.017.

[28] Pfefferle S. *et al.*, (2011), 'The SARS-Coronavirus-host interactome: Identification of cyclophilins as target for pan-Coronavirus inhibitors', PLoS Pathogens, 7 (10), , doi: 10.1371/journal.ppat.1002331.

[29] Van Hemert M. J., Van Den Worm S. H. E., Knoops K., Mommaas A. M., Gorbalenya A. E., Snijder E. J., (2008), 'SARS-coronavirus replication/transcription complexes are membrane-protected and need a host factor for activity in vitro', PLoS Pathogens, 4 (5), , doi: 10.1371/journal.ppat.1000054.

[30] Angelini M. M., Akhlaghpour M., Neuman B. W., Buchmeier M. J., (2013), 'Severe acute respiratory syndrome coronavirus nonstructural proteins 3, 4, and 6 induce double-membrane vesicles', mBio, 4 (4), , doi: 10.1128/mBio.00524-13.

[31] Snijder E. J., van der Meer Y., Zevenhoven-Dobbe J., Onderwater J. J. M., van der Meulen J., Koerten H. K., Mommaas A. M., (2006), 'Ultrastructure and Origin of Membrane Vesicles Associated with the Severe Acute Respiratory Syndrome Coronavirus Replication Complex', Journal of Virology, 80 (12), 5927–5940, doi: 10.1128/jvi.02501-05.

[32] Shin D. *et al.*, (2020), 'Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity', Nature, 587 (7835), 657–662, doi: 10.1038/s41586-020-2601-5.

[33] Rut W. *et al.*, 'Activity profiling and crystal structures of inhibitor-bound SARS-CoV-2 papain-like protease: A framework for anti-COVID-19 drug design'. [Online]. Available: http://advances.sciencemag.org/

[34] Shen Z. *et al.*, (2022), 'Design of SARS-CoV-2 PLpro Inhibitors for COVID-19 Antiviral Therapy Leveraging Binding Cooperativity', Journal of Medicinal Chemistry, 65 (4), 2940–2955, doi: 10.1021/acs.jmedchem.1c01307.

[35] Sargsyan K., Lin C. C., Chen T., Grauffel C., Chen Y. P., Yang W. Z., Yuan H. S., Lim C., (2020), 'Multi-targeting of functional cysteines in multiple conserved SARS-CoV-2 domains by clinically safe Zn-ejectors', Chemical Science, 11 (36), 9904–9909, doi: 10.1039/d0sc02646h.

[36] Tan H., Hu Y., Jadhav P., Tan B., Wang J., 'Progress and Challenges in Targeting the SARS-CoV-2 Papain-like Protease', *Journal of Medicinal Chemistry*, 65 (11), . American Chemical Society, 7561–7580, Jun. 09, 2022. doi: 10.1021/acs.jmedchem.2c00303.

[37] Ratia K., Singh Saikatendu K., Santarsiero B. D., Barretto N., Baker S. C., Stevens R. C., Mesecar A. D., 'Severe acute respiratory syndrome coronavirus papain-like protease: Structure of a viral deubiquitinating enzyme'. [Online]. Available: www.pnas.orgcgidoi10.1073pnas.0510851103

[38] Harrigan J. A., Jacq X., Martin N. M., Jackson S. P., 'Deubiquitylating enzymes and drug discovery: Emerging opportunities', *Nature Reviews Drug Discovery*, 17 (1), . Nature Publishing Group, 57–77, Jan. 01, 2018. doi: 10.1038/nrd.2017.152.

[39] Clementz M. A., Kanjanahaluethai A., O'Brien T. E., Baker S. C., (2008), 'Mutation in murine coronavirus replication protein nsp4 alters assembly of double membrane vesicles', Virology, 375 (1), 118–129, doi: 10.1016/j.virol.2008.01.018.

[40] Oostra M., te Lintelo E. G., Deijs M., Verheije M. H., Rottier P. J. M., de Haan C. A. M., (2007), 'Localization and Membrane Topology of Coronavirus Nonstructural Protein 4: Involvement of the Early Secretory Pathway in Replication', Journal of Virology, 81 (22), 12323–12336, doi: 10.1128/jvi.01506-07.

[41] Xu X. *et al.*, (2009), 'Crystal structure of the C-terminal cytoplasmic domain of non-structural protein 4 from mouse hepatitis virus A59', PLoS ONE, 4 (7), , doi: 10.1371/journal.pone.0006217.

[42] Manolaridis I., Wojdyla J. A., Panjikar S., Snijder E. J., Gorbalenya A. E., Berglind H., Nordlund P., Coutard B., Tucker P. A., (2009), 'Structure of the C-terminal domain of nsp4 from feline coronavirus', Acta Crystallographica Section D: Biological Crystallography, 65 (8), 839–846, doi: 10.1107/S0907444909018253.

[43] Cottam E. M., Maier H. J., Manifava M., Vaux L. C., Chandra-Schoenfelder P., Gerner W., Britton P., Ktistakis N. T., Wileman T., (2011), 'Coronavirus nsp6 proteins generate autophagosomes from the endoplasmic reticulum via an omegasome intermediate', Autophagy, 7 (11), 1335–1347, doi: 10.4161/auto.7.11.16642.

[44] Wolff G. *et al.*, (2020), 'A molecular pore spans the double membrane of the coronavirus replication organelle', Science, 369 (6509), 1395–1398, doi: 10.1126/SCIENCE.ABD3629.

[45] Anand K., Ziebuhr J., Wadhwani P., Mesters J. R., Hilgenfeld R., 'Coronavirus Main Proteinase (3CL pro) Structure: Basis for Design of Anti-SARS Drugs'. [Online]. Available: www.sciencemag.org/cgi/content/full/300/5626/1758/

[46] Yang H. *et al.*, (2003), 'The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor', Proceedings of the National Academy of Sciences of the United States of America, 100 (23), 13190–13195, doi: 10.1073/pnas.1835675100.

[47] Xue X. *et al.*, (2008), 'Structures of Two Coronavirus Main Proteases: Implications for Substrate Binding and Antiviral Drug Design', Journal of Virology, 82 (5), 2515–2527, doi: 10.1128/jvi.02114-07.

[48] Zhao Q., Li S., Xue F., Zou Y., Chen C., Bartlam M., Rao Z., (2008), 'Structure of the Main Protease from a Global Infectious Human Coronavirus, HCoV-HKU1', Journal of Virology, 82 (17), 8647–8655, doi: 10.1128/jvi.00298-08.

[49] Ren Z., Yan L., Zhang N., Guo Y., Yang C., Lou Z., Rao Z., (2013), 'The newly emerged SARS-Like coronavirus HCoV-EMC also has an "Achilles" heel": Current effective inhibitor targeting a 3C-like protease"', Protein and Cell, 4 (4), 248–250, doi: 10.1007/s13238-013-2841-3.

[50] Masters P. S., ‘The Molecular Biology of Coronaviruses’, *Advances in Virus Research*, 65 . 193–292, 2006. doi: 10.1016/S0065-3527(06)66005-3.

[51] Lu Y., Lu X., Denison M. R., ‘Identification and Characterization of a Serine-Like Proteinase of the Murine Coronavirus MHV-A59’. [Online]. Available: https://journals.asm.org/journal/jvi

[52] Snijder E. J., Decroly E., Ziebuhr J., ‘The Nonstructural Proteins Directing Coronavirus RNA Synthesis and Processing’, in *Advances in Virus Research*, Academic Press Inc., 2016, 59–126. doi: 10.1016/bs.aivir.2016.08.008.

[53] Jin Z. *et al.*, (2020), ‘Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors’, Nature, 582 (7811), 289–293, doi: 10.1038/s41586-020-2223-y.

[54] Tay M. Z., Poh C. M., Rénia L., MacAry P. A., Ng L. F. P., ‘The trinity of COVID-19: immunity, inflammation and intervention’, *Nature Reviews Immunology*, 20 (6), . Nature Research, 363–374, Jun. 01, 2020. doi: 10.1038/s41577-020-0311-8.

[55] Jin Z. *et al.*, (2020), ‘Structural basis for the inhibition of SARS-CoV-2 main protease by antineoplastic drug carmofur’, Nature Structural and Molecular Biology, 27 (6), 529–532, doi: 10.1038/s41594-020-0440-6.

[56] Yang H. *et al.*, (2005), ‘Design of wide-spectrum inhibitors targeting coronavirus main proteases’, PLoS Biology, 3 (10), , doi: 10.1371/journal.pbio.0030324.

[57] Ghosh A. K. *et al.*, (2008), ‘Design, synthesis and antiviral efficacy of a series of potent chloropyridyl ester-derived SARS-CoV 3CLpro inhibitors’, Bioorganic and Medicinal Chemistry Letters, 18 (20), 5684–5688, doi: 10.1016/j.bmcl.2008.08.082.

[58] Yang H., Bartlam M., Rao Z., ‘Drug Design Targeting the Main Protease, the Achilles’ Heel of Coronaviruses’.

[59] Bartlam M., Yang H., Rao Z., ‘Structural insights into SARS coronavirus proteins’, *Current Opinion in Structural Biology*, 15 (6), . 664–672, Dec. 2005. doi: 10.1016/j.sbi.2005.10.004.

[60] Lou Z., Sun Y., Rao Z., ‘Current progress in antiviral strategies’, *Trends in Pharmacological Sciences*, 35 (2), . 86–102, Feb. 2014. doi: 10.1016/j.tips.2013.11.006.

[61] Owen D. R. *et al.*, ‘An oral SARS-CoV-2 M pro inhibitor clinical candidate for the treatment of COVID-19’. [Online]. Available: https://www.science.org

[62] Zhang L., Lin D., Sun X., Curth U., Drosten C., Sauerhering L., Becker S., Rox K., Hilgenfeld R., ‘Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved a-ketoamide inhibitors’. [Online]. Available: https://www.science.org

[63] Cooper M. S. *et al.*, (2022), ‘Diastereomeric Resolution Yields Highly Potent Inhibitor of SARS-CoV-2 Main Protease’, Journal of Medicinal Chemistry, 65 (19), 13328–13342, doi: 10.1021/acs.jmedchem.2c01131.

[64] De Clercq E., Li G., ‘Approved antiviral drugs over the past 50 years’, *Clinical Microbiology Reviews*, 29 (3), . American Society for Microbiology, 695–747, Jul. 01, 2016. doi: 10.1128/CMR.00102-15.

[65] Ghosh A. K., Osswald H. L., Prato G., ‘Recent Progress in the Development of HIV-1 Protease Inhibitors for the Treatment of HIV/AIDS’, *Journal of Medicinal Chemistry*, 59 (11), . American Chemical Society, 5172–5208, Jun. 09, 2016. doi: 10.1021/acs.jmedchem.5b01697.

[66] Uraki R. *et al.*, (2022), ‘Characterization and antiviral susceptibility of SARS-CoV-2 Omicron BA.2’, Nature, 607 (7917), 119–127, doi: 10.1038/s41586-022-04856-1.

[67] Greasley S. E. *et al.*, (2022), ‘Structural basis for the in vitro efficacy of nirmatrelvir against SARS-CoV-2 variants’, Journal of Biological Chemistry, 298 (6), , doi: 10.1016/j.jbc.2022.101972.

[68] Malla T. R. *et al.*, (2022), ‘Penicillin Derivatives Inhibit the SARS-CoV-2 Main Protease by Reaction with Its Nucleophilic Cysteine’, Journal of Medicinal Chemistry, 65 (11), 7682–7696, doi: 10.1021/acs.jmedchem.1c02214.

[69] Amporndanai K. *et al.*, (2021), 'Inhibition mechanism of SARS-CoV-2 main protease by ebselen and its derivatives', Nature Communications, 12 (1), , doi: 10.1038/s41467-021-23313-7.

[70] Ghosh A. K. *et al.*, (2021), 'Indole Chloropyridinyl Ester-Derived SARS-CoV-2 3CLpro Inhibitors: Enzyme Inhibition, Antiviral Efficacy, Structure-Activity Relationship, and X-ray Structural Studies', Journal of Medicinal Chemistry, 64 (19), 14702–14714, doi: 10.1021/acs.jmedchem.1c01214.

[71] Dampalla C. S. *et al.*, (2022), 'Structure-Guided Design of Potent Spirocyclic Inhibitors of Severe Acute Respiratory Syndrome Coronavirus-2 3C-like Protease', Journal of Medicinal Chemistry, 65 (11), 7818–7832, doi: 10.1021/acs.jmedchem.2c00224.

[72] Su H. *et al.*, (2021), 'Identification of pyrogallol as a warhead in design of covalent inhibitors for the SARS-CoV-2 3CL protease', Nature Communications, 12 (1), , doi: 10.1038/s41467-021-23751-3.

[73] Unoh Y. *et al.*, (2022), 'Discovery of S-217622, a Noncovalent Oral SARS-CoV-2 3CL Protease Inhibitor Clinical Candidate for Treating COVID-19', Journal of Medicinal Chemistry, , doi: 10.1021/acs.jmedchem.2c00117.

[74] Imbert I., Guillemot J. C., Bourhis J. M., Bussetta C., Coutard B., Egloff M. P., Ferron F., Gorbalenya A. E., Canard B., (2006), 'A second, non-canonical RNA-dependent RNA polymerase in SARS coronavirus', EMBO Journal, 25 (20), 4933–4942, doi: 10.1038/sj.emboj.7601368.

[75] Te Velthuis A. J. W., Van Den Worm S. H. E., Snijder E. J., (2012), 'The SARS-coronavirus nsp7+nsp8 complex is a unique multimeric RNA polymerase capable of both de novo initiation and primer extension', Nucleic Acids Research, 40 (4), 1737–1747, doi: 10.1093/nar/gkr893.

[76] Zhai Y., Sun F., Li X., Pang H., Xu X., Bartlam M., Rao Z., (2005), 'Insights into SARS-CoV transcription and replication from the structure of the nsp7-nsp8 hexadecamer', Nature Structural and Molecular Biology, 12 (11), 980–986, doi: 10.1038/nsmb999.

[77] Li S., Zhao Q., Zhang Y., Zhang Y., Bartlam M., Li X., Rao Z., (2010), 'New nsp8 isoform suggests mechanism for tuning viral RNA synthesis', Protein and Cell, 1 (2), 198–204, doi: 10.1007/s13238-010-0028-8.

[78] Gao Y. *et al.*, (2020), 'Structure of the RNA-dependent RNA polymerase from COVID-19 virus', Science, 368 (6492), 779–782, doi: 10.1126/science.abb7498.

[79] Hillen H. S., Kokic G., Farnung L., Dienemann C., Tegunov D., Cramer P., (2020), 'Structure of replicating SARS-CoV-2 polymerase', Nature, 584 (7819), 154–156, doi: 10.1038/s41586-020-2368-8.

[80] Wang Q. *et al.*, (2020), 'Structural Basis for RNA Replication by the SARS-CoV-2 Polymerase', Cell, 182 (2), 417-428.e13, doi: 10.1016/j.cell.2020.05.034.

[81] Egloff M.-P. *et al.*, 'The severe acute respiratory syndrome-coronavirus replicative protein nsp9 is a single-stranded RNA-binding subunit unique in the RNA virus world'. [Online]. Available: www.pnas.orgcgidoi10.1073pnas.0307877101

[82] Sutton G. *et al.*, (2004), 'The nsp9 Replicase Protein of SARS-Coronavirus, Structure and Functional Insights', Structure, 12 (2), 341–353, doi: 10.1016/j.str.2004.01.016.

[83] Frieman M. *et al.*, (2012), 'Molecular Determinants of Severe Acute Respiratory Syndrome Coronavirus Pathogenesis and Virulence in Young and Aged Mouse Models of Human Disease', Journal of Virology, 86 (2), 884–897, doi: 10.1128/jvi.05957-11.

[84] Littler D. R., Gully B. S., Colson R. N., Rossjohn J., 'Crystal Structure of the SARS-CoV-2 Non-structural Protein 9, Nsp9', doi: 10.1016/j.isci.

[85] Kleiger G., Grothe R., Mallick P., Eisenberg D., (2002), 'GXXXG and AXXXA: Common α-helical interaction motifs in proteins, particularly in extremophiles', Biochemistry, 41 (19), 5990–5997, doi: 10.1021/bi0200763.

[86] Miknis Z. J., Donaldson E. F., Umland T. C., Rimmer R. A., Baric R. S., Schultz L. W., (2009), 'Severe Acute Respiratory Syndrome Coronavirus nsp9 Dimerization Is Essential for Efficient Viral Growth', Journal of Virology, 83 (7), 3007–3018, doi: 10.1128/jvi.01505-08.

[87] Yan L. *et al.*, (2021), 'Coupling of N7-methyltransferase and 3′-5′ exoribonuclease with SARS-CoV-2 polymerase reveals mechanisms for capping and proofreading', Cell, 184 (13), 3474-3485.e11, doi: 10.1016/j.cell.2021.05.033.

[88] Su D. *et al.*, (2006), 'Dodecamer Structure of Severe Acute Respiratory Syndrome Coronavirus Nonstructural Protein nsp10', Journal of Virology, 80 (16), 7902–7908, doi: 10.1128/jvi.00483-06.

[89] Decroly E. *et al.*, (2011), 'Crystal structure and functional analysis of the SARS-coronavirus RNA cap 2′-o-methyltransferase nsp10/nsp16 complex', PLoS Pathogens, 7 (5), , doi: 10.1371/journal.ppat.1002059.

[90] Ma Y. *et al.*, (2015), 'Structural basis and functional analysis of the SARS coronavirus nsp14-nsp10 complex', Proceedings of the National Academy of Sciences of the United States of America, 112 (30), 9436–9441, doi: 10.1073/pnas.1508686112.

[91] Kirchdoerfer R. N., Ward A. B., (2019), 'Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors', Nature Communications, 10 (1), , doi: 10.1038/s41467-019-10280-3.

[92] Ziebuhr J., 'The Coronavirus Replicase', Springer-Verlag.

[93] Mcdonald S. M., (2013), 'RNA synthetic mechanisms employed by diverse families of RNA viruses', Wiley Interdisciplinary Reviews: RNA, 4 (4), 351–367, doi: 10.1002/wrna.1164.

[94] Lehmann K. C. *et al.*, (2015), 'Discovery of an essential nucleotidylating activity associated with a newly delineated conserved domain in the RNA polymerase-containing protein of all nidoviruses', Nucleic Acids Research, 43 (17), 8416–8434, doi: 10.1093/nar/gkv838.

[95] Walker A. P., Fan H., Keown J. R., Grimes J. M., Fodor E., William S., 'Title: Identification of guanylyltransferase activity in the SARS-CoV-2 RNA polymerase Running title: Identification of the SARS-CoV-2 guanylyltransferase', doi: 10.1101/2021.03.17.435913.

[96] Jia Z. *et al.*, (2019), 'Delicate structural coordination of the Severe Acute Respiratory Syndrome coronavirus Nsp13 upon ATP hydrolysis', Nucleic Acids Research, 47 (12), 6538–6550, doi: 10.1093/nar/gkz409.

[97] Chen J. *et al.*, (2020), 'Structural Basis for Helicase-Polymerase Coupling in the SARS-CoV-2 Replication-Transcription Complex', Cell, 182 (6), 1560-1573.e13, doi: 10.1016/j.cell.2020.07.033.

[98] Yan L. *et al.*, (2020), 'Architecture of a SARS-CoV-2 mini replication and transcription complex', Nature Communications, 11 (1), , doi: 10.1038/s41467-020-19770-1.

[99] Seybert A., Van Dinten L. C., Snijder E. J., Ziebuhr J., 'Biochemical Characterization of the Equine Arteritis Virus Helicase Suggests a Close Functional Relationship between Arterivirus and Coronavirus Helicases'.

[100] Seybert A., Hegyi A., Siddell S. G., Ziebuhr J., (2000), 'The human coronavirus 229E superfamily 1 helicase has RNA and DNA duplex-unwinding activities with 5'-to-3' polarity', RNA, 6 (7), 1056–1068, doi: 10.1017/S1355838200000728.

[101] Ferron F. *et al.*, (2017), 'Structural and molecular basis of mismatch correction and ribavirin excision from coronavirus RNA', Proceedings of the National Academy of Sciences of the United States of America, 115 (2), E162–E171, doi: 10.1073/pnas.1718806115.

[102] Case J. B., Ashbrook A. W., Dermody T. S., Denison M. R., (2016), ' Mutagenesis of S -Adenosyl- l -Methionine-Binding Residues in Coronavirus nsp14 N7-Methyltransferase Demonstrates Differing Requirements for Genome Translation and Resistance to Innate Immunity ', Journal of Virology, 90 (16), 7248–7256, doi: 10.1128/jvi.00542-16.

[103] Munster V. J. *et al.*, (2020), 'Respiratory disease in rhesus macaques inoculated with SARS-CoV-2', Nature, 585 (7824), 268–272, doi: 10.1038/s41586-020-2324-7.

[104] Ogando N. S., Ferron F., Decroly E., Canard B., Posthuma C. C., Snijder E. J., 'The Curious Case of the Nidovirus Exoribonuclease: Its Role in RNA Synthesis and

Replication Fidelity', *Frontiers in Microbiology*, 10 . Frontiers Media S.A., Aug. 07, 2019. doi: 10.3389/fmicb.2019.01813.

[105] Bouvet M., Imbert I., Subissi L., Gluais L., Canard B., Decroly E., (2012), 'RNA 3′-end mismatch excision by the severe acute respiratory syndrome coronavirus nonstructural protein nsp10/nsp14 exoribonuclease complex', Proceedings of the National Academy of Sciences of the United States of America, 109 (24), 9372–9377, doi: 10.1073/pnas.1201130109.

[106] Pillon M. C. *et al.*, (2021), 'Cryo-EM structures of the SARS-CoV-2 endoribonuclease Nsp15 reveal insight into nuclease specificity and dynamics', Nature Communications, 12 (1), , doi: 10.1038/s41467-020-20608-z.

[107] Xu X. *et al.*, (2006), 'New Antiviral Target Revealed by the Hexameric Structure of Mouse Hepatitis Virus Nonstructural Protein nsp15', Journal of Virology, 80 (16), 7909–7917, doi: 10.1128/jvi.00525-06.

[108] Zhang L., Li L., Yan L., Ming Z., Jia Z., Lou Z., Rao Z., (2018), 'Structural and Biochemical Characterization of Endoribonuclease Nsp15 Encoded by Middle East Respiratory Syndrome Coronavirus', Journal of Virology, 92 (22), , doi: 10.1128/jvi.00893-18.

[109] Kang H., Bhardwaj K., Li Y., Palaninathan S., Sacchettini J., Guarino L., Leibowitz J. L., Kao C. C., (2007), 'Biochemical and Genetic Analyses of Murine Hepatitis Virus Nsp15 Endoribonuclease', Journal of Virology, 81 (24), 13587–13597, doi: 10.1128/jvi.00547-07.

[110] Bhardwaj K., Guarino L., Kao C. C., (2004), 'The Severe Acute Respiratory Syndrome Coronavirus Nsp15 Protein Is an Endoribonuclease That Prefers Manganese as a Cofactor', Journal of Virology, 78 (22), 12218–12224, doi: 10.1128/jvi.78.22.12218-12224.2004.

[111] Viswanathan T. *et al.*, (2020), 'Structural basis of RNA cap modification by SARS-CoV-2', Nature Communications, 11 (1), , doi: 10.1038/s41467-020-17496-8.

[112] Krafcikova P., Silhan J., Nencka R., Boura E., (2020), 'Structural analysis of the SARS-CoV-2 methyltransferase complex involved in RNA cap creation bound to sinefungin', Nature Communications, 11 (1), , doi: 10.1038/s41467-020-17495-9.

[113] Lai M. M. C., Stohlman S. A., 'Comparative Analysis of RNA Genomes of Mouse Hepatitis Viruses'. [Online]. Available: https://journals.asm.org/journal/jvi

[114] Peng Y. H., Lin C. H., Lin C. N., Lo C. Y., Tsai T. L., Wu H. Y., (2016), 'Characterization of the role of hexamer AGUAAA and poly(A) tail in coronavirus polyadenylation', PLoS ONE, 11 (10), , doi: 10.1371/journal.pone.0165077.

[115] Knoops K., Kikkert M., Van Den Worm S. H. E., Zevenhoven-Dobbe J. C., Van Der Meer Y., Koster A. J., Mommaas A. M., Snijder E. J., (2008), 'SARS-coronavirus replication is supported by a reticulovesicular network of modified endoplasmic reticulum', PLoS Biology, 6 (9), 1957–1974, doi: 10.1371/journal.pbio.0060226.

[116] Snijder E. J., Limpens R. W. A. L., de Wilde A. H., de Jong A. W. M., Zevenhoven-Dobbe J. C., Maier H. J., Faas F. F. G. A., Koster A. J., Bárcena M., (2020), 'A unifying structural and functional model of the coronavirus replication organelle: Tracking down RNA synthesis', PLoS Biology, 18 (6), , doi: 10.1371/journal.pbio.3000715.

[117] Ulasli M., Verheije M. H., de Haan C. A. M., Reggiori F., (2010), 'Qualitative and quantitative ultrastructural analysis of the membrane rearrangements induced by coronavirus', Cellular Microbiology, 12 (6), 844–861, doi: 10.1111/j.1462-5822.2010.01437.x.

[118] Kandeel M., Ibrahim A., Fayez M., Al-Nazawi M., (2020), 'From SARS and MERS CoVs to SARS-CoV-2: Moving toward more biased codon usage in viral structural and nonstructural genes', Journal of Medical Virology, 92 (6), 660–666, doi: 10.1002/jmv.25754.

[119] Neuman B. W. *et al.*, (2011), 'A structural analysis of M protein in coronavirus assembly and morphology', Journal of Structural Biology, 174 (1), 11–22, doi: 10.1016/j.jsb.2010.11.021.

[120] McBride R., van Zyl M., Fielding B. C., 'The coronavirus nucleocapsid is a multifunctional protein', *Viruses*, 6 (8), . MDPI AG, 2991–3018, Aug. 07, 2014. doi: 10.3390/v6082991.

[121] Schoeman D., Fielding B. C., 'Coronavirus envelope protein: Current knowledge', *Virology Journal*, 16 (1), . BioMed Central Ltd., May 27, 2019. doi: 10.1186/s12985-019-1182-0.

[122] Song H. C. *et al.*, (2004), 'Synthesis and Characterization of a Native, Oligomeric Form of Recombinant Severe Acute Respiratory Syndrome Coronavirus Spike Glycoprotein', Journal of Virology, 78 (19), 10328–10335, doi: 10.1128/jvi.78.19.10328-10335.2004.

[123] Belouzard S., Millet J. K., Licitra B. N., Whittaker G. R., 'Mechanisms of coronavirus cell entry mediated by the viral spike protein.', *Viruses*, 4 (6), . 1011–1033, 2012. doi: 10.3390/v4061011.

[124] Reis C. A., Tauber R., Blanchard V., (2092), 'Glycosylation is a key in SARS-CoV-2 infection', doi: 10.1007/s00109-021-02092-0/Published.

[125] Walls A. C., Park Y. J., Tortorici M. A., Wall A., McGuire A. T., Veesler D., (2020), 'Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein', Cell, 181 (2), 281-292.e6, doi: 10.1016/j.cell.2020.02.058.

[126] Majumder J., Minko T., 'Recent Developments on Therapeutic and Diagnostic Approaches for COVID-19', *AAPS Journal*, 23 (1), . Springer Science and Business Media Deutschland GmbH, Jan. 01, 2021. doi: 10.1208/s12248-020-00532-2.

[127] Wei C. *et al.*, (2020), 'HDL-scavenger receptor B type 1 facilitates SARS-CoV-2 entry', Nature Metabolism, 2 (12), 1391–1400, doi: 10.1038/s42255-020-00324-0.

[128] Wruck W., Adjaye J., (2020), 'SARS-CoV-2 receptor ACE2 is co-expressed with genes related to transmembrane serine proteases, viral entry, immunity and cellular stress', Scientific Reports, 10 (1), , doi: 10.1038/s41598-020-78402-2.

[129] Zhai X. *et al.*, (2020), 'Comparison of Severe Acute Respiratory Syndrome Coronavirus 2 Spike Protein Binding to ACE2 Receptors from Human, Pets, Farm

Animals, and Putative Intermediate Hosts', Journal of Virology, 94 (15), , doi: 10.1128/jvi.00831-20.

[130] Shang J., Wan Y., Luo C., Ye G., Geng Q., Auerbach A., Li F., 'Cell entry mechanisms of SARS-CoV-2', doi: 10.1073/pnas.2003138117/-/DCSupplemental.

[131] Ziebuhr J., Siddell S. G., 'Processing of the Human Coronavirus 229E Replicase Polyproteins by the Virus-Encoded 3C-Like Proteinase: Identification of Proteolytic Products and Cleavage Sites Common to pp1a and pp1ab'.

[132] Ziebuhr J., Herold J., Siddell S. G., 'Characterization of a Human Coronavirus (Strain 229E) 3C-Like Proteinase Activity'.

[133] Harcourt B. H., Jukneliene D., Kanjanahaluethai A., Bechill J., Severson K. M., Smith C. M., Rota P. A., Baker S. C., (2004), 'Identification of Severe Acute Respiratory Syndrome Coronavirus Replicase Products and Characterization of Papain-Like Protease Activity', Journal of Virology, 78 (24), 13600–13612, doi: 10.1128/jvi.78.24.13600-13612.2004.

[134] Wang M. Y., Zhao R., Gao L. J., Gao X. F., Wang D. P., Cao J. M., 'SARS-CoV-2: Structure, Biology, and Structure-Based Therapeutics Development', *Frontiers in Cellular and Infection Microbiology*, 10 . Frontiers Media S.A., Nov. 25, 2020. doi: 10.3389/fcimb.2020.587269.

[135] Chun-Chieh Hsu J., Laurent-Rolle M., Pawlak J. B., Wilen C. B., Cresswell P., Garcia-Blanco M., Horner S. M., 'Translational shutdown and evasion of the innate immune response by SARS-CoV-2 NSP14 protein', doi: 10.1073/pnas.2101161118/-/DCSupplemental.

[136] Nieto-Torres J. L., Verdiá-Báguena C., Jimenez-Guardeño J. M., Regla-Nava J. A., Castaño-Rodriguez C., Fernandez-Delgado R., Torres J., Aguilella V. M., Enjuanes L., (2015), 'Severe acute respiratory syndrome coronavirus E protein transports calcium ions and activates the NLRP3 inflammasome', Virology, 485 330–339, doi: 10.1016/j.virol.2015.08.010.

[137] Jean Millet A. K., Jaimes J. A., Whittaker G. R., (2020), 'Title Molecular diversity of coronavirus host cell entry receptors One-sentence summary This review

reframes and puts into perspective SARS-CoV-2 receptor usage in the broader landscape of coronavirus spike interactions with animal host cell receptors. Corresponding author', doi: 10.1093/femsre/fuaa057/5942658.

[138] Bian J., Li Z., 'Angiotensin-converting enzyme 2 (ACE2): SARS-CoV-2 receptor and RAS modulator', *Acta Pharmaceutica Sinica B*, 11 (1), . Chinese Academy of Medical Sciences, 1–12, Jan. 01, 2021. doi: 10.1016/j.apsb.2020.10.006.

[139] Wirnsberger G. *et al.*, (2021), 'Clinical grade ACE2 as a universal agent to block SARS-CoV-2 variants', bioRxiv, 2021.09.10.459744, doi: 10.1101/2021.09.10.459744.

[140] Shin Y. H., Jeong K., Lee J., Lee H. J., Yim J., Kim J., Kim S., Park S. B., (2022), 'Inhibition of ACE2-Spike Interaction by an ACE2 Binder Suppresses SARS-CoV-2 Entry', Angewandte Chemie - International Edition, 61 (11), , doi: 10.1002/anie.202115695.

[141] Hwang S. S. *et al.*, (2020), 'MRNA destabilization by BTG1 and BTG2 maintains T cell quiescence', Science, 367 (6483), 1255–1260, doi: 10.1126/science.abb2507.

[142] Zhou T. *et al.*, (2020), 'Cryo-EM Structures of SARS-CoV-2 Spike without and with ACE2 Reveal a pH-Dependent Switch to Mediate Endosomal Positioning of Receptor-Binding Domains', Cell Host and Microbe, 28 (6), 867-879.e5, doi: 10.1016/j.chom.2020.11.004.

[143] Hoffmann M. *et al.*, (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.e8, doi: 10.1016/j.cell.2020.02.052.

[144] Koch J., Uckeley Z. M., Doldan P., Stanifer M., Boulant S., Lozach P., (2021), 'TMPRSS2 expression dictates the entry route used by SARS-CoV-2 to infect host cells', The EMBO Journal, 40 (16), , doi: 10.15252/embj.2021107821.

[145] Jocher G. *et al.*, (2022), 'ADAM10 and ADAM17 promote SARS-CoV-2 cell entry and spike protein-mediated lung cell fusion', EMBO reports, 23 (6), , doi: 10.15252/embr.202154305.

[146] Cheng Y. W. *et al.*, (2020), 'Furin Inhibitors Block SARS-CoV-2 Spike Protein Cleavage to Suppress Virus Production and Cytopathic Effects', Cell Reports, 33 (2), , doi: 10.1016/j.celrep.2020.108254.

[147] Jackson C. B., Farzan M., Chen B., Choe H., 'Mechanisms of SARS-CoV-2 entry into cells', *Nature Reviews Molecular Cell Biology*, 23 (1), . Nature Research, 3–20, Jan. 01, 2022. doi: 10.1038/s41580-021-00418-x.

[148] Peacock T. P. *et al.*, (2021), 'The furin cleavage site in the SARS-CoV-2 spike protein is required for transmission in ferrets', Nature Microbiology, 6 (7), 899–909, doi: 10.1038/s41564-021-00908-w.

[149] Karoyan P., Vieillard V., Gómez-Morales L., Odile E., Guihot A., Luyt C. E., Denis A., Grondin P., Lequin O., (2021), 'Human ACE2 peptide-mimics block SARS-CoV-2 pulmonary cells infection', Communications Biology, 4 (1), , doi: 10.1038/s42003-021-01736-8.

[150] Laporte M. *et al.*, (2021), 'The SARS-CoV-2 and other human coronavirus spike proteins are fine-tuned towards temperature and proteases of the human airways', PLoS Pathogens, 17 (4), , doi: 10.1371/journal.ppat.1009500.

[151] Hui K. P. Y. *et al.*, (2022), 'SARS-CoV-2 Omicron variant replication in human bronchus and lung ex vivo', Nature, 603 (7902), 715–720, doi: 10.1038/s41586-022-04479-6.

[152] Meng B. *et al.*, (2022), 'Altered TMPRSS2 usage by SARS-CoV-2 Omicron impacts infectivity and fusogenicity', Nature, 603 (7902), 706–714, doi: 10.1038/s41586-022-04474-x.

[153] Li K., Meyerholz D. K., Bartlett J. A., McCray P. B., (2021), 'The tmprss2 inhibitor nafamostat reduces sars-cov-2 pulmonary infection in mouse models of covid-19', mBio, 12 (4), , doi: 10.1128/mBio.00970-21.

[154] Brevini T. *et al.*, (2023), 'FXR inhibition may protect from SARS-CoV-2 infection by reducing ACE2', Nature, 615 (7950), 134–142, doi: 10.1038/s41586-022-05594-0.

[155] Samelson A. J. *et al.*, 'BRD2 inhibition blocks SARS-CoV-2 infection by reducing transcription of the host cell receptor ACE2', doi: 10.1101/2021.01.19.427194.

[156] Chu H. *et al.*, (2022), 'Coronaviruses exploit a host cysteine-aspartic protease for replication', Nature, 609 (7928), 785–792, doi: 10.1038/s41586-022-05148-4.

[157] Geng J. *et al.*, (2021), 'CD147 antibody specifically and effectively inhibits infection and cytokine storm of SARS-CoV-2 and its variants delta, alpha, beta, and gamma', Signal Transduction and Targeted Therapy, 6 (1), , doi: 10.1038/s41392-021-00760-8.

[158] White K. M. *et al.*, 'Plitidepsin has potent preclinical efficacy against SARS-CoV-2 by targeting the host protein eEF1A'. [Online]. Available: https://www.science.org

[159] Xiaolei Liua *et al.*, (2021), 'Targeting the coronavirus nucleocapsid protein through GSK-3 inhibition', PNAS Direct Submission, 118 , doi: 10.1073/pnas.2113401118/-/DCSupplemental.

[160] Braga L. *et al.*, (2021), 'Drugs that inhibit TMEM16 proteins block SARS-CoV-2 spike-induced syncytia', Nature, 594 (7861), 88–93, doi: 10.1038/s41586-021-03491-6.

[161] Barnette K. G. *et al.*, (2022), 'Oral Sabizabulin for High-Risk, Hospitalized Adults with Covid-19: Interim Analysis', NEJM Evidence, 1 (9), , doi: 10.1056/evidoa2200145.

[162] Bian H. *et al.*, (2023), 'Meplazumab in hospitalized adults with severe COVID-19 (DEFLECT): a multicenter, seamless phase 2/3, randomized, third-party double-blind clinical trial', Signal Transduction and Targeted Therapy, 8 (1), , doi: 10.1038/s41392-023-01323-9.

[163] Pantaleo G., Correia B., Fenwick C., Joo V. S., Perez L., 'Antibodies to combat viral infections: development strategies and progress', *Nature Reviews Drug Discovery*, 21 (9), . Nature Research, 676–696, Sep. 01, 2022. doi: 10.1038/s41573-022-00495-3.

[164] Mullard A., ‘FDA approves 100th monoclonal antibody product’, *Nature reviews. Drug discovery*, 20 (7), . NLM (Medline), 491–495, Jul. 01, 2021. doi: 10.1038/d41573-021-00079-7.

[165] Bhimraj A. *et al.*, (2020), ‘Infectious Diseases Society of America Guidelines on the Treatment and Management of Patients with COVID-19’, Clinical infectious diseases : an official publication of the Infectious Diseases Society of America, , doi: 10.1093/cid/ciaa478.

[166] Cox M. G. *et al.*, ‘SARS-CoV-2 variant evasion of monoclonal antibodies based on in vitro studies’, *Nature Reviews Microbiology*, 21 (2), . Nature Research, 112–124, Feb. 01, 2023. doi: 10.1038/s41579-022-00809-7.

[167] Saunders K. O., ‘Conceptual approaches to modulating antibody effector functions and circulation half-life’, *Frontiers in Immunology*, 10 (JUN), . Frontiers Media S.A., 2019. doi: 10.3389/fimmu.2019.01296.

[168] Focosi D., Franchini M., Pirofski L. A., Burnouf T., Paneth N., Joyner M. J., Casadevall A., ‘COVID-19 Convalescent Plasma and Clinical Trials: Understanding Conflicting Outcomes’, *Clinical Microbiology Reviews*, 35 (3), . American Society for Microbiology, Sep. 01, 2022. doi: 10.1128/cmr.00200-21.

[169] Senefeld J. W., Casadevall A., Joyner M. J., ‘Convalescent plasma to deliver therapeutic antibodies against COVID-19’, *Trends in Molecular Medicine*, 28 (5), . Elsevier Ltd, 435–436, May 01, 2022. doi: 10.1016/j.molmed.2022.02.005.

[170] Barnes C. O. *et al.*, (2020), ‘SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies’, Nature, 588 (7839), 682–687, doi: 10.1038/s41586-020-2852-1.

[171] Piccoli L. *et al.*, (2020), ‘Mapping Neutralizing and Immunodominant Sites on the SARS-CoV-2 Spike Receptor-Binding Domain by Structure-Guided High-Resolution Serology’, Cell, 183 (4), 1024-1042.e21, doi: 10.1016/j.cell.2020.09.037.

[172] Taylor P. C., Adams A. C., Hufford M. M., de la Torre I., Winthrop K., Gottlieb R. L., ‘Neutralizing monoclonal antibodies for treatment of COVID-19’, *Nature*

Reviews Immunology, 21 (6), . Nature Research, 382–393, Jun. 01, 2021. doi: 10.1038/s41577-021-00542-x.

[173] Prof Peter W Horby, Prof Martin J Landray, RECOVERY Central Coordinating Office, 'Casirivimab and imdevimab in patients admitted to hospital with COVID-19 (RECOVERY): a randomised, controlled, open-label, platform trial'. doi: 10.1016/S0140-6736(22)00163-5.

[174] Majumdar P., Niyogi S., 'SARS-CoV-2 mutations: The biological trackway towards viral fitness', *Epidemiology and Infection*. Cambridge University Press, 2021. doi: 10.1017/S0950268821001060.

[175] Harvey W. T. *et al.*, 'SARS-CoV-2 variants, spike mutations and immune escape', *Nature Reviews Microbiology*, 19 (7), . Nature Research, 409–424, Jul. 01, 2021. doi: 10.1038/s41579-021-00573-0.

[176] Shaqra A. M. *et al.*, (2022), 'Defining the substrate envelope of SARS-CoV-2 main protease to predict and avoid drug resistance', Nature Communications, 13 (1), , doi: 10.1038/s41467-022-31210-w.

[177] Pawson Tony, (1995), 'Protein modules and signaling networks', Nature, 373 , doi: 10.1038/373573a0.

[178] Pawson T., Raina M., Nash P., 'Interaction domains: From simple binding events to complex cellular behavior', *FEBS Letters*, 513 (1), . 2–10, Feb. 20, 2002. doi: 10.1016/S0014-5793(01)03292-6.

[179] Pawson T., Nash P., 'Assembly of Cell Regulatory Systems Through Protein Interaction Domains'. [Online]. Available: http://www.sciencemag.org/cgi/content/full/300/5618/445/DC1.

[180] Warren L DeLano, 'Unraveling hot spots in binding interfaces: progress and challenges', San Francisco. doi: 10.1016/s0959-440x(02)00283-x.

[181] Shoemaker B. A., Panchenko A. R., 'Deciphering protein-protein interactions. Part I. Experimental techniques and databases', *PLoS Computational Biology*, 3 (3), . Public Library of Science, 0337–0344, 2007. doi: 10.1371/journal.pcbi.0030042.

[182] London N., Movshovitz-Attias D., Schueler-Furman O., (2010), 'The Structural Basis of Peptide-Protein Binding Strategies', Structure, 18 (2), 188–199, doi: 10.1016/j.str.2009.11.012.

[183] Vanhee P., Stricher F., Baeten L., Verschueren E., Lenaerts T., Serrano L., Rousseau F., Schymkowitz J., (2009), 'Protein-Peptide Interactions Adopt the Same Structural Motifs as Monomeric Protein Folds', Structure, 17 (8), 1128–1136, doi: 10.1016/j.str.2009.06.013.

[184] Ren R., Mayer B. J., Cicchetti P., Baltimore D., 'Identification of a Ten-Amino Acid Proline-Rich SH3 Binding Site'.

[185] Castagnoli L., Costantini A., Dall'Armi C., Gonfloni S., Montecchi-Palazzi L., Panni S., Paoluzi S., Santonico E., Cesareni G., 'Selectivity and promiscuity in the interaction network mediated by protein recognition modules', in *FEBS Letters*, Jun. 2004, 74–79. doi: 10.1016/j.febslet.2004.03.116.

[186] Encinar J. A., Fernandez-Ballester G., Sánchez I. E., Hurtado-Gomez E., Stricher F., Beltrao P., Serrano L., (2009), 'ADAN: A database for prediction of protein-protein interaction of modular domains mediated by linear motifs', Bioinformatics, 25 (18), 2418–2424, doi: 10.1093/bioinformatics/btp424.

[187] Gorelik M., Davidson A. R., (2012), 'Distinct peptide binding specificities of Src homology 3 (SH3) protein domains can be determined by modulation of local energetics across the binding interface', Journal of Biological Chemistry, 287 (12), 9168–9177, doi: 10.1074/jbc.M111.330753.

[188] Zarrinpar A., Bhattacharyya R. P., Lim W. A., 'The structure and function of proline recognition domains.', *Science's STKE : signal transduction knowledge environment*, 2003 (179), . 2003. doi: 10.1126/stke.2003.179.re8.

[189] Ali Zarrinpar, Sang-Hyun Park, Wendell A. Lim, (2003), 'Optimization of specificity in a cellular protein interaction network by negative selection', Nature, 426 , doi: 10.1038/nature02178.

[190] Stollar E. J., Garcia B., Chong P. A., Rath A., Lin H., Forman-Kay J. D., Davidson A. R., (2009), 'Structural, functional, and bioinformatic studies demonstrate the

crucial role of an extended peptide binding site for the SH3 domain of yeast Abp1p', Journal of Biological Chemistry, 284 (39), 26918–26927, doi: 10.1074/jbc.M109.028431.

[191] Dalgarno D. C., Botfield M. C., Rickles R. J., 'SH3 Domains and Drug', John Wiley & Sons, Inc. Biopoly.

[192] Mayer B. J., 'SH3 domains: complexity in moderation COMMENTARY Signal Transduction and Cellular Organization'. [Online]. Available: http://pfam.wustl.edu/]

[193] Gfeller D. *et al.*, (2011), 'The multiple-specificity landscape of modular peptide recognition domains', Molecular Systems Biology, 7 , doi: 10.1038/msb.2011.18.

[194] Wang C., Nguyen P. H., Pham K., Huynh D., Le T. B. N., Wang H., Ren P., Luo R., (2016), 'Calculating protein-ligand binding affinities with MMPBSA: Method and error analysis', Journal of computational chemistry, 37 (27), 2436–2446, doi: 10.1002/jcc.24467.

[195] Wang E., Liu H., Wang J., Weng G., Sun H., Wang Z., Kang Y., Hou T., (2020), 'Development and evaluation of MM/GBSA based on a variable dielectric GB model for predicting protein-ligand binding affinities', Journal of Chemical Information and Modeling, 60 (11), 5353–5365, doi: 10.1021/acs.jcim.0c00024.

[196] Ravindranathan K., Tirado-Rives J., Jorgensen W. L., Guimarães C. R. W., (2011), 'Improving MM-GB/SA scoring through the application of the variable dielectric model', Journal of Chemical Theory and Computation, 7 (12), 3859–3865, doi: 10.1021/ct200565u.

[197] Ravindranathan K., Tirado-Rives J., Jorgensen W. L., Guimarães C. R. W., (2011), 'Improving MM-GB/SA Scoring through the Application of the Variable Dielectric Model', J. Chem. Theory Comput., 7 3859.

[198] Akkus E., Tayfuroglu O., Yildiz M., Kocak A., (2022), 'Accurate Binding Free Energy Method from End-State MD Simulations', Journal of Chemical Information and Modeling, 62 (17), 4095–4106, doi: 10.1021/acs.jcim.2c00601.

[199] Levy R. M., Zhang L. Y., Gallicchio E., Felts A. K., (2003), 'On the nonpolar hydration free energy of proteins: Surface area and continuum solvent models for the solute-solvent interaction energy', Journal of the American Chemical Society, 125 (31), 9523–9530, doi: 10.1021/ja029833a.

[200] Zacharias M., (2003), 'Continuum solvent modeling of nonpolar solvation: Improvement by separating surface area dependent cavity and dispersion contributions', Journal of Physical Chemistry A, 107 (16), 3000–3004, doi: 10.1021/jp027598c.

[201] Gallicchio E., Levy R. M., (2004), 'AGBNP: An Analytic Implicit Solvent Model Suitable for Molecular Dynamics Simulations and High-Resolution Modeling', Journal of Computational Chemistry, 25 (4), 479–499, doi: 10.1002/jcc.10400.

[202] Su Y., Gallicchio E., (2004), 'The non-polar solvent potential of mean force for the dimerization of alanine dipeptide: The role of solute-solvent van der Waals interactions', Biophysical Chemistry, 109 (2), 251–260, doi: 10.1016/j.bpc.2003.11.007.

[203] Dzubiella J., Swanson J. M. J., McCammon J. A., (2006), 'Coupling nonpolar and polar solvation free energies in implicit solvent models', Journal of Chemical Physics, 124 (8), , doi: 10.1063/1.2171192.

[204] Tan C., Tan Y. H., Luo R., (2007), 'Implicit nonpolar solvent models', Journal of Physical Chemistry B, 111 (42), 12263–12274, doi: 10.1021/jp073399n.

[205] Wagoner J. A., Baker N. A., 'Assessing implicit models for nonpolar mean solvation forces: The importance of dispersion and volume terms'. [Online]. Available: https://www.pnas.org

[206] Roe D. R., Cheatham T. E., (2013), 'PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data', Journal of Chemical Theory and Computation, 9 (7), 3084–3095, doi: 10.1021/ct400341p.

BIOGRAPHY

Hilal Sena Taşçı, pursuing her academic journey, she graduated with distinction from the Genetics and Bioengineering department at Istanbul Bilgi University between 2014 and 2019. Between 2020 and 2022, she dedicated her expertise as a molecular biologist at a COVID-19 diagnostic center. Concurrently, in 2020, she embarked on her master's journey in Molecular Biology and Genetics at Gebze Technical University, successfully completing the program in August 2023. Her commitment to the field and her relentless pursuit of knowledge have been remarkable features of her academic and professional journey.