

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

**PREDICTING NOVEL SMALL INHIBITORS OF SARS-COV-2:
TARGETING SARS-COV-2 SPIKE PROTEIN, HUMAN ACE2 PROTEIN
AND SARS-COV-2 NSP16 VIA MOLECULAR DOCKING**



M.Sc. THESIS

Onur ÖZER

Department of Molecular Biology-Genetics and Biotechnology

Molecular Biology-Genetics and Biotechnology Programme

FEBRUARY 2022

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

**PREDICTING NOVEL SMALL INHIBITORS OF SARS-COV-2:
TARGETING SARS-COV-2 SPIKE PROTEIN, HUMAN ACE2 PROTEIN
AND SARS-COV-2 NSP16 VIA MOLECULAR DOCKING**

M.Sc. THESIS

**Onur ÖZER
(521181117)**

Department of Molecular Biology-Genetics and Biotechnology

Molecular Biology-Genetics and Biotechnology Programme

**Thesis Advisor : Assoc. Prof. Mert GÜR
Thesis Co-Advisor: Assist. Prof Sefer BADAY**

FEBRUARY 2022

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ LİSANSÜSTÜ EĞİTİM
ENSTİTÜSÜ

**SARS-COV-2 İÇİN YENİ KÜÇÜK İNHİBİTÖR MOLEKÜLLERİN TAHMİNİ:
MOLEKÜLER YANAŞTIRMA YÖNTEMİYLE SARS-COV-2 SPIKE PROTEİNİ,
İNSAN ACE2 PROTEİNİ VE SARS-COV-2 NSP16 HEDEFLLENMESİ**

YÜKSEK LİSANS TEZİ

Onur ÖZER
(521181117)

Moleküler Biyoloji-Genetik ve Biyoteknoloji Anabilim Dalı

Moleküler Biyoloji-Genetik ve Biyoteknoloji Programı

Tez Danışmanı: Doç. Dr. Mert GÜR
Eş Danışman: Dr. Öğr. Üyesi Sefer BADAY

ŞUBAT 2022

Onur ÖZER, a M.Sc student of ITU Graduate School student ID 521181117, successfully defended the thesis/dissertation entitled “PREDICTING NOVEL SMALL INHIBITORS OF SARS-COV-2: TARGETING SARS-COV-2 SPIKE PROTEIN, HUMAN ACE2 PROTEIN AND SARS-COV-2 NSP16 VIA MOLECULAR DOCKING”, which he prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor : **Assoc. Prof. Mert GÜR**
Istanbul Technical University

Co-Advisor : **Assist. Prof. Sefer BADAY**
Istanbul Technical University

Jury Members : **Prof. Dr. Gizem Dinler DOĞANAY**
Istanbul Technical University

Assoc. Prof. Nurcan TUNÇBAĞ
Koç University

Assist. Prof. Özge ŞENSOY
Medipol University

Date of Submission :14 January 2022

Date of Defense : 16 February 2022





To my family,



FOREWORD

This thesis was conducted under the supervision of Assoc. Prof. Mert GÜR and Assist. Prof. Sefer BADAY in Istanbul Technical University between 2020 and 2022.

I would like to thank my supervisor Assoc. Prof. Mert GÜR and my co-advisor Assist. Prof. Sefer BADAY for giving me a greater understanding of a new area in molecular biology, which is the molecular dynamics and biocomputational drug development; and broadening my horizons on these particular topics. They showed me a different perspective for molecular biology, which is different and diverse than wet-lab research, by providing a deeper insight. Thanks to this opportunity, this branch of science became one of my main scientific career focuses to pursue.

Additionally, I am thankful to all of my lab colleagues from both GÜR and BADAY labs to help my research problems over the course of this three years graduate school period. Moreover, I specially want to thank my dear friend Ayberk ÖZKAN for his tremendous support and helping me getting through all the hardships of graduate school life.

Furthermore, I want to thank TÜBİTAK 2210-A National Scholarship Programme for M.Sc Students (TÜBİTAK BİDEB 2210-A Genel Yurt İçi Yüksek Lisans Burs Programı) for their financial support during my thesis process. Additionally, computing resources used in this work were provided by the National Center for High Performance Computing of Turkey (UHcM) under grant numbers 1004022016 and 5009682021.

Lastly, I would like to thank my father Melih ÖZER, my mother Nuriye ÖZER and especially my sister Selin ÖZER for their huge support during my graduate life.

February 2022

Onur ÖZER

TABLE OF CONTENTS

	<u>Page</u>
FOREWORD	ix
TABLE OF CONTENTS	xi
ABBREVIATIONS	xiii
LIST OF TABLES	xv
LIST OF FIGURES	xvii
SUMMARY	xxi
ÖZET	xxv
1. INTRODUCTION	1
1.1. Coronaviruses and SARS-CoV-2	1
1.2. Structure of SARS-CoV-2	2
1.3. Spike Protein and Receptor Binding Domain	4
1.4. Infection Mechanism of SARS-CoV-2	6
1.5. Nsp16-Nsp10 Interaction Mechanism in SARS-CoV-2	7
1.6. Molecular Modeling	8
1.7. Small-molecule Inhibitor and Drug Development	13
1.8. Molecular Docking Research Against SARS-CoV-2	14
1.9. Purpose of the Thesis	15
2. MATERIALS AND METHODS	17
2.1. Selection of Target Proteins for Molecular Docking Simulations	17
2.2. Ligand Preparation for Molecular Docking	21
2.3. Protein Preparation for Molecular Docking	22
2.4. Molecular Docking Simulations using AutoDock Vina	22
2.5. Extra-precision Docking using Schrödinger Maestro Glide XP	26
2.6. Fingerprint Clustering using Schrödinger CANVAS	31
2.7. ADME/T Analyses using Schrödinger QikProp	31
2.8. Analyses of Interaction between Target Proteins and Ligands	33
2.9. Molecular Dynamic Simulations using Schrödinger Desmond	33
3. RESULTS	35
3.1. Small-molecule Inhibitor Prediction for SARS-CoV-2 Spike RBD	35
3.2. Small-molecule Inhibitor Prediction for Human ACE2 Receptor	47
3.3. Small-molecule Inhibitor Prediction for Nsp16-Nsp10 Interaction	60
3.4. Small-molecule Inhibitor Prediction for Nsp16-SAM Interaction	73
3.5. Molecular Dynamics Simulations	87
4. DISCUSSION	101
REFERENCES	105
APPENDICES	111
CURRICULUM VITAE	121



ABBREVIATIONS

ACE2	:Angiotensin converting enzyme 2
ADME/T	:Absorption, distribution, metabolism, and excretion/toxicity
ADT	:AutoDock Tools
CHARMM	:Chemistry at Harvard macromolecular mechanics
CoV	:Coronavirus
COVID-19	:2019 Coronavirus Disease
E	:Envelope protein
HTVS	:High-throughput screening
KD	:Equilibrium diassociation constant
M	:Membrane/matrix protein
MAE	:Meastro file format
MD	:Molecular Dynamics
N	:Nucleocapsid protein
NSP	:Non-structural protein
OPLS	:Optimized Potentials for Liquid Simulations
PDB	:Protein data bank
PDBQT	:Protein data bank, partial charge & atom type
RBD	:Receptor binding domain
RMSD	:Root-mean-square deviation
S	:Spike protein
SAM	:S-Adenosyl methionine
SARS	:Severe acute respiratory syndrome
SDF	:Structural data file
SMI	:SMILES format
SP	:Standard-precision
XP	:Extra-precision



LIST OF TABLES

	<u>Page</u>
Table 1.1: Infographic of three major Coronavirus outbreaks	2
Table 2.1: Binding profiles between SARS-CoV-2 RBD and ACE2	20
Table 2.2: Important pharmaceutical properties of ligands for ADME/T analysis ...	32
Table 3.1: Low-energy and high binding-affinity of 10 ligands with the lowest state penalty, the highest docking scores and optimum ADME/T values against ACE2 binding site of Spike RBD	39
Table 3.2: Protein-ligand interaction profile of Spike RBD and top-scored 10 small-molecule ligands.....	46
Table 3.3: Low-energy and high binding-affinity of 10 ligands with the lowest state penalty, the highest docking scores and optimum ADME/T values against RBD binding site of human ACE2.....	52
Table 3.4: Protein-ligand interaction profile of ACE2 and top-scored 10 small-molecule ligands.....	59
Table 3.5: Low-energy and high binding-affinity of 10 ligands with the lowest state penalty, the highest docking scores and optimum ADME/T values against Nsp10 binding site of Nsp16.....	64
Table 3.6: Protein-ligand interaction profile of Nsp16 for its Nsp10 binding site and top-scored 10 small-molecule ligands.....	72
Table 3.7: Low-energy and high binding-affinity of 10 ligands with the lowest state penalty, the highest docking scores and optimum ADME/T values against SAM binding site of Nsp16.....	78
Table 3.8: Protein-ligand interaction profile of Nsp16 for its SAM binding site and top-scored 10 small-molecule ligands.....	85
Table A.1: 163 ligand clusters generated by CANVAS dendritic fingerprinting hierarchical clustering for Spike RBD	112
Table A.2: 166 ligand clusters generated by CANVAS dendritic fingerprinting hierarchical clustering for ACE2.....	112
Table A.3: 161 ligand clusters generated by CANVAS dendritic fingerprinting hierarchical clustering for Nsp10 binding site of Nsp16.....	112
Table A.4: 164 ligand clusters generated by CANVAS dendritic fingerprinting hierarchical clustering for SAM binding site of Nsp16	112



LIST OF FIGURES

	<u>Page</u>
Figure 1.1: Schematic representation of SARS-CoV-2 structure.....	3
Figure 1.2: ‘Down/inactive’ and ‘up/active’ states of SARS-CoV-2 structures	5
Figure 1.3: The interaction of SARS-Cov-2 in its ‘up’/active conformation with human ACE2 receptor.....	6
Figure 1.4: The ss (+) mRNA genome (top) and the 3-D structures of Nsp10-Nsp16 complex and SAM.....	8
Figure 1.5: The summary of molecular docking process	12
Figure 2.1: The 3-D structure of homotrimeric SARS-CoV-2 Spike protein in its ‘up’/active conformation, visualized in UCSF Chimera.....	17
Figure 2.2: The 3-D structure of single monomer of fully glycosylated Spike protein, visualized in UCSF Chimera.....	18
Figure 2.3: The 3-D Structure of human ACE2, visualized in UCSF Chimera	19
Figure 2.4: Spike RBD and ACE2 binding interface, visualized in UCSF Chimera .	19
Figure 2.5: The 3-D Structure of Nsp16 and Nsp10 in complex with SAM in binding pocket of Nsp16	20
Figure 2.6: Selection of ZINC library consisting of 6.2 million ligands.....	21
Figure 2.7: Docking box on RBD of Spike protein for AutoDock Vina calculations, visualized in UCSF Chimera.....	23
Figure 2.8: Docking box on human ACE2 protein for AutoDock Vina calculations, visualized in UCSF Chimera.....	24
Figure 2.9: Docking box on Nsp10 binding site of Nsp16 protein for AutoDock Vina calculations, visualized in UCSF Chimera.....	25
Figure 2.10: Docking box on SAM binding site of Nsp16 protein for AutoDock Vina calculations, visualized in UCSF Chimera.....	26
Figure 2.11: The structure of 6VSB_1_1_1 integrated with grid box on Schrödinger Maestro.....	28
Figure 2.12: The structure of 6VW1 integrated with grid box on Schrödinger Maestro	29
Figure 2.13: The structure of 6W4H integrated with grid box on Schrödinger Maestro	30
Figure 2.14: The structure of 6W4H integrated with grid box on Schrödinger Maestro	30
Figure 3.1: Distribution of AutoDock Vina docking scores (x-axis) of 400.000 ligands (y-axis) against ACE2 binding site on Spike RBD	35
Figure 3.2: Distribution of AutoDock Vina docking scores (x-axis) of 50.000 ligands (y-axis) against ACE2 binding site on Spike RBD	36
Figure 3.3: Distribution of AutoDock Vina docking scores (x-axis) of 5.000 ligands (y-axis) against ACE2 binding site on Spike RBD	36
Figure 3.4: Distribution of docking scores (x-axis) of 5.000 ligands (y-axis) against ACE2 binding site on Spike RBD from Glide XP docking simulation ..	37

Figure 3.5: Three clusters from dendritic fingerprinting hierarchical clustering of 200 ligands against ACE2 binding site on Spike RBD, identified by CANVAS Software.....	38
Figure 3.6: Top-scored 10 ligands in complex with Spike RBD, visualized on Maestro	41
Figure 3.7: Ligand interaction diagrams of top-scored 10 ligands with Spike RBD, visualized by Schrödinger Maestro	44
Figure 3.8: Distribution of AutoDock Vina docking scores (x-axis) of 400.000 ligands (y-axis) against Spike RBD binding site of human ACE2	47
Figure 3.9: Distribution of AutoDock Vina docking scores (x-axis) of 50.000 ligands (y-axis) against Spike RBD binding site of human ACE2	48
Figure 3.10: Distribution of AutoDock Vina docking scores (x-axis) of 5.000 ligands (y-axis) against Spike RBD binding site of human ACE2	49
Figure 3.11: Distribution of docking scores (x-axis) of 5.000 ligands (y-axis) against Spike RBD binding site of human ACE2 from Glide XP docking simulation	50
Figure 3.12: Three clusters from dendritic fingerprinting hierarchical clustering of 200 ligands against Spike RBD binding site of human ACE2, identified by CANVAS Software	50
Figure 3.13: Top-scored 10 ligands in complex with human ACE2, visualized on Maestro	54
Figure 3.14: Ligand interaction diagrams of top-scored 10 ligands with human ACE2, visualized by Schrödinger Maestro	57
Figure 3.15: Distribution of AutoDock Vina docking scores (x-axis) of 400.000 ligands (y-axis) against Nsp10 binding site of Nsp16.....	60
Figure 3.16: Distribution of AutoDock Vina docking scores (x-axis) of 50.000 ligands (y-axis) against Nsp10 binding site of Nsp16	61
Figure 3.17: Distribution of AutoDock Vina docking scores (x-axis) of 5.000 ligands (y-axis) against Nsp10 binding site of Nsp16	61
Figure 3.18: Distribution of docking scores (x-axis) of 5.000 ligands (y-axis) against Nsp10 binding site of Nsp16 from Glide XP docking simulation	62
Figure 3.19: Three clusters from dendritic fingerprinting hierarchical clustering of 200 ligands against Nsp10 binding site of Nsp16, identified by CANVAS Software.....	63
Figure 3.20: Top-scored 10 ligands within the Nsp16-Nsp10 binding interface, visualized on Maestro.....	66
Figure 3.21: Ligand interaction diagrams of top-scored 10 ligands with Nsp16, visualized by Schrödinger Maestro	70
Figure 3.22: Distribution of AutoDock Vina docking scores (x-axis) of 400.000 ligands (y-axis) against SAM binding site of Nsp16	74
Figure 3.23: Distribution of AutoDock Vina docking scores (x-axis) of 50.000 ligands (y-axis) against SAM binding site of Nsp16	74
Figure 3.24: Distribution of AutoDock Vina docking scores (x-axis) of 5.000 ligands (y-axis) against SAM binding site of Nsp16	75
Figure 3.25: Distribution of docking scores (x-axis) of 5.000 ligands (y-axis) against SAM binding site of Nsp16 from Glide XP docking simulation	76
Figure 3.26: Three clusters from dendritic fingerprinting hierarchical clustering of 200 ligands against SAM binding site of Nsp16, identified by CANVAS Software.....	77

Figure 3.27: Top-scored 10 ligands in complex within SAM binding pocket of Nsp16, visualized on Maestro.....	80
Figure 3.28: Ligand interaction diagrams of top-scored 10 ligands with Nsp16, visualized by Schrödinger Maestro	83
Figure 3.29: Changes in RMSD of Spike RBD C α and side chains and ligand ZINC000067920242.....	88
Figure 3.30: Changes in RMSD of Spike RBD C α and side chains and ligand ZINC000009828805.....	88
Figure 3.31: Changes in RMSD of Spike RBD C α and side chains and ligand ZINC000014228954.....	89
Figure 3.32: Changes in RMSD of Spike RBD C α and side chains and ligand ZINC000009134748.....	90
Figure 3.33: Changes in RMSD of Spike RBD C α and side chains and ligand ZINC000008772453.....	90
Figure 3.34: Changes in RMSD of ACE2 C α and side chains and ligand ZINC000014879822.....	91
Figure 3.35: Changes in RMSD of ACE2 C α and side chains and ligand ZINC000000813934.....	92
Figure 3.36: Changes in RMSD of ACE2 C α and side chains and ligand ZINC000019742660.....	92
Figure 3.37: Changes in RMSD of ACE2 C α and side chains and ligand ZINC000019022189.....	93
Figure 3.38: Changes in RMSD of ACE2 C α and side chains and ligand ZINC000020149756.....	94
Figure 3.39: Changes in RMSD of Nsp16 C α and side chains and ligand ZINC000009751488.....	94
Figure 3.40: Changes in RMSD of Nsp16 C α and side chains and ligand ZINC000002286111.....	95
Figure 3.41: Changes in RMSD of Nsp16 C α and side chains and ligand ZINC000019497136.....	96
Figure 3.42: Changes in RMSD of Nsp16 C α and side chains and ligand ZINC000009252898.....	96
Figure 3.43: Changes in RMSD of Nsp16 C α and side chains and ligand ZINC000009485446.....	97
Figure 3.44: Changes in RMSD of Nsp16 C α and side chains and ligand ZINC000019373581.....	98
Figure 3.45: Changes in RMSD of Nsp16 C α and side chains and ligand ZINC000096137984.....	98
Figure 3.46: Changes in RMSD of Nsp16 C α and side chains and ligand ZINC000257313608.....	99
Figure 3.47: Changes in RMSD of Nsp16 C α and side chains and ligand ZINC000022289086.....	100
Figure 3.48: Changes in RMSD of Nsp16 C α and side chains and ligand ZINC000072395414.....	100
Figure B.1: Comparison of the first and the last frame of ZINC000067920242 within the ACE2 binding site of Spike RBD.....	113
Figure B.2: Comparison of the first and the last frame of ZINC000009828805 within the ACE2 binding site of Spike RBD.....	113
Figure B.3: Comparison of the first and the last frame of ZINC000014228954 within the ACE2 binding site of Spike RBD.....	114

Figure B.4: Comparison of the first and the last frame of ZINC000009134748 within the ACE2 binding site of Spike RBD.....	114
Figure B.5: Comparison of the first and the last frame of ZINC000008772453 within the ACE2 binding site of Spike RBD.....	114
Figure B.6: Comparison of the first and the last frame of ZINC000014879822 within the Spike RBD binding site of ACE2.....	115
Figure B.7: Comparison of the first and the last frame of ZINC000000813934 within the Spike RBD binding site of ACE2.....	115
Figure B.8: Comparison of the first and the last frame of ZINC000019742660 within the Spike RBD binding site of ACE2.....	115
Figure B.9: Comparison of the first and the last frame of ZINC000019022189 within the Spike RBD binding site of ACE2.....	116
Figure B.10: Comparison of the first and the last frame of ZINC000020149756 within the Spike RBD binding site of ACE2.....	116
Figure B.11: Comparison of the first and the last frame of ZINC000009751488 within the Nsp10 binding site of Nsp16	116
Figure B.12: Comparison of the first and the last frame of ZINC000002286111 within the Nsp10 binding site of Nsp16	117
Figure B.13: Comparison of the first and the last frame of ZINC000019497136 within the Nsp10 binding site of Nsp16	117
Figure B.14: Comparison of the first and the last frame of ZINC000009252898 within the Nsp10 binding site of Nsp16	117
Figure B.15: Comparison of the first and the last frame of ZINC000009485446 within the Nsp10 binding site of Nsp16	118
Figure B.16: Comparison of the first and the last frame of ZINC000019373581 within the SAM binding site of Nsp16.....	118
Figure B.17: Comparison of the first and the last frame of ZINC000096137984 within the SAM binding site of Nsp16.....	118
Figure B.18: Comparison of the first and the last frame of ZINC000257313608 within the SAM binding site of Nsp16.....	119
Figure B.19: Comparison of the first and the last frame of ZINC000022289086 within the SAM binding site of Nsp16.....	119
Figure B.20: Comparison of the first and the last frame of ZINC000072395414 within the SAM binding site of Nsp16.....	119

PREDICTING NOVEL SMALL INHIBITORS OF SARS-COV-2: TARGETING SARS-COV-2 SPIKE PROTEIN, HUMAN ACE2 PROTEIN AND SARS-COV-2 NSP16 VIA MOLECULAR DOCKING

SUMMARY

In 2019, a virus with an unknown origin first emerged in Wuhan, China and was identified as a type of severe acute respiratory syndrome (SARS) coronavirus and named 2019 novel coronavirus (2019-nCoV) by World Health Organization (WHO). The virus was later named SARS-CoV-2 and WHO declared the global pandemic on 11 March 2020. Coronaviruses are a subfamily of RNA viruses (Orthocoronavirinae) which belongs to family of Coronaviridae and order of Nidovirales. Worldwide, SARS-CoV-2 has infected 308 million people and caused 5.49 million deaths, resulting in 1.79% mortality rate.

Four major structural and functional proteins of SARS-CoV-2 are envelope (E), spike (S), nucleocapsid (N) and matrix/membrane (M) proteins. The genome is protected by N protein. Main structural protein of viral envelope is M protein. It is a transmembrane protein that interacts with other proteins; and mainly regulates viral assembly, budding and release after infection. E proteins are lipid bilayer-embedded proteins that promote assembly and pathogenesis through its ion channels. Given its similarity to mRNA, (+) stranded RNA genome of SARS-CoV-2 directs host transcriptory and translatory machinery to produce its replicase gene which encodes for two proteins which will be cleaved into non-structural proteins called NSPs. NSPs function in host immune response suppression, viral RNA replication-transcription and the host cell exit. S proteins are homotrimeric type I membrane glycoproteins, consisting of two functional domains, which are S1 and S2. 22 N-glycosylation sites on the surface of S protein are required for stabilizing the virus and maintaining its structural integrity, viral interaction with the host cell receptor, ACE-2 and promoting the proper folding of S protein.

The receptor-binding domain (RBD) of S1 subunit has two structural conformations which are 'down' conformation (inactive state) and 'up' conformation (active state). Receptor accessible 'up' conformation RBD recognizes and attaches to human cell membrane receptor angiotensin converting enzyme 2 (ACE2). This allows S1 subunit to catalyze viral attachment to the host cell. After initial attachment, host cell peptidases (e.g. transmembrane protease serine 2 [TMPRSS2]) cleave S protein into its two subunits. Lastly, S2 subunit promotes the fusion with the host membrane. Non-structural protein (Nsp16) methylates viral RNA cap using its 2'O-methyltransferase activity, carrying a methyl group from Sadenosylmethionine (SAM) onto RNA. This requires the activation of Nsp16 by its cofactor Nsp10. This action promotes replication and also required for immune evasion of SARS-CoV-2.

Molecular docking is a computational simulation method in which a ligand (e.g. small molecule, etc.) is bound to a target in a predefined binding pocket (e.g. the binding site of a protein receptor, etc.) to generate a stable complex. A library of ligands is virtually screened and docked onto the target site through molecular docking tools. Scanned

candidate ligands are scored and ranked based on their binding affinity to the target site. The ligands with more negative binding energy and higher binding affinity result in the generation of more stable bound complexes. These ligands can be considered promised inhibitory molecules for the target protein activity. There have been many molecular docking studies against SARS-CoV-2 recently, in order to find strong inhibitory small molecules that can be utilized as drugs against SARS-CoV-2. Major targets for SARS-CoV-2 in these studies have mostly focused on S protein RBD, human ACE2, viral main protease and non-structural proteins as they are key players in viral functions.

In this thesis, the studies were conducted to propose possible small-molecule inhibitor candidates for the treatment of SARS-CoV-2. 400,000 commercially available small ligands from ZINC database were virtually screened against three major SARS-CoV-2 targets (RBD of 'up' conformation Spike protein in its fully glycosylated form (PDB ID: 6VSB), human ACE2 (PDB ID: 6VW1) and Nsp16 (PDB ID: 6W4H). The docking of small ligands on the Spike RBD and ACE2 interaction surface interface was conducted to predict candidate small inhibitory molecules to block the interaction between these two proteins and thus viral cell entry. Additionally, the same library of small molecules was separately docked on the Nsp16-Nsp10 interaction site to block the full activation of Nsp16 by Nsp10; and SAM binding pocket of Nsp16 to prevent its methylation activity, thus preventing the cell replication.

Ligands and proteins were prepared for molecular docking by generating their 3-D structures and adding polar hydrogen atoms using AutoDock Tools. AutoDock Vina was used to perform molecular docking calculations on the binding interfaces of Spike RBD-ACE2, Nsp16-Nsp10 and Nsp16-SAM. Ligands were docked onto grid docking boxes, which were generated based on these sites, by AutoDock Vina using Python to calculate their binding energies and binding affinities. Ligands were scored and ranked based on their binding energy-docking scores. This method was done for a second time to narrow down 400,000 ligands to 5,000 ligands with the highest affinities to their respective targets. These 5,000 ligands and protein targets were then prepared by using LigPrep and Protein Preparation Wizard modules of Schrödinger Suite Maestro. These 5,000 ligands were screened by using extra-precision (XP) docking of Schrödinger Suite Glide XP to predict top 200 ligands with the highest docking scores. Top 200 ligands were clustered by using CANVAS software of Schrödinger Suite. Dendritic fingerprinting and hierarchical clustering were applied. From each cluster, a single ligand with the lowest state penalty and the highest docking score was identified to prevent the resemblance of ligands and increase the ligand variety for future steps. These resulting ligands were screened for their ADME/T (absorption, distribution, metabolism, and excretion/toxicity) properties to define optimum ligands for their possible pharmaceutical use. Furthermore, top-docked ligands were analyzed for their intermolecular interactions with their respective targets to see whether they can form strong and desired hydrogen-bonds and hydrophobic interactions at target-binding interfaces. Lastly, top 10 ligands among most drug-like ones were applied to molecular dynamic simulations to calculate binding free energy using OSPL2005 force field and TIP3P water model of Schrödinger Desmond Software. The most stable, high-binding affinity and most drug-like 5 ligands were identified for each target protein.

AutoDock Vina docking calculations of ligands resulted in high binding scores, indicating that selected ligands can bind to target proteins with high affinity and potentially inhibit their interactions. Additionally, two-step AutoDock Vina docking calculations increased the accuracy of results by providing consistency as the similar

binding scores were calculated for the same ligands between two docking runs. Schrödinger Suite Glide XP docking calculations resulted in a different range of slightly lower docking scores due to the fact that Glide XP utilized its own and more extensive scoring function. Despite these, Glide XP scores showed high or adequate binding affinities -in the standards of Glide XP scoring function- for top-docked ligands which are mostly the same top ones predicted in AutoDock Vina calculations. This indicates that high consistency between all docking calculation steps. Resulting 200 top ligands with highest binding affinities can be considered as potential small inhibitory molecules. Moreover, clustering results indicated that top-docked ligand candidates are different from each other in terms of chemical and structural properties. This diversity enables the increase in the number of molecule options for possible future lab experiments to test. ADME/T analyses showed that top-docked ligands, which are the representative of each cluster, were statistically closer to optimum in terms of drug-likeness since all of their pharmaceutical properties were found to be in desired range. Furthermore, ligand interaction analyses suggested that the use of these ligands can block the binding site of their respective targets as they can antagonistically overlap and bind to the same or similar residues on these binding interfaces with high affinity by forming high number of hydrogen bonds and hydrophobic interactions. Lastly, molecular dynamics simulations on these ligands showed ligand-target complex stability over the course of simulation time, confirming that identified ligands can retain their positions within binding pocket of their respective targets for possible long-time inhibition. Ligands discovered in this study can provide a good insight for the treatment of SARS-CoV-2 by being used as possible inhibitor molecules or drugs.



**SARS-COV-2 İÇİN YENİ KÜÇÜK İNHİBİTÖR MOLEKÜLLERİN TAHMİNİ:
MOLEKÜLER YANAŞTIRMA YÖNTEMİYLE SARS-COV-2 SPIKE PROTEİNİ,
İNSAN ACE2 PROTEİNİ VE SARS-COV-2 NSP16 HEDEFLENMESİ**

ÖZET

2019 yılında, bilinmeyen kökenli bir virus, Çin'in Wuhan bölgesinde ortaya çıktı ve şiddetli akut solunum yolu sendromu (SARS) koronavirüs olarak isimlendirildi. Dünya Sağlık Örgütü (WHO), bu virüsü 2019 yeni koronavirüsü (2019-nCoV) olarak isimlendirdi ve salgın, küresel pandemi olarak 11 Mart 2020 yılında ilan edildi. Koronavirüsleri, Orthocoronavirinae RNA virüslerinden bir alt aile olmakla birlikte Coronaviridae ailesine and Nidovirales üst kimliğine mensuplardır. Dünya çapında SARS-CoV-2, 308 milyon insanı enfekte etmiş ve 5,49 milyon ölüme yol açarak %1,79 ölüm oranıyla sonuçlanmıştır ki hala daha etkisini göstermeye devam etmektedir.

SARS-CoV-2'nin dört ana yapısal ve görevsel proteini, viral zarf (E), spike (S), nükleokapsid (N) ve matrix/zar (M) proteinleridir. Genomu, N proteini tarafından muhafaza edilmektedir. Ana yapısal proteini olan M proteini virüsün zarf adı verilen koruyucu dış duvarını oluşturmakta olan bir transmembran proteinidir. Diğer proteinlerle etkileşerek virüsün oluşmasını, tomurcuklanmasını ve enfeksiyon sonrası hücrelerden salınarak çoğalmasını kontrol eder. Lipid-duvarı üzerine gömülü olan E proteinleri de üzerindeki iyon kanalları yoluyla virüsün oluşumunu ve pathogenezi kontrol etmektedir. Sahip olduğu pozitif (+) tek zincirli RNA genomu, konak hücrenin RNA yapısına benzerliğinden ötürü konak hücrenin transkripsiyon ve translasyon mekanizmasını kullanarak kendi replikaz geni üretir. Üretilen replikaz geninin kodladığı iki protein, kesilerek yapısal olmayan proteinleri (NSP) oluşturmaktadır. Bu proteinler, konak hücrenin bağışıklık mekanizmasının baskılanmasında, virüs RNA'sının replikasyonunda, transkripsiyonunda ve konak hücreden çıkışında görev almaktadır. S proteinleri, homotrimer yapıda olan tip I hücre zarı glikoproteinleridir. S1 ve S2 olmak üzere iki ana kısımdan oluşmaktadır. Üzerlerinde bulunan 22 adet N-glikolizasyon bölgesi, virüsün yapısının korunması, konak hücreyi enfeksiyonu ve S proteinin uygun bir şekilde katlanması için gereklidir.

S1 alt kısmında bulunan reseptör bağlanma bölgesi (RBD), iki ana yapısal konformasyona sahiptir. Bunlar aktif olan yukarı ve inaktif olan aşağı konformasyonlarıdır. Reseptöre erişebilir olan yukarı konformasyonu, konak hücrenin ACE2 reseptörünün tanıyarak bağlanır. Bu bağlanma, S1 biriminin virüsün konak hücreye tutunmasını katalize eder. Konak TMPRSS2 gibi hücre peptidazları, S proteinini iki alt birimi olan S1 ve S2'ye ayırır. Bu olay, S2 biriminin aktifleşerek virüsün konak hücre zarı ile füzyonunu ve dolayısıyla hücreye girişini sağlar. NSP proteinlerinden Nsp16, virüs RNA başını sahip olduğu 2'O metiltransferaz aktivitesi aracılığı ile sadenozilmetioninden (SAM) aldığı metil grubu ile metile ederek replikasyonunu ve konak bağışıklık cevabından kaçışını sağlar. Nsp16 aktivitesi için kofaktörü olan Nsp10 ile etkileşimi gereklidir.

Moleküler yanaştırma, bilgisayar yöntemleriyle yapılan simülasyon metotları olmakla birlikte bu yöntemlerde ligand adı verilen küçük moleküller, protein reseptörler gibi hedef makromoleküllere bağlatılır. Ligand kütüphaneleri, moleküler yanaştırma araçlarıyla hedef bölgeye uygunlukları için taranır. Taranan aday ligandlar, bağlanma enerjilerine göre puanlanarak sıralanır. Daha negatif bağlanma enerjisine sahip olan ligandlar, daha stabil ligand-protein yapılarının oluşumuna izin verdiklerinden hedef proteine karşı aday küçük inhibitör moleküller olarak düşünülebilir. SARS-CoV-2 içinde son zamanlar birçok sayıda moleküler yanaştırma çalışmaları önemli hedef proteinleri (S protein RBD, ACE2, NSP proteinleri) üzerinde uygulanarak potansiyel inhibitörler keşfedilmiştir.

Bu tezde sunulan çalışmalar, SARS-CoV-2'ye karşı inhibitör olarak kullanılabilecek aday küçük moleküllerin tahmini için yapılmıştır. ZINC data bankasından alınan 400.000 ligand; 'yukarı' konformasyonundaki aktif ve tamamiyle glikolize edilmiş S proteinin RBD bölgesine (PDB: 6VSB), insan ACE2 proteinine (PDB: 6VW1) ve Nsp16 proteinine (6W4H) karşı taranmıştır. Küçük molekül ligandların, Spike RBD ve ACE2 etkileşim bölgesine karşı taranıp yanaştırılması, virüsün hücreye girişini engellemek için yapılmıştır. İkincil olarak yapılan Nsp16 proteininin NSP10 aktifleşme bölgesine ve SAM bağlanma cebine ayrı ayrı yapılan yanaştırma hesaplamaları, virüs genomunun konak hücre içinde replikasyonunu önlemek amacıyla yapılmıştır.

Ligandlar ve hedef proteinlerin 3 boyutlu yapıları polar hidrojen atomları da eklenerek AutoDock Tools programında hazırlanmıştır. AutoDock Vina kullanılarak, ligandlar S Protein RBD-ACE2, Nsp16-Nsp10 ve Nsp16-SAM bağlanma arayüzlerine gelecek şekilde hedef proteinlere yanaştırılmıştır. Yanaştırılmanın yapıldığı bağlanma kutuları, AutoDock Tools ile belirlenmiş ve bağlanma enerjileri AutoDock Vina aracılığı ile Python üzerinden hesaplanmıştır. Bu metot, 400.000 ligand arasından en iyi bağlanma enerjisine sahip 5.000 ligand bulunana kadar tekrarlanmıştır. Bu 5.000 ligand ve hedef proteinler, Schrödinger Maestro'nun LigPrep ve Protein Preparation Wizard modülleri ile bir sonraki bağlanma hesabı için uygun hale getirilmiştir. Daha sonra, Schrödinger Glide XP moleküler yanaştırma programı ile yüksek tutarlılıklı yanaştırma hesaplamalarına tabi tutulmuşlardır ve en iyi bağlanma enerjisine sahip 200 ligand belirlenmiştir. Belirlenen ligandlar, CANVAS programı ile dendritil parmakizi yöntemi ve hiyerarşik kümülendirilme ile kümülendirilmiştir. Her kümeye ait en düşük durum cezasına ve en yüksek bağlanma skoruna sahip moleküller, ait oldukları kümenin temsilcisi olarak seçilerek ileride deneylerde kullanılacak ilaç çeşitliliği artırılması hedeflenmiştir. Bulunan ligandlar, ADME/T analizleriyle farmasötikal özellikleri için taranmıştır. Ayrıca, bu ligandlar moleküller arası etkileşimleri açısından incelenerek, bağlandıkları hedef proteinleri arasında hidrojen bağları ve hidrofobik etkileşimler yapıp yapmadığı belirlenmiştir. Son olarak, en yüksek ilaç benzerliği özelliklerine sahip olan ligandların bağlanma enerjileri, en son olarak Schrödinger Desmond programında OSPL2005 kuvvet alanı ve TIP3P su modeli kullanılarak hesaplanmıştır.

AutoDock Vina yanaştırma hesaplamaları, seçilen ligandlar için hedef bölgelerin inhibasyonu için yüksek bağlanma skorları göstermiştir. Ek olarak, iki basamaklı tekrarlanan AutoDock Vina bağlanma hesapları sonuçların tutarlılığını artırmıştır çünkü aynı iki farklı hesaplama arasında ligandlar için benzer bağlanma skorları elde edilmiştir. Schrödinger Suite Glide XP yanaştırma hesaplamaları daha farklı bir aralıkta ve biraz daha düşük olmak üzere bağlanma skorları göstermesine rağmen, bulunan ligandlar Glide XP'nin kendi hesaplama fonksiyonuna standartlarında uygun ve iyi kabul edilebilecek düzeyde optimum bağlanma skorları vermiştir. En iyi

bağlanma enerjisine sahip olanların, AutoDock bağlanma hesaplamalarında bulunanlar ile benzerlik gösterdiği de belirlenmiştir. Bu bağlanma hesaplarının tutarlılığını göstermekte olup en iyi skora sahip olan 200 ligandın, potansiyel olarak inhibitör molekül olarak kullanılabileceğine işaret etmektedir. Kümelenendirme işlemleri de bulunan yüksek bağlanma skorlu ligandların, deneylerde çeşitlilik artıracak düzeyde olmasını sağlayacak sonuçlar vermiştir. ADME/T analizleri yüksek bağlanma enerjisine sahip en yüksek skorlu ligandlar için optimum aralıklarda farmasötik değerler vererek, bunların ilaç olarak kullanılabileceği uygunluğunu göstermiştir. Bir sonraki ligand etkileşim analizleri, en yüksek değerli ligandların kullanıldığı takdirde hedeflendikleri bölgeyi bloke edebileceklerini göstermiştir; çünkü ligandlar hedeflendikleri bölgede simülasyonlarda yüksek sayıda hidrojen bağı ve hidrofobik etkileşim oluşturmuştur. Son olarak, moleküler dinamik simülasyonlarında, simülasyon boyunca ligand-hedef protein komplekslerinin stabilitelerinin korunduğu gözlemlenmiştir ki bu, ligandların bağlanma bölgelerinden çıkmadan pozisyonlarını koruyabileceğini öne sürmektedir. Bu çalışmada tahmin edilen küçük-molekül inhibitörler, SARS-CoV-2'ye karşı potansiyel bir savunma ve tedavi aracı olarak ileride kullanılmak üzere laboratuvar deneylerinde test edilebilir.



1. INTRODUCTION

1.1. Coronaviruses and SARS-COV-2

In 2019, a virus with an unknown origin was assumed to be first emerged in Wuhan, China. The outbreak started locally but rapidly spread to the entire world. The virus has a rapid transmission rate given its multiple ways of infection such as air, respiratory droplets, contact and saliva. Due to its evolutionary similarities, the virus was identified as a type of severe acute respiratory syndrome (SARS) coronavirus and named 2019 novel coronavirus (2019-nCoV) which causes coronavirus disease 2019 (COVID-19). The International Committee of Taxonomy of Viruses gave its novel name as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (World Health Organization, 2020). After the eventual global spread of SARS-CoV-2, the World Health Organization declared the global pandemic on 11 March 2020.

Coronaviruses are a subfamily of RNA viruses (Orthocoronavirinae) which belongs to family of Coronaviridae and order of Nidovirales (Banerjee et al., 2019). These enveloped viruses have helical nucleocapsids. Their RNA genome is single-stranded and positive sense (+) (Coleman and Frieman, 2014). This subfamily of coronaviruses consists of four main genera which are alpha, beta, gamma and delta coronaviruses. During their evolutionary history, alpha and beta species originated from bats and rodents (mice, rat, etc.) while delta and gamma species originated from avian species (Woo et al., 2010). Given their highly infectious nature, they can cause inter-species/cross-species infections among various species such as bats, humans, birds, cats and camels. They can cause a wide range of respiratory tract infections from common cold to bronchitis and pneumonia as seen in lethally effective SARS, MERS and COVID-19 (Banerjee et al., 2019).

There are seven strains of coronaviruses that can infect humans. Four of them generally cause mild symptoms. These are HCoV-OC43, HCoV-229E, HCoV-HKU1 and HCoV-NL63 (Lim et al., 2016). Other three coronaviruses resulted in three major outbreaks which are SARS-CoV, MERS and COVID-19 (SARS-CoV-2). Severe acute respiratory syndrome coronavirus (SARS-CoV) outbreak occurred in 2002 and led to

the mortality rate of ~9%. SARS-CoV caused pneumonia and additionally acute respiratory distress syndrome, leading to 8096 cases and 774 deaths. MERS-CoV (Middle East respiratory syndrome coronavirus) emerged in 2012 in Arabian Peninsula. It led to 2494 cases and the deaths of 858 people with the mortality rate of ~35% by causing similar lethal diseases such as acute respiratory distress syndrome (Al-Qahtani, 2020). The last and current major outbreak is due to more contagious SARS-CoV-2 which was first reported in 2019 in Wuhan, China. Worldwide. As of today, it has infected 308 million people and caused 5.49 million deaths, resulting in 1.79% mortality rate (World Health Organization, 2022). Table 1.1 shows the features of three major outbreaks. As a result of phylogenetic and genotyping studies, the virus was associated with 2002 SARS-CoV as the evolutionary successor and its zoonotic origin was attributed to bats. Current medical research including vaccine and antiviral agent/drug development heavily focuses on its diagnosis and treatment.

Table 1.1: Infographic of Three Major Coronavirus Outbreaks.

SARS-CoV	MERS-CoV	SARS-CoV-2
Guangdong, China	Saudi Arabia	Wuhan, China
November 2002	June 2012	December 2019
8096 Cases	2494 Cases	308 million Cases
774 Deaths	858 Deaths	5.9 million Deaths
~9% Mortality Rate	~35% Mortality Rate	~1.79% Mortality Rate
Bat Origin	Bat Origin	Bat Origin
Cat Intermediate	Camel Intermediate	Unknown Intermediate
Angiotensin Converting Enzyme 2 (ACE2) Receptor as the Target Protein	Dipeptidyl Peptidase 4 (DDP4) as the Target Protein	Angiotensin Converting Enzyme 2 (ACE2) Receptor as the Target Protein

1.2. Structure of SARS-CoV-2

The structure of SARS-CoV-2 is highly similar to other members of coronavirus family. They are spherical viruses with the molecular mass of 40,000 kDa. Their size is in the range from 50 to 200 nm in diameter (Lalchhandama, 2020). Similar to all coronaviruses, they consist of four major structural and functional proteins which are envelope (E), spike (S), nucleocapsid (N) and matrix/membrane (M) proteins, (Santos

et al., 2020) as shown in Figure 1.1 The viral envelope is formed of a lipid bilayer containing anchored and embedded E, M and S proteins. Its length is 85 nm in diameter. On the other hand, the genome is protected by N protein.

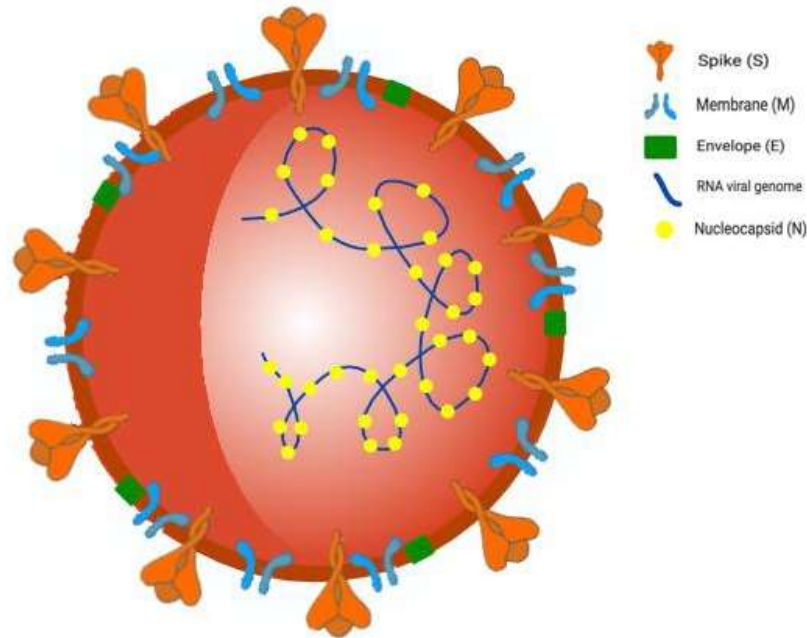


Figure 1.1: Schematic representation of SARS-CoV-2 structure.

Main structural protein of envelope is M protein. It is a transmembrane protein that interacts with other proteins. M protein mainly regulates viral assembly, budding and release after infection. M protein has ectodomain in its N-terminal and endodomain in its C-terminal along with a membrane-spanning part. E proteins are lipid bilayer-embedded/anchored proteins with external C-terminal domain. E protein also promotes assembly and pathogenesis via its ion channels. Together, M and E proteins give the virus its shape and maintain its integrity (Kuo et al., 2016).

SARS-CoV-2 has single-stranded positive-sense (+) RNA genome with the approximate length of 33 kilobases (kb). The viral RNA genome has 5' cap and 3' poly(A) tail (polyadenylated tail) (Wang et al., 2020). Due to its similarity to mRNA, it directs host transcriptory and translatory machinery to produce its replicase gene which encodes for two proteins which will be cleaved into non-structural proteins called NSPs. NSPs function in host immune response suppression, viral RNA replication-transcription and the host cell exit (Yadav, 2021).

S proteins are type I membrane glycoproteins, spanning a single protruding transmembrane domain towards extracellular part. They consist of two functional

domains, which are S1 and S2. S1 subunit catalyzes viral attachment to the host cell while S2 subunit promotes its fusion with the host membrane. These S proteins which are present on the surface of envelope give the virus a halo/crown-like shape (Wang et al., 2020).

1.3. Spike Protein and Receptor Binding Domain

The most unique feature of SARS-CoV-2 is its spike (S) proteins on the viral envelope. S proteins are homotrimeric transmembrane glycoproteins with the length of 20 nm. They consist of two main domains which are N-terminal S1 and C-terminal S2 subunits. S proteins mediate the attachment of SARS-CoV-2 to the surface of host cell membrane and initiates the viral infection via the function and regulation of these two subunits. S1 subunit is the head of S protein, and contains the N-terminal domain (NTD), the receptor-binding domain (RBD), subdomain 1 and 2. Subdomain 1 and 2 protects fusion peptide contained within S2 subunit. S2 subunit allows S protein to be embedded in the envelope (Henderson et al., 2020).

There are 22 N-glycosylation sites on the surface of S protein. These glycans are produced and attached to the virion during the replication inside the host cell. They mainly stabilize the virus and maintain its structural integrity. Additionally, they are required for viral interaction with the host cell receptor, ACE-2 and promotes the proper folding of spike protein. Since S protein has a high density of these N-glycans, it can also easily evade immune cells and their epitopes (Ramirez et al., 2021).

RBD of S1 subunit has two structural conformations which are ‘down’ conformation (inactive state) and ‘up’ conformation (active state) as shown in Figure 1.2. S protein RBD is only accessible to the host receptor in its up conformation. During the viral infection, S1 facilitates the binding of viral particle to the host surface through RBD-host cell receptor interaction. ‘Up conformation’ RBD of S1 subunit recognizes human angiotensin converting enzyme 2 (ACE2) membrane receptor as the target receptor and binds to it with high affinity as shown (Henderson et al., 2020) (Yan et al., 2021) in Figure 1.3. After initial attachment, host cell peptidases (e.g. peptidase domain of ACE2, transmembrane protease serine 2 [TMPRSS2]) cleave S protein into its two subunits.

This is followed by the activation of S2 subunit, and thus the fusion of SARS-CoV-2 and host cell membrane. In down conformation, it can mask itself from the immune system (Yan et al., 2021).

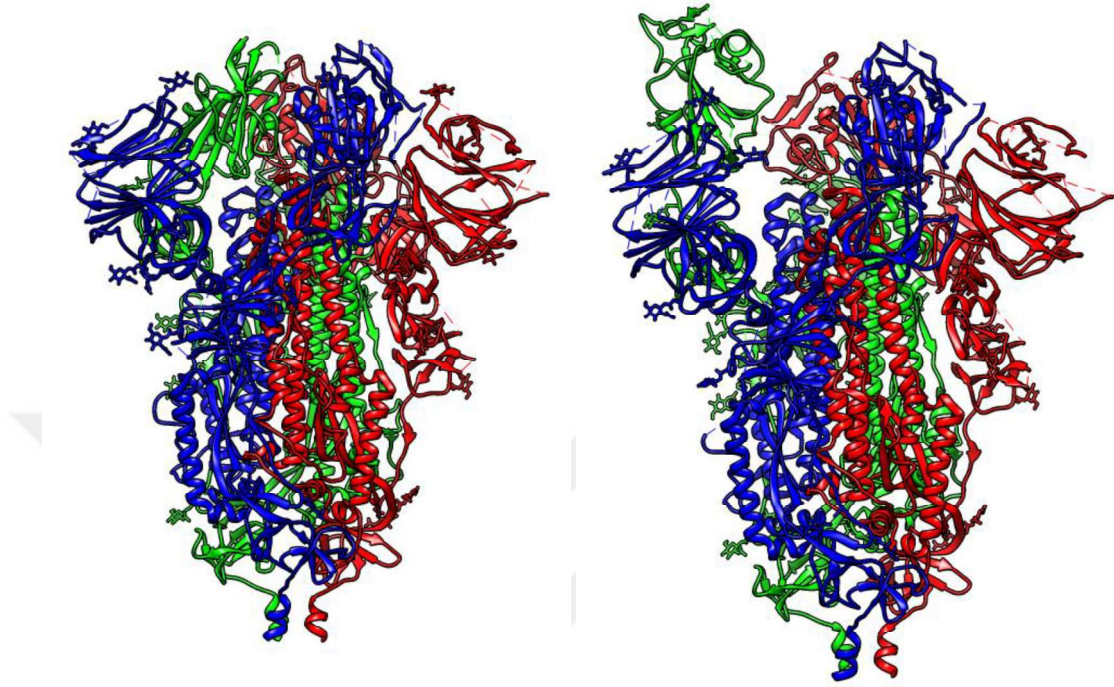


Figure 1.2: ‘Down/inactive’ (left) and ‘up/active’ (right) states of SARS-CoV-2 structures (PDB IDs: 6VXX [left] and 6VYB [right]). Color-coded top green part represents RBD on its down and up state respectively.

The RBD and its interactions with ACE2 have been studied via many cryo-electron microscopy (cryo-EM) and X-ray crystallography methods to generate 3D structures. These structures are stored in the Protein Data Bank (PDB) such as 6VSB, 6M0J and 6VXX. Moreover, recent cryo-EM researches have demonstrated that SARS-CoV-2 binds to human ACE2 with 10-20 times higher affinity compared to 2002 SARS-CoV resulting in more potent infection rate (Nguyen et al., 2020). This suggests that SARS-CoV-2 is more infectious than its ancestor and Spike RBD is almost the most critical part for its interaction with the host receptor.

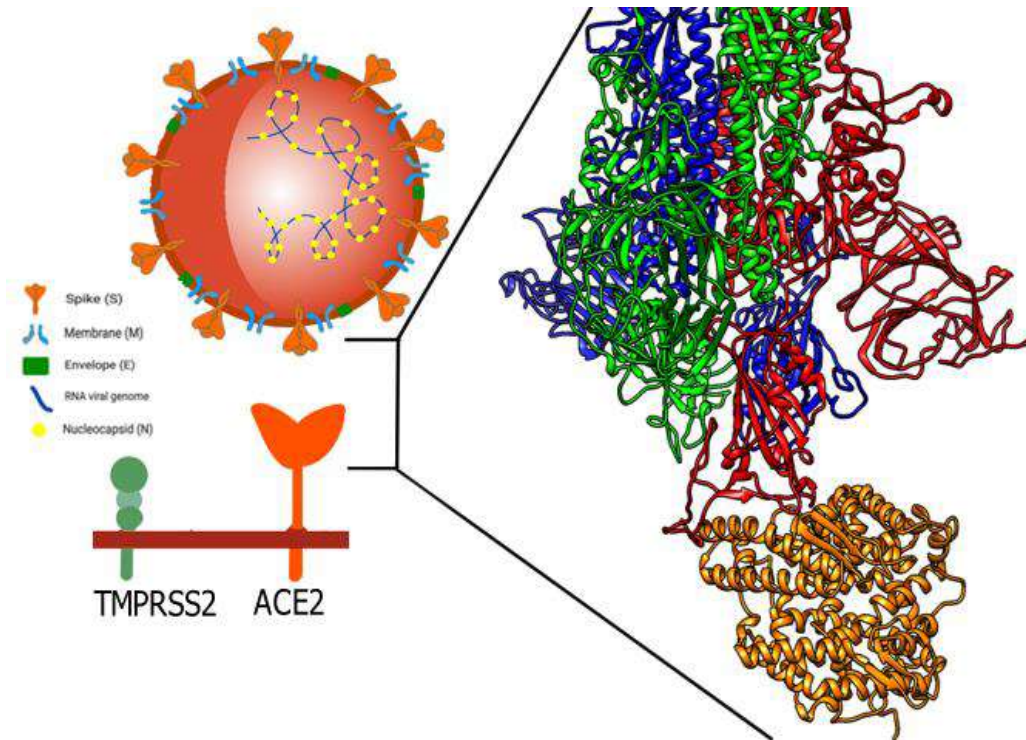


Figure 1.3: The interaction of SARS-Cov-2 in its ‘up’/active conformation (PDB ID: 6VSB, red/green/blue homotrimer) with human ACE2 receptor (PDB ID: 6VW1, orange).

1.4. Infection Mechanism of SARS-CoV-2

The infection mechanism is initiated by binding of viral particles to host surface cellular receptors. SARS-CoV-2 S proteins bind to human angiotensin converting enzyme 2 (ACE2) with high affinity to initiate the cell entry. After viral attachment to the host cell surface, transmembrane protease serine 2 (TMPRSS2) causes priming of S protein leading to exposure of pre-attachment preserved fusion peptide, S2 subunit to ACE2 receptor. This event promotes the cell entry through direct fusion of two membranes or endocytosis which forms endosome around viral particle (Hoffmann et al., 2020) (Yan et al., 2021). The viral particle escapes upon pH drop from endosome. After its escape, the viral particle sheds its coating and releases its (+) single stranded RNA into the cell (Petersen et al., 2020). Since the viral RNA has a similar structure to mRNA due to the presence of 5' methyl cap and 3' poly(A) tail, it can use host translational mechanism. The host ribosomes translate open reading frames of coronavirus RNA (ORF1a and ORF1b) into polypeptides pp1a and pp1ab. These polypeptides are differentially cleaved by papain like protease (Plpro, Nsp3) and the main protease, 3CL protease (3Clpro, Nsp5), resulting in the formation of non-

structural proteins (NSP) (Fehr and Perlman, 2015). Many of these NSPs form replicase transcriptase complex (RTC) which is an RNA-dependent RNA polymerase. This multi-protein complex replicates and transcriptase nascent RNAs from viral RNA to be translated by ribosomes to produce viral accessory proteins (Sexton et al., 2016). The coronavirus accessory proteins S, M and E are synthesized in endoplasmic reticulum and are assembled within Golgi by the main role of M protein. This assembly is followed by nucleocapsid formation around viral RNA to protect and stabilize it. This is followed by the transport of viral RNA genome to viral envelope where S proteins are embedded/anchored. The matured virion is then released from the host cell. This entire SARS-CoV-2 infection-maturation-release cycle is essential for SARS-CoV-2 spread (Cevik et al., 2020).

1.5. Nsp16-Nsp10 Interaction Mechanism in SARS-CoV-2

The cleavage of polypeptides pp1a and pp1ab, which are encoded from ORF1a and ORF1b region of SARS-CoV-2, by their associated proteases (Nsp3 and Nsp5, respectively) results in the formation of non-structural proteins (NPSs) which have various functions for SARS-CoV-2 such as intracellular host cell signaling suppression, host translation blockage, viral RNA replication, providing virulence factor and uridylyte-specific endoribonuclease activity (Raj, 2021). Among these non-structural proteins, two of them have critical importance for viral mRNA capping which is utilized to facilitate viral replication and mask viral RNA against host immune system and prevent recognition by immune toll-like receptors 7 and 8. This is achieved by non-structural protein 16 (Nsp16) by methylating viral RNA cap. For the proper activity of Nsp16, it requires non-structural protein 10 (Nsp10) which is the cofactor of Nsp16 2'O-methyltransferase activity. During the viral mRNA maturation, Nsp10 and Nsp16 form a heterodimer. Activated Nsp16 of this complex transfers a methyl group from S-adenosylmethionine (SAM) to 2' hydroxyl of ribose sugar in the viral mRNA. The unique zinc binding protein of nsp10 binds nonspecifically to RNA and stabilizes the SAM binding pocket in Nsp16 to form a stable complex (Rossa-Lemus et al., 2020) (Morales et al., 2020). The genomic region of SARS-CoV-2 encoding for Nsp10 and Nsp16 is shown in their 3-D structures along with with SAM in Figure 1.4.

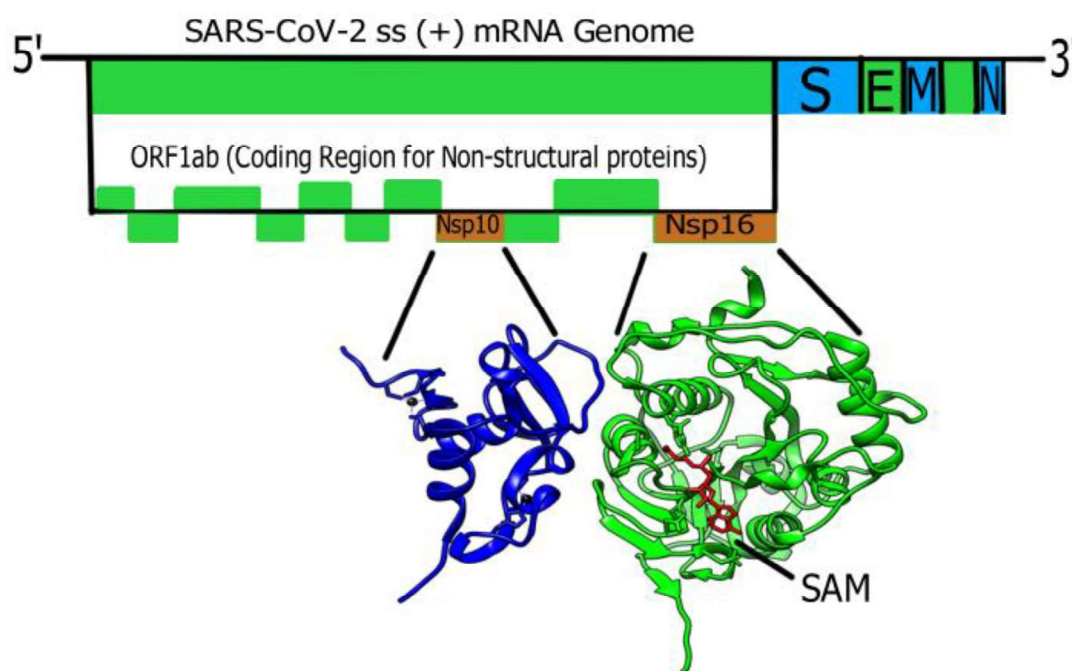


Figure 1.4: The ss (+) mRNA genome (top) and the 3-D structures of Nsp10-Nsp16 complex and SAM.

1.6. Molecular Modeling

Molecular modeling is the combination of computational approaches to simulate and observe the interactions between molecules. Among these approaches, molecular docking is a computational simulation method in which a ligand (e.g. small molecule, small peptide, nucleic acids, etc.) is bound to a target in a predefined binding pocket to generate a stable complex (Lengauer and Rarey, 1996). In the context of molecular modeling, a ligand is a molecule that can bind to a macromolecule and trigger a biochemical response. These target macromolecules are generally target protein receptors. There are two types of ligands which are response-triggering agonists and receptor-blocking antagonists (Berg and Clarke, 2018).

There are two major molecular docking approaches to deduce the information from resulting interaction between a ligand and a target molecule. The first approach is ligand-based in which there is a ligand with known structure and functions for its target protein. In this method, structurally or biochemically similar ligands are selected or designed to test possible interactions with target protein that can surpass the potency and affinity of the original ligand. Structural similarity is quantified to assess the ligand candidates with respect to each other (Sliwoski et al., 2014). For this, structural scoring

methods are used such as root-mean-square deviation (RMSD) which is the measure of average distance between the atoms of superimposed proteins or small molecules (Carugo and Pongor, 2018). On the other hand, chemical properties (e.g. absorption, distribution, metabolism and excretion values) are assessed to predict the effect of screened ligands on target protein with respect to starting ligand (Vrbanac and Slauter, 2017).

The second method is structure-based in which the structure of target protein is produced and predetermined by laboratory procedures such as cryo-EM, nuclear magnetic resonance (NMR) and x-ray crystallography (Salmaso and Moro, 2018). Generated 3-D structures of target proteins are generally stored in a database called Protein Data Bank which is freely accessible web-based database for 3-D structures of large molecules (Berman et al., 2000). In structure-based method, a target site on the protein is identified as a possible ligand binding site. These sites are generally active/catalytic sites of the protein, and present on the surfaces or pockets, having the possibility to generate strong hydrogen bond or hydrophobic interactions. After the identification of the candidate target site, a library of ligands is virtually screened and docked onto the target site (Salmaso and Moro, 2018). These small molecules are taken from databases like ZINC (Irwin and Shoichet, 2005). Search algorithm is used to search conformational space for docking on proteins by identifying and analyzing all possible orientations of ligand on the active site. Scanned candidate ligands are scored and ranked based on the strength of their bond-interactions with the target site, which is called binding affinity. The binding affinity is closely associated with the equilibrium dissociation constant (KD) between ligand and protein (Pantsar and Poso, 1899). KD is calculated with following equation (1.1).

$$KD = [\text{molar concentration of ligand}] \times \frac{[\text{molar concentration of protein}]}{[\text{molar concentration of complex}]} \quad (1.1)$$

KD is inversely related to binding affinity. KD represents the amount of ligand required to effectively bind to target protein and thus, lower KD value indicates higher binding affinity. Ligands with higher binding affinity are considered as possible binding molecules for target protein. During molecular docking simulations, ligands are allowed to find its optimum position on the active site of target protein after serial transitions including bond rotations, torsional angle rotations, residue translations.

During these transitions, the total energy of the system actively changes and it is calculated during each transition. The optimum binding pose can be inferred through the minimization of energy and flexibility of ligands. Molecular docking tools generate a large number of poses for ligands in different orientations via minimization of free energy of the complex (Ferreira et al., 2015). Energy minimization is performed to reduce the potential energy of the bound complex. Biological systems have lower potential energies and are highly dynamic to spontaneously interact with other molecules so the minimization enables the generation of conformations that promote such dynamics, closer to the ones observed in real-life biological systems. Resulting poses are ranked with a scoring function to identify the ones with more negative binding energy and the highest binding affinity. The ligands with more negative binding energy score result in the generation of more stable bound complexes (Kastritis and Bonvin, 2013).

The scoring function, which is the mathematical functions based on free-energy calculations, then uses poses as input and returns a numeric score which demonstrates the odds for most favorable binding interactions. It additionally ranks all ligands based on their docking score and binding affinity to target protein relative to each other. Many scoring functions use force fields which are computational methods to estimate the forces between atoms within and between molecules and parameters to calculate the potential energy of system (e.g. van der Waals, electrostatic interactions and torsional forces) (Li et al., 2019) (Torres et al., 2019). Another significant factor in these methods is the confirmation of docking accuracy. It is measured by evaluating the fidelity of docking program and comparing prediction of resulting right pose of ligand to experimentally observed one (Wang et al., 2016).

There are two bases for molecular docking algorithms which are rigid-body and flexible docking. In rigid-body docking, ligand flexibility is not considered and a large number of ligands are directly docked onto determined target site. This method is fast but inaccurate. On the other hand, flexible docking uses a larger search space and analyze all possible conformations with all possible rotations. This leads to intense but highly accurate docking (Andrusier et al., 2008).

Most molecular docking tools consider ligands flexible and receptors rigid body. Two of these tools are AutoDock and Schrödinger Suite Maestro. AutoDock is a molecular modeling simulation software which allows ligands to flexibly find their optimum

positions on a rigid protein receptor. Search space is specified in the form of coordinate system of the receptor, as a grid box in which many ligand positions are considered. Randomized ligands through genetic algorithm, in which different poses are generated from multiple pairings of ligands based on Darwinian natural selection theory, are docked onto given search space box. Independent number of docking runs are performed and after each step, ligand optimizations are applied by torsion, position and orientation rotations and analyses followed by scoring function. Conformation-based scoring is used through force-field calculations in which the strengths of intermolecular interactions are summed to find chemical potentials and binding affinities (Morris et al., 2009) (Trott and Olson, 2010).

Schrödinger Suite is the amalgamation of different softwares and modules for chemical uses such as visualization, docking and molecular dynamic simulations (Schrödinger, 2021). These softwares and modules are: a. Maestro: Molecular visualizer for chemical and biochemical compound (Schrödinger, 2021), b. Canvas: The similarity analysis between molecules based on the fingerprints which represent the existence of 2-D core structural fragments of a molecules by screening similarities between pairs of these fragments (Duan et al., 2010), c. Desmond: molecular dynamic simulation tool (Bowers et al., 2006), d. Epik: the tool for pKa (acid disassociation constant) and tautomeric state (isomer states with protein relocation and ionization) prediction (Shelley et al., 2007), e. LigPrep: the tool for ligand preparation in molecular docking (Schrödinger, 2021), f. Protein Preparation Wizard: the tool for protein preparation in molecular docking (Sastry et al., 2013), g. Glide: Molecular docking tool for standard and high-precision docking (Friesner et al., 2004), h. QikProp: the tool for ADME/T analysis (Schrödinger, 2021). The entire molecular docking proceses is summarized in Figure 1.5. After molecular docking calculations, molecular dynamics (MD) simulations can be done to observe physical movements of molecules and atoms over a predefined time period. The atom and molecule trajectories are analyzed as the evolution of the system. The forces and velocities between molecules and atoms are calculated along with their potential energies using force fields and calculations based on Newton's equations of motion. For ligand-docked bound complexes, MD trajectories give insight about the displacement of ligand within the system so one can see whether a particular ligand can maintain its

interaction with the target binding site or not over a period of time. This information is used to identify whether the ligand is stable or not (Hollingsworth and Dror, 2018).

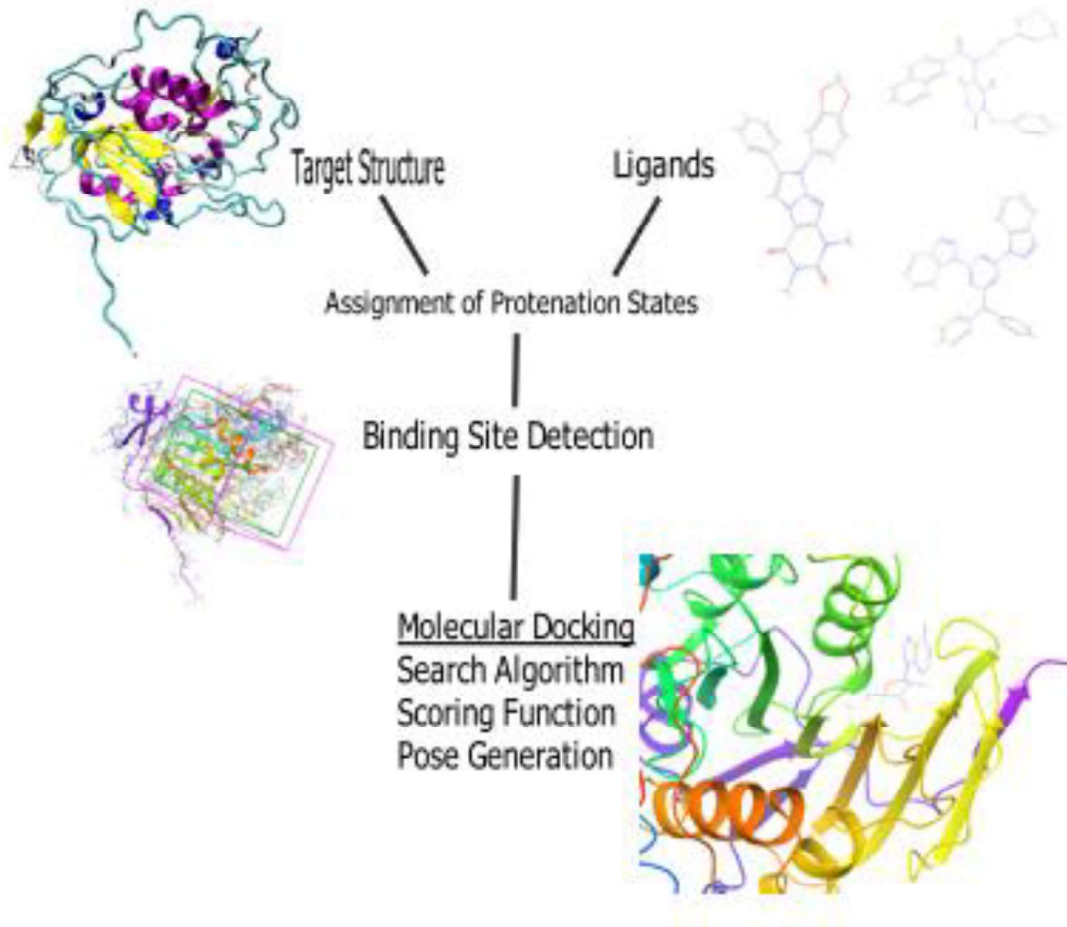


Figure 1.5: The summary of molecular docking process.

This is generated by RMSD over the trajectory time by the equation (1.2) where N is the number of selected atoms, t_{ref} is reference time (the first frame), r' is position of selected atoms in frame number x after superimposing on reference frame and frame x is the frame at time x (t_x). At each frame, RMSD is and thus, gives the changes in atom displacement for each frame compared to reference frame. If the changes in RMSD value for ligand and target protein generally fluctuates with less than 3 Å between, that can mean that ligand and protein are stable within the bound complex and the ligand retain its bound status within binding pocket (Guterres and Im, 2020).

$$RMSD_x = \sqrt{\frac{1}{N} \sum_{i=1}^N (r'_i(t_x)) - r_i(t_{ref}))^2} \quad (1.2)$$

1.7. Small-molecule Inhibitor and Drug Development

The process from the development of a drug to its entry to the market nearly takes 11-15 years and 2-3 billion dollars. Moreover, success rate of drug development procedure is only 2%. (Scannell et al., 2012). Drug discovery and development procedure consists of six steps which are screening for lead-compound identification (a compound which has biological or pharmacological activity for potential therapeutic usage and can be used as starting molecule for the design of drugs with proper chemical modifications), preclinical studies, applications for clinical studies, phase 1, 2 and 3 trials, applications for novel drug and PMS (post-marketing safety surveillance) (Xue et al., 2018) (Dhir et al., 2020).

One of the fastest and the most cost-effective approaches to shorten drug discovery process is development of small-molecule inhibitors or drugs through molecular dynamics analyses. Small-molecule ligands can be docked onto drug targets and screened via high-throughput docking simulations. Based on scoring of docked ligands, ligands with best binding affinities can be directly selected to be tested on laboratories. On the other hand, drugs can be also screened directly instead of ligands with unknown potential to find whether screened drugs can be used for new purposes against the target. This process is called drug repurposing (drug repositioning). It is a process in which new uses of existing drugs are discovered. Both techniques shorten the entire process of drug discovery which normally takes a decade from target discovery to clinical research step since market-available ligands or existing drugs could be already studied, have already established production and delivery chain in the market, and have proper amount of clinical data (data regarding toxicity, drug safety, adverse effects, administration, bioavailability, etc) (Dhir et al., 2020). Many drugs tested were also FDA-approved. There are many successful examples for drug development. For instance, a phosphodiesterase type 5 inhibitor Sildenafil which was produced by Pfizer in 1980 can be utilized in coronary artery disease treatment and hypertension (Shim and Liu, 2014).

Despite being cost- and time-efficient way of drug discovery, there are some challenges regarding these techniques such as the need of new clinical trials for proposed inhibitors or repurposed drug on different diseases, the availability of drug, possible need of new targeted delivery mechanism to body. Computational molecular

docking methods based on virtual screening overcomes these challenges by screening large number of ligands for target proteins and comparing their efficiencies for the optimum drug candidate.

1.8. Molecular Docking Research Against SARS-CoV-2

Although there have been vaccine development and a huge effort in the development of potent treatment for SARS-CoV-2, the emergency of pandemic, high transmissibility and mutagenesis rate of SARS-CoV-2 demands more rapid approaches for novel treatments and drug production. Computational studies which focuses on molecular docking, small inhibitory molecule discovery and drug repurposing have become major players in this need of rapid drug development.

Major key proteins in SARS coronavirus replication, maturation and infection have been considered as promising targets for drug development. As mentioned before, these targets are mainly RBD of spike protein which recognizes and binds to the host cell receptor, ACE2 which is the human host cell receptor, SARS-CoV-2 main protease and associated proteins which are necessary for the viral cell entry.

There are many in silico molecular docking and repurposing studies for SARS-CoV-2 with effective results. Following examples indicate the importance of molecular docking simulations in potential drug development of SARS-CoV-2: a. Structure-based virtual screening of existing drugs against ACE2 by AutoDock Vina where Lividomycin, Quisinostat, Spirofylline, Burixafor and Pretrexed showed high binding affinity for ACE2 (Teralı et al., 2020), b. The use of Schrödinger Suite Maestro and Glide molecular docking tool to identify Remdesivir, Indinavir, Zanamivir and Saquinavir as possible drugs against spike protein and main protease 3CL (Hall et al., 2020), c. The computational drug discovery study in which RBD, papain-like protease and main protease 3CL were targeted, showed favorable binding energies and high binding affinities for Simeprevir, Glecaprevir, Ledipasvir and Paritaprevir (Mahdian et al., 2020), d. AutoDock Vina docking study where Remdesivir, Favipiravir, Ritonavir and Ribavirin demonstrated high affinity against main protease and spike glycoprotein (Narkhede et al., 2020), e. Repurposing of a selective estrogen-receptor modulator, Toremifene against S protein and Nsp14 which could potentially block viral cell entry and viral replication respectively based on AutoDock Vina results (Martin and Cheng, 2020), f. AutoDock Vina and Schrödinger Glide XP docking

calculations of small compounds from 10 spices predicted small-molecule inhibitors against S protein and main protease and their drug-likeness were verified by ADME analysis (Sen et al., 2020), and g. Glide virtual screening and Lipinski's rule of five-based ADME/T analysis of small antiviral and antiprotease ligands against Nsp10/Nsp16 and main protease showed a great range of docking binding free energy and drug-likeness for ligands (Maurya et al., 2020).

1.9. Purpose of the Thesis

Since the emergence of SARS-CoV-2, there has been an urgent need for the development of new treatments, inhibitor molecules and drugs. The studies which were conducted for this thesis aim to predict possible small-molecule inhibitor candidates for the treatment of SARS-CoV-2. Many commercially available small ligands were virtually screened against three major targets which are receptor binding domain (RBD) of Spike protein, human ACE2 receptor and a non-structural protein Nsp16. The studies focused on the inactivation of two significant pathways in viral life cycle, the entry to host cells and cell replication. Firstly, the docking of small-molecule ligands on the Spike RBD and ACE2 interaction surface interface was done to provide candidate small inhibitory molecules/drugs to block the interaction between these two proteins. Such blockage can result in the prevention of viral cell entry. Additionally, these small molecules were docked on the Nsp16-Nsp10 interaction site to stop the full activation of Nsp16 by its cofactor Nsp10. Lastly, SAM binding pocket of Nsp16 was targeted to prevent its methylation activity. The aim of both docking calculations was to develop small-molecule inhibitors to prevent viral the cell replication. During the study, AutoDock Vina and Schrödinger Suite Maestro-Glide XP software were used to conduct high-precision docking calculations so docking results could represent accurate binding affinities (closer to experimental findings) for tested ligands against the target proteins. Moreover, fingerprinting clustering method was applied on ligands having high docking scores to eliminate similar ligands so this way, ligand variety was intended to increase for higher significance in results and possible future laboratory experiments. Additionally, the study aims to predict pharmaceutical properties of tested ligands by ADME/T experiments so it can be identified whether proposed ligands can be used as possible drugs or not. Furthermore, the interactions of ligands with their respective target binding sites were analyzed to see if there are larger number

of hydrogen bonds and hydrophobic interactions and such interactions are overlapping with the interactions of competitive molecules. Lastly, molecular dynamics simulations were done to see the possible ways in which tested inhibitors can behave when they are bound to their respective targets and to analyze if docked ligands retain their positions within their respective binding pockets so they can be properly used as long-term inhibitors. With the combination of docking calculations, structural analysis, drug-likeness studies and molecular dynamics simulations, this thesis was conducted to propose optimum ligands that can be used for SARS-CoV-2 inhibition as such ligands would have strong binding affinities, stability and drug-likeness.



2. MATERIALS AND METHOD

2.1. Selection of Target Proteins for Molecular Docking Simulations

Three major proteins were selected as the targets of molecular docking simulations which are SARS-CoV-2 Spike receptor binding domain (RBD), human angiotensin converting enzyme 2 (ACE2) receptor and non-structural protein 16 (Nsp16).

Spike protein was selected as the target protein as its RBD is the key receptor binding domain that recognizes human ACE2 receptor, allowing viral attachment on the host cell surface and following cell entry. The 3-D structure of Spike protein was selected as homotrimeric Spike in its prefusion ‘up’ (active) conformation which lack glycans on its glycosylation sites (PDB ID: 6VSB) (Figure 2.1) (Wrapp et al., 2020).

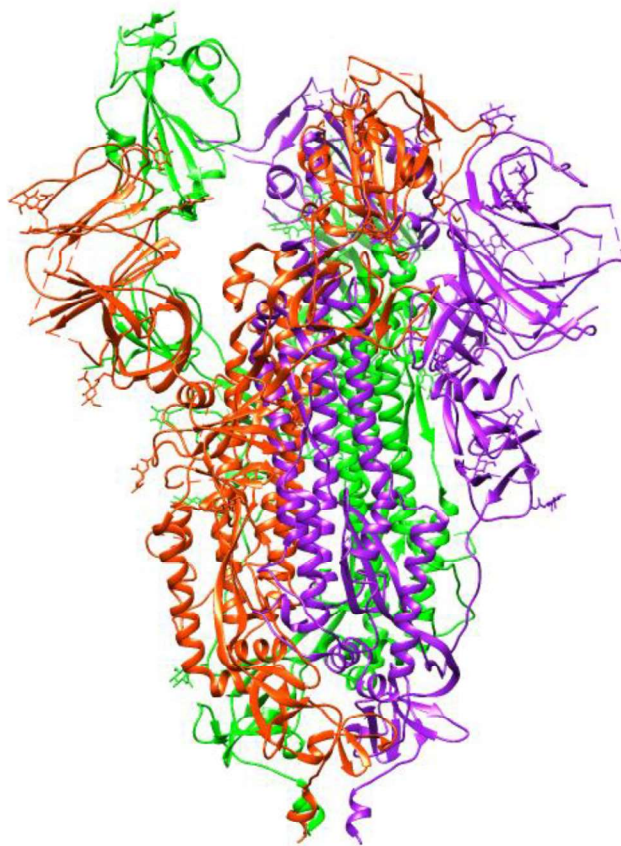


Figure 2.1: The 3-D structure of homotrimeric SARS-CoV-2 Spike protein in its ‘up’/active conformation (PDB ID: 6VSB), visualized in UCSF Chimera.

The closest form of Spike protein to its wild-type state is required for high accuracy ligand docking. N-glycosylation sites are also essential to maintain the structural integrity of virus, promote the interaction with ACE-2, regulate the proper folding and provide immune evasion and docking. For these two reasons, a fully glycosylated version of 6VSB was utilized. 6VSB_1_1_1 which is fully glycosylated form of 6VSB was taken from CHARMM-GUI Archive. It was produced via template-based and loop modeling methods in GALAXY modeling suite. CHARMM-GUI Glycan Reader&Modeler was used to add 22 N-glycans. Molecular dynamics simulations were performed to assess stability of the structure on a membrane built by CHARMM-GUI Membrane Builder (Jo et al., 2008) (Woo et al., 2020). In this method, a single monomer was taken among the entire structure by manually modifying PDB file (Figure 2.2).

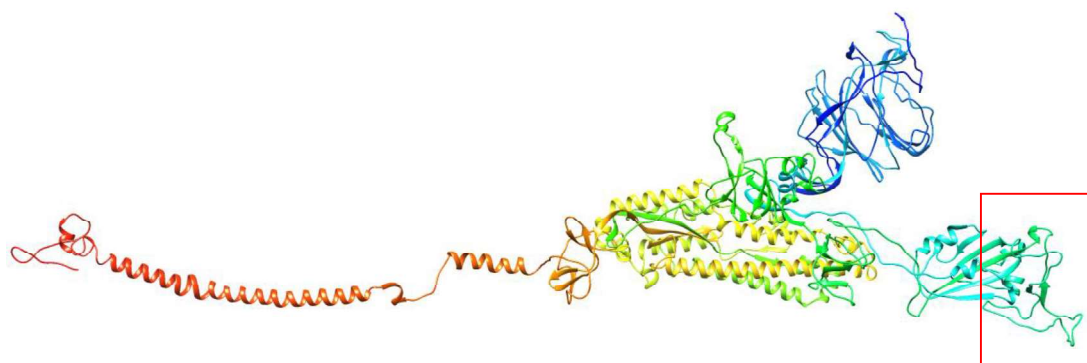


Figure 2.2: The 3-D structure of single monomer of fully glycosylated Spike protein (6VSB_1_1_1), visualized in UCSF Chimera. RBD is indicated by the red rectangular.

The second target structure was selected as human ACE2 receptor due to its aforementioned role in viral entry. For 3-D structure of ACE2 receptor, the one, which was experimentally bound to 6VSB_1_1_1 in CHARMM-GUI Archive, was taken to maintain the accuracy between docking studies (PDB ID: 6VW1) (Figure 2.3) (Shang et al., 2020) (Woo et al., 2020). This way, both docking calculations can consistently predict ligands for Spike RBD-ACE2 binding interface.

The critical amino-acid residues on both Spike protein RBD and ACE2 binding interface (Figure 2.4) were identified from the established literature through the binding research based on their ability to create strong hydrogen bonding and hydrophobic interactions between their surfaces (Veeramachaneni et al., 2020). These sites were used for the creation of docking space. They are listed in Table 2.1.

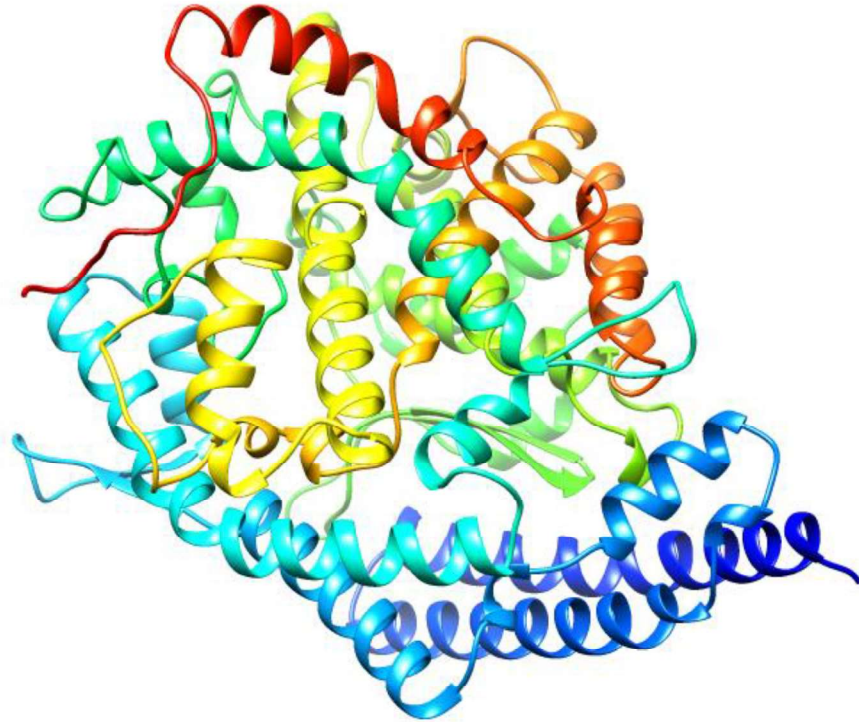


Figure 2.3: The 3-D Structure of human ACE2 (6VW1), visualized in UCSF Chimera. Dark blue parts are binding site of Spike RBD.

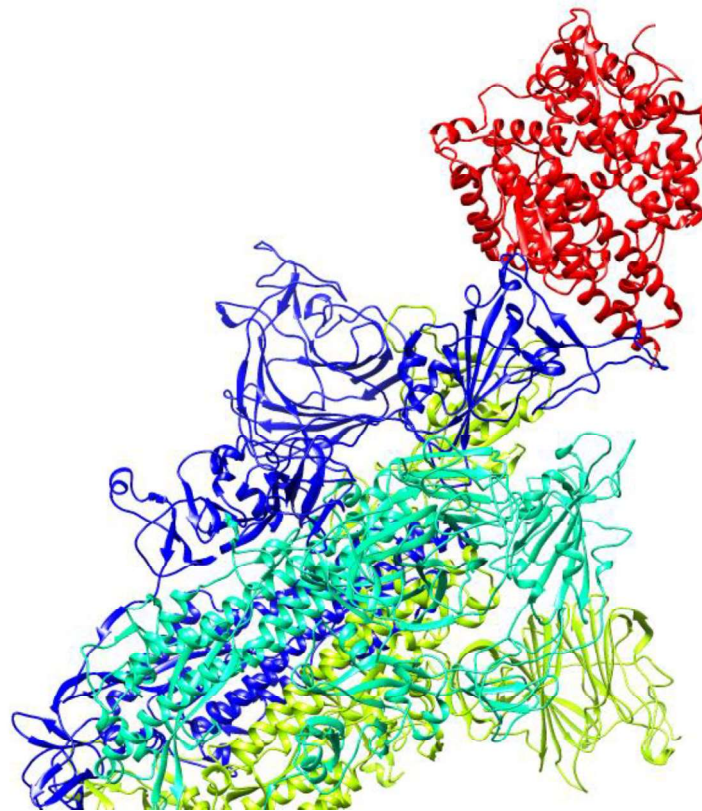
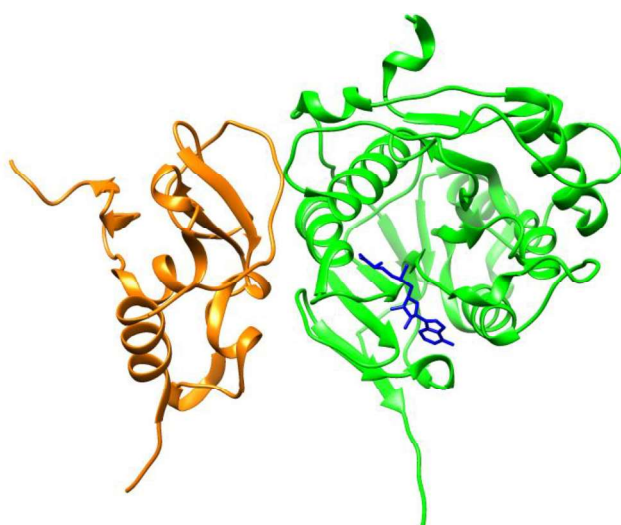


Figure 2.4: Spike RBD (multicolor homotrimer, PDB ID: 6VSB_1_1_1) and ACE2 (red, PDB ID: 6VW1) binding interface, visualized in UCSF Chimera.

Table 2.1: Binding profiles between SARS-CoV-2 RBD and ACE2

S Protein RDB	ACE2
Tyr 408	Ala 387
Lys 417	Asp 30, His 34
Gly 446	Gln 42
Tyr 449	Gln 24, Asp 38, Gln 42
Tyr 453	His 34
Asn 487	Gln 24, Tyr 83
Tyr 489	Gln 24, Thr 27, Tyr 83
Gln 493	Lys 31, His 34, Glu 35
Gln 498	Asp 38, Tyr 41, Gln 42, Lys 353
Thr 500	Tyr 41, Gln 42, Asn 330, Asp 355, Arg 357
Gly 502	Lys 353, Gly 354
Tyr 505	Glu 37, Lys 353, Gly 354, Arg 393

The last docking targets were selected as two binding sites on non-structural protein 16 (Nsp16) (PDB ID: 6W4H) (Figure 2.5). One of them binds to Nsp10 which is required for full Nsp16 activation for its methyltransferase catalytic activity. The other site is S-adenosylmethionine (SAM) binding pocket which is used for methyl transferase activity of Nsp16 as it is the methyl donor of Nsp16. Both sites were differentially targeted to provide small inhibitors that can be used to inactivate Nsp16 and thus viral replication (Rosas-Lemus, 2020). 6W4H only contains Nsp16.

**Figure 2.5:** The 3-D Structure of Nsp16 (green) and Nsp10 (orange) in complex with SAM (PDB ID: 6W4H), visualized in UCSF Chimera. Only Nsp16 was used.

2.2. Ligand Preparation for Molecular Docking

6,2 million ligands were selected among 540 tranches of ZINC Database. Selection criterias were based on their property (drug-like ligands are selected to get most drug properties and capabilities), reactivity (clean/non-reactive drugs are selected to prevent possible adverse effects and undesired metabolic activities), availability (only in-stock drugs were selected in case of their future experimental tests after docking simulations) (Figure 2.6). This ligand library was downloaded as 2D structures in SMILES format.

ZINCSubstancesCatalogsTranchesBiological +More +

Ref.2D3D

ResetClean

FiltersIn-Stock

pHNA

ChargeNA

Molecular Weight (up to, Daltons)

	200	250	300	325	350	375	400	425	450	500	>500	Totals, by LogP
-1	1,000	4,327	5,896	2,815	2,922	2,051	1,373	1,084	851	1,562	1,000	22,591
0	12,910	12,910	19,304	11,545	12,199	9,119	5,293	2,933	2,311	1,069	1,000	77,345
1	50,931	50,931	81,976	52,029	55,981	55,430	24,334	12,608	7,434	6,857	1,000	347,580
2	118,077	118,077	198,585	139,099	166,539	133,956	84,924	48,364	32,398	26,535	1,000	948,477
2.5	69,334	69,334	138,321	105,579	137,131	117,573	82,147	53,794	37,935	51,880	10,000	793,694
3	58,720	58,720	137,812	112,375	159,257	144,655	112,190	78,683	58,237	55,632	10,000	917,561
3.5	37,835	37,835	110,596	100,344	146,376	150,188	129,826	101,654	80,733	81,812	10,000	939,164
4	18,376	18,376	72,943	73,503	112,561	130,565	127,528	114,052	135,744	107,153	10,000	892,425
4.5	6,278	6,278	37,952	46,501	74,726	96,466	107,681	108,612	138,283	121,772	10,000	738,271
5	1,658	1,658	14,893	23,440	42,144	60,225	74,745	83,297	119,716	123,880	10,000	543,978
>5	0	0	0	0	0	0	0	0	0	0	0	0
Totals, by Weight	0	378,346	818,068	667,230	909,836	900,219	750,041	604,961	613,642	578,723	0	6,221,086 Substances 540 Tranches

Figure 2.6: Selection of ZINC library consisting of 6.2 million ligands.

After the initial library generation, 400.000 ligands were randomly selected from initial 6,2 million ligands by using randomization command in Windows PowerShell. Chosen ligands were converted to 2D files in SDF format from SMILES with OpenBabel. Then, these SDF files were converted to 3D ligand structures in the format of PDBQT with following OpenBabel command by executing Python bash command on Python sh files.

```
Obabel -i sdf ./directory-name/folder-name/2Dliganlistnumbers.sdf -gen3D -h -
DelNonPolarH -O ./target-directory-name/target-folder-name/ 3Dligandlistnumbers -
errorlevel 1 -m
```

This command generated 400.000 proper 3D ligand structures by sequentially generating 3D structure, adding hydrogen atoms and deleting non-polar hydrogen atoms.

2.3. Protein Preparation for Molecular Docking

Three target proteins which are a single monomer of Spike protein (PDB ID: 6VSB_1_1_1), human ACE2 (PDB ID: 6VW1) and Nsp16 (PDB ID: 6W4H) were loaded to AutoDock Tools (ADT) in PDB format. All non-amino acid molecules (excluding essential glycans of S protein) and water molecules were deleted from structures. Since PDB files lack hydrogen atoms, polar hydrogens atoms were added to each structure through following menu selection: “Edit-Hydrogens-add” and then “polar only”. Polar hydrogen-added PDB files were converted to PDBQT files (files containing partial charges and AutoDock atom types in addition to PDB data) by using the selection of “Grid-Macramolecule”.

2.4. Molecular Docking Simulations using AutoDock Vina

The binding site of human ACE2 on RBD of Spike protein were determined by choosing the RBD area covering critical amino-acid residues which make strong hydrogen bonds and hydrophobic interactions with human ACE2 as the center of docking space. These were selected as Tyr 408, Lys 417, Gly 446, Tyr 449, Tyr 453, Asn 487, Tyr 489, Gln 493, Gln 498, Thr 500, Gly 502 and Tyr 505. The center of docking space was determined by taking the area away from 5 Å of these amino acids and calculated by measure center command of Visual Molecular Dynamics Software (VMD). According to calculated region, docking space was generated using Grid Box function of ADT with following coordinates and sizes and written into configuration file of docking simulation. The center of docking box was selected as center_x = -34.3700, center_y = 8.5660 and center_z = 282.1210. The dimensions of docking box were set as size_x = 32, size_y = 32 and size_z = 22 (Fig 2.7). The exhaustiveness parameter was selected as 24 for docking calculations. This parameter was specifically selected to be high since it represents the amount of computational-effort in a docking calculation and a higher exhaustiveness value provides greter atom flexibility, consistent docking score and more stable-low energy resulting ligand poses. Docking calculations of 400.000 ligands on ACE2 binding area of RBD were performed via AutoDock Vina using Python.

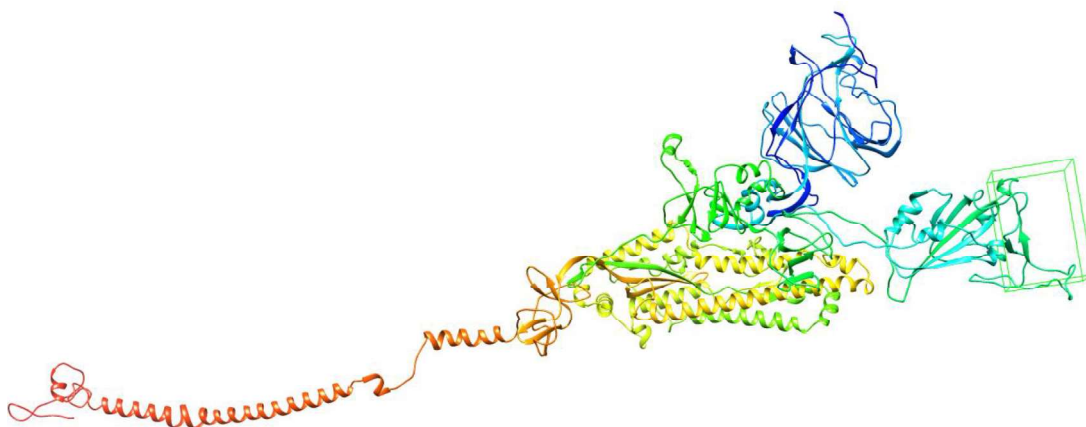


Figure 2.7: Docking box on RBD of Spike protein for AutoDock Vina calculations, visualized in UCSF Chimera.

Ligands were ranked based on their AutoDock Vina binding scores. Top 50.000 ligands with the highest binding affinity/more negative binding energy score were selected for a second docking calculation. The same setup and settings/docking parameters were used to dock these 50.000 ligands to the same docking box. Top 5.000 ligands with the highest binding affinity/more negative binding energy score were selected for the next step.

The binding site of SARS-CoV-2 on human ACE2 Spike protein were determined by choosing the ACE2 area covering critical amino-acid residues which make strong hydrogen bonds and hydrophobic interactions with RBD surface as the center of docking space. These were selected as Gln 24, Thr 27, Asp 30, Lys 31, His 34, Glu 35, Glu 37, Asp 38, Tyr 41, Gln 42, Tyr 83, Asn 330, Lys 353, Gly 354, Asp 355, Arg 357, Ala 387 and Arg 393. The center of docking space was determined by taking the area away from 5 Å of these amino acids and calculated by measure center command of VMD. According to calculated region, docking space was generated using Grid Box function of ADT with following coordinates and sizes and written into configuration file of docking simulation. The center of docking box was selected as center_x = -27.5767, center_y = 5.9075 and center_z = 285.4953. The dimensions of docking box were set as size_x = 31, size_y = 33 and size_z = 15 (Fig 2.8). The exhaustiveness parameter was selected as 24 for docking calculations. Docking calculations of 400.000 ligands on RBD binding area of ACE2 were performed via AutoDock Vina using Python.

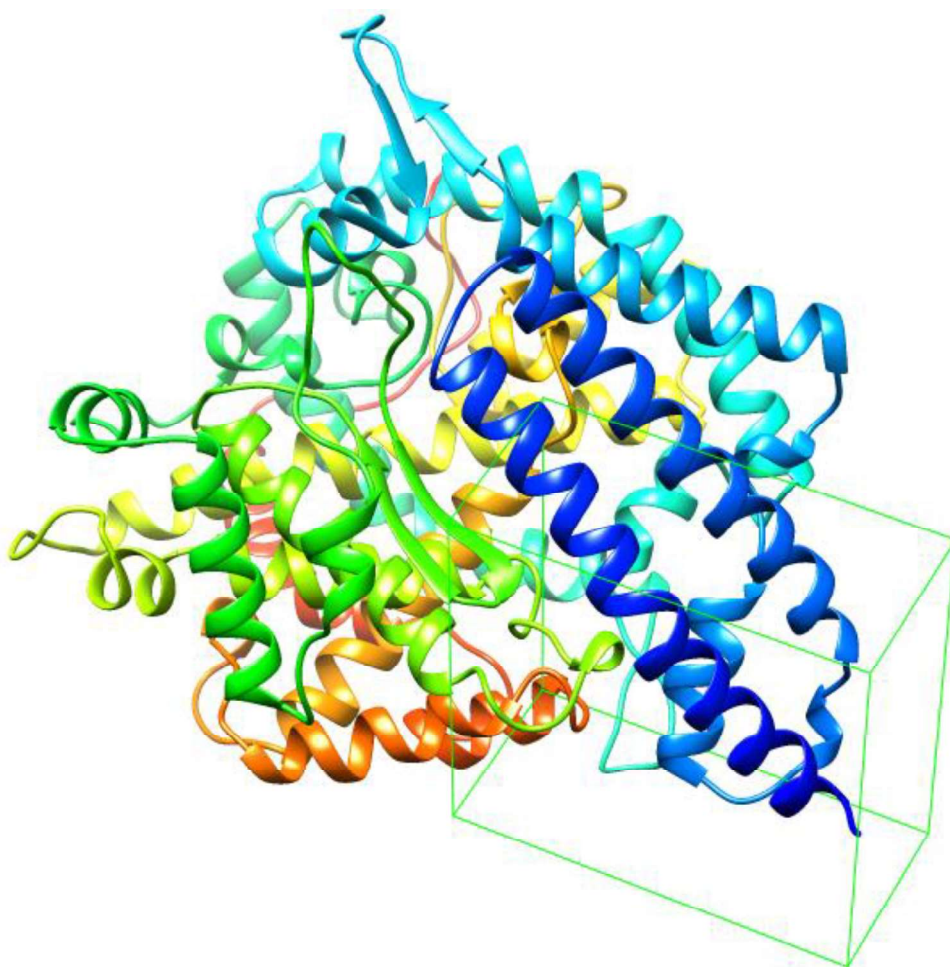


Figure 2.8: Docking box on human ACE2 protein for AutoDock Vina calculations, visualized in UCSF Chimera.

Ligands were ranked based on their AutoDock Vina binding scores. Top 50,000 ligands with the highest binding affinity/more negative binding energy score were selected for a second docking calculation. The same setup and settings/docking parameters were used to dock these 50,000 ligands to the same docking box. Top 5,000 ligands with the highest binding affinity/more negative binding energy score were selected for the next step.

Two docking simulations were performed for Nsp16 structure which are for the interaction site of Nsp10 on Nsp16 and SAM binding pocket of Nsp16. The docking spaces were generated using Grid Box function of ADT with following coordinates and sizes and written into configuration file of docking simulation. The center of docking box for Nsp16-Nsp10 docking was selected as center_x = 76.0000, center_y = 25.0000 and center_z = 16.0000. The dimensions of docking box were set as size_x = 15, size_y = 35 and size_z = 25 (Fig 2.9). This site includes significant residues

making hydrophobic interactions and hydrogen bonds with Nsp10 such as Pro 6835, Ile 6838, Met 6839, Val 6842, Thr 6846, Ala 6871, Val 6876, Pro 6878, Val 6882, Glu 6885, Leu 7042 and Met 7048 (Krafcikova et al., 2020). The center of docking box for Nsp16-SAM docking was selected as center_x = 83.0000, center_y = 15.0000 and center_z = 27.0000. The dimensions of docking box were set as size_x = 25, size_y = 25 and size_z = 25 (Fig 2.10). This site includes significant residues making hydrophobic interactions and hydrogen bonds with SAM such as Asn 6841, Tyr 6845, Phe 6868, Gly 6869, Gly 6879, Ser 6896, Asp 6897, Leu 6898, Asn 6899, Cys 6913, Asp 6928, Met 6929 and Tyr 6930 (Lin et al., 2020). The exhaustiveness parameter was selected as 24 for docking calculations. Docking calculations of 400.000 ligands on NPS16 were performed via AutoDock Vina using Python.

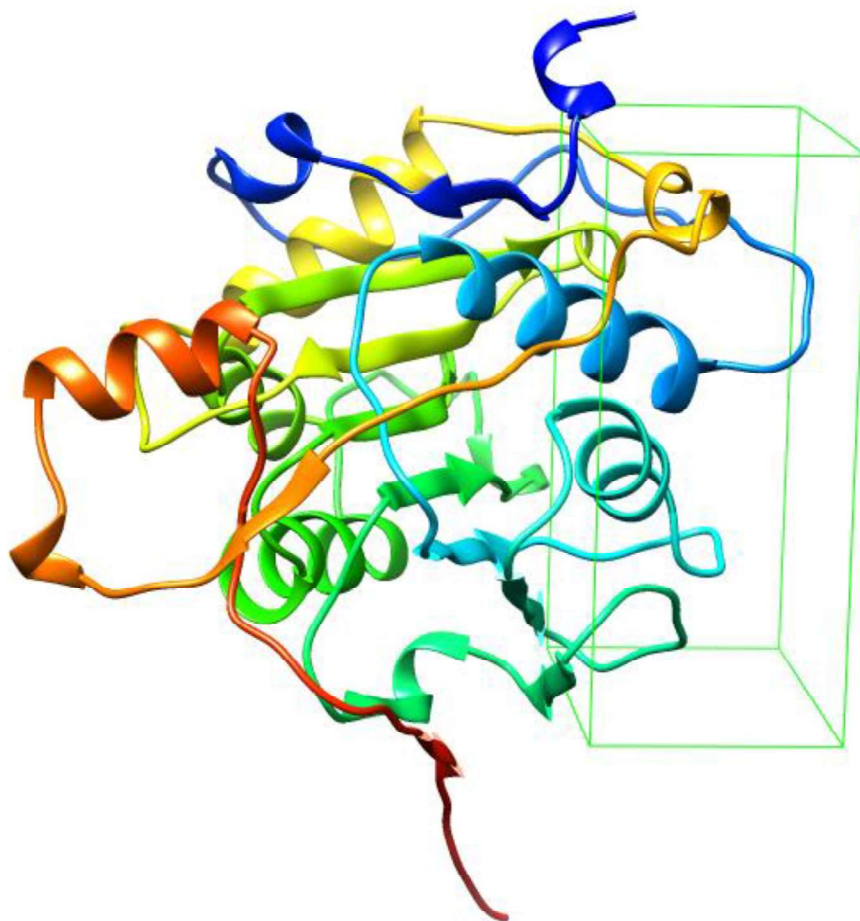


Figure 2.9: Docking box on Nsp10 binding site of Nsp16 protein for AutoDock Vina calculations, visualized in UCSF Chimera.

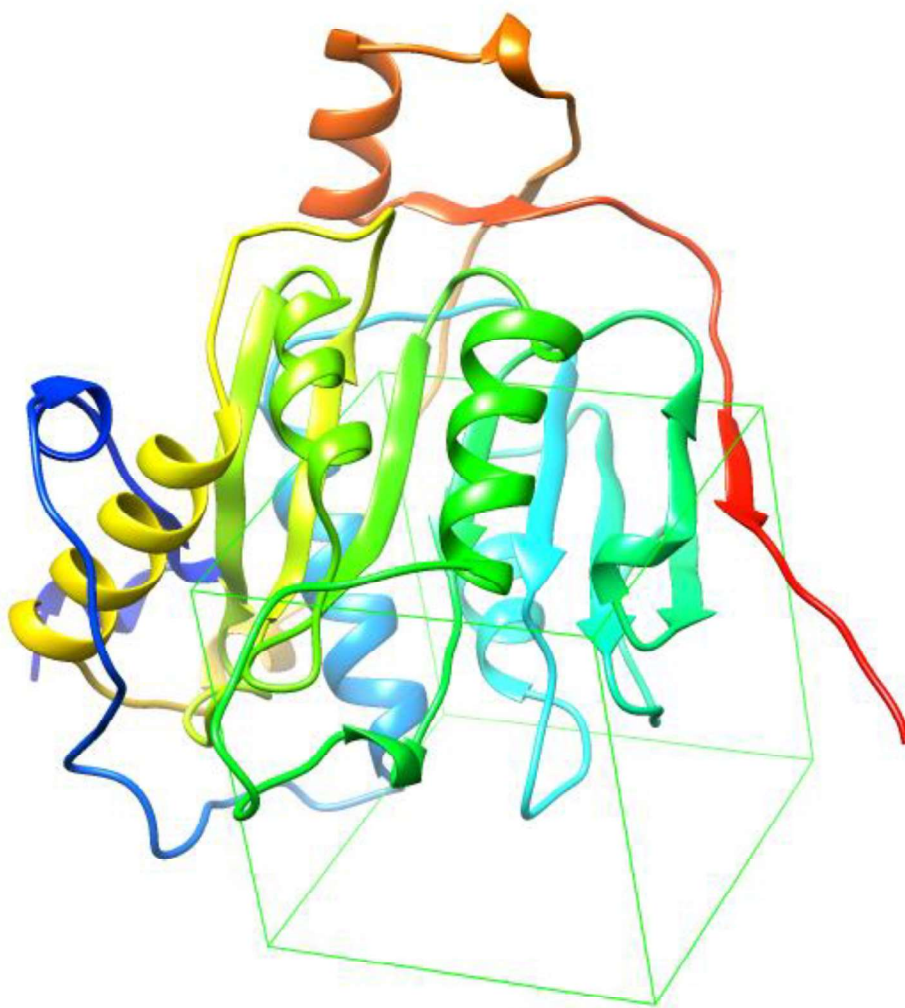


Figure 2.10: Docking box on SAM binding site of Nsp16 protein for AutoDock Vina calculations, visualized in UCSF Chimera.

Ligands were ranked based on their AutoDock Vina binding scores. Top 50.000 ligands with the highest binding affinity/more negative binding energy score were selected for a second docking calculation. The same setup and settings/docking parameters were used to dock these 50.000 ligands to the same docking box. Top 5.000 ligands with the highest binding affinity/more negative binding energy score were selected for the next step for both Nsp16/Nsp10 and Nsp16/SAM docking calculations.

2.5. Extra-precision Docking using Schrödinger Maestro Glide XP

2D structures of 5.000 ligands were again taken from ZINC database in SMILES format. Schrödinger Suite Maestro programme was used for the preparation of ligands molecules for Glide extra-precision (XP) docking. By using LigPrep module of Maestro in conjunction with Epik software program, tautomerization and ionization

calculations were performed. In LigPrep, ionization was done by generating possible states at specified pH:7.0+/-2.0 using Epik. Epik predicted all possible isomers for given ligand at a time by generating its probable protonation states and pKa value. Structures were desalted and tautomers were generated. Specified chiralities and chiral centers were retained. At most 32 variations (stereoisomers) per ligands were generated in MAE format. Among the stereoisomers, one stereoisomer with the lowest state penalty and ionization penalty was selected for each of these 5,000 ligands based on pKa values. Minimization was performed by Optimized Potentials for Liquid Simulations (OPLS_2005) force field. In this way, the most favorable and biochemically sensible conformation with low energy that can promote proper ligand binding were acquired for each ligand.

For the preparation of Spike protein, 6VSB_1_1_1 was imported to “Protein-Preparation-Wizard” module of Maestro as the input. During “preprocess” step, parameters were set by assigning bond orders, using CCD database, adding explicit hydrogens, creating zero-order bonds between metals and near atoms, generating disulfide bonds and deleting waters from 5 Å distance of het groups. Missing side chains were filled by Prime. Het states were generated at pH:7.0+/-2.0 by using Epik and preprocessing was run. This was done to generate the most favorable protonation states. During “refine” step, H-bond optimization was done at pH: 7.0. After optimization, water molecules that make less than 3 hydrogen bonds to non-water molecules were removed. Then, the energy of protein was minimized by restrained minimization, in which heavy atoms are restrained and hydrogen are non-restrained using OPLS_2005 force field and bringing RMSD to 0.30 Å. Minimized S protein was used as input for “Receptor Grid Generation” module. “Receptor” settings were kept at default in which van der Waals radius scaling factor was set to be 1.0 and partial charge cutoff set to be 0.25 to decrease some potential of non-polar atoms. In “Site” settings, “Center” parameters were adjusted according to the X, Y, Z coordinates which were previously used for AutoDock Vina docking box generation of 6VSB_1_1_1. The length of ligands to be docked was set to be less than 10 Å as this threshold was needed to be closer to grid box size for consistent docking and this value was found to be optimum after a series of benchmarking trials with different lengths. The module was run to generate the structure of 6VSB_1_1_1 integrated with grid box for docking (Figure 2.11).

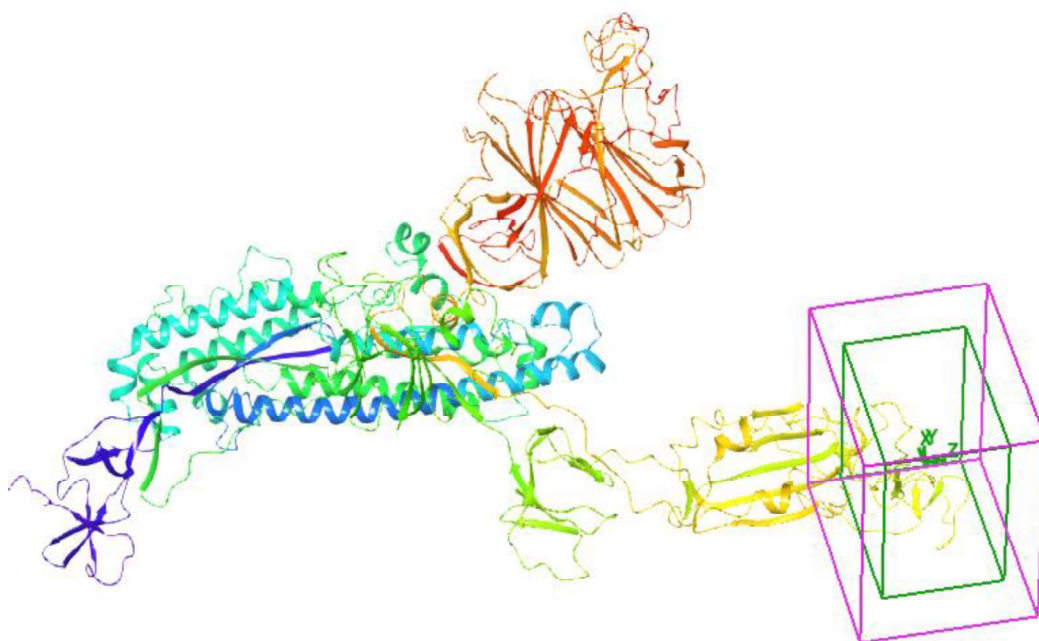


Figure 2.11: The structure of 6VSB_1_1_1 integrated with grid box on Schrödinger Maestro. Green box represents docking box (center_x = -34.3700, center_y = 8.5660 and center_z = 282.1210; size_x = 32, size_y = 32 and size_z = 22). Pink box represents the length of ligands to be docked (<10 Å).

Glide docking calculation was performed by using Glide XP. In “Ligand Docking” panel, 5,000 ligands in their most favorable stereoisomer form with the lowest state penalty and grid box-integrated 6VSB_1_1_1 were imported as the input. To reduce the potential of non-polar atoms in ligands, van der Waals radii scaling factor was set to be 0.80 and partial charge cutoff was set to be 0.15. Glide normally use high-throughput screening (HTVS) and standard precision (SP) to search all possible binding sites for ligands. In this study, “Precision” parameter was set to extra-precision (XP) to run Glide XP. Glide XP conducts more extensive classification and analysis compared to HTVS and SP. False-positive ligands can be identified and a greater scoring function is applied to generate more accurate docking results. “Ligand sampling” was set to be “flexible” as this results in more accurate ligand poses by sampling different conformations and orientations of ligands. For output settings, at most 10 binding poses were requested per ligand. Epik state penalties were added to final docking score as it more accurately reflects true docking score of ligands by excluding energetically unfavourable states. During docking of ligands to Spike RBD grid box site, different ligand conformations analyzed using ligand torsions and rotations. For each stereoisomer, sequentially more favorable ligand poses with

different orientations were generated with better scoring during each iteration. After Glide-XP, post-docking minimization was done using OPLS_2005 force field to reduce energy potential of the entire docked system. All docked ligand poses were filtered, scored and ranked based on their docking scores (Docking score=Glide XP score+Epik penalty). According to this analysis, more negative score means greater binding score and higher binding affinity for a particular ligand. Top 200 ligands with the best docking score (more negative binding energy) were selected for the next step. The same procedure in this section was applied during other three Glide XP docking calculations: ACE2 on its RBD binding site (Figure 2.12), Nsp16 on its Nsp10 binding site (Figure 2.13) and Nsp16 on its SAM binding pocket (Figure 2.14). For each docking calculation, top 200 ligands with the best docking score (more negative binding energy) were selected for the next step. Additionally, SAM (ZINC4214738) was redocked onto the same grid box used for docking calculations on SAM binding pocket of Nsp16 to see the binding potency of tested ligands compared to SAM.

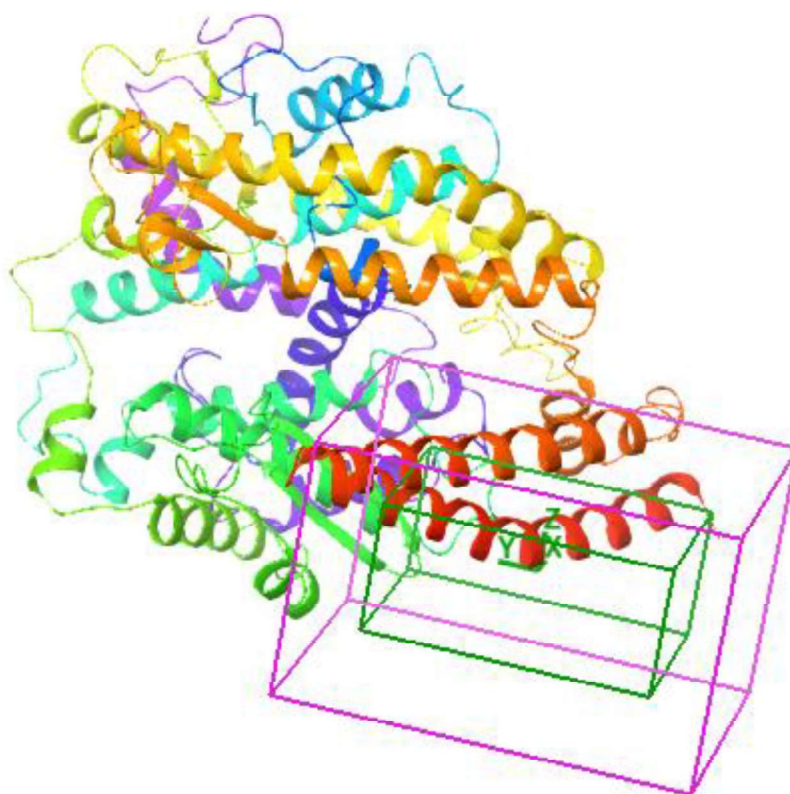


Figure 2.12: The structure of 6VW1 integrated with grid box on Schrödinger Maestro. Green box represents docking box (center_x = -31.5767, center_y = 7.9075 and center_z = 283.4953; size_x = 23, size_y = 29 and size_z = 13). Pink box represents the length of ligands to be docked (<13 Å).

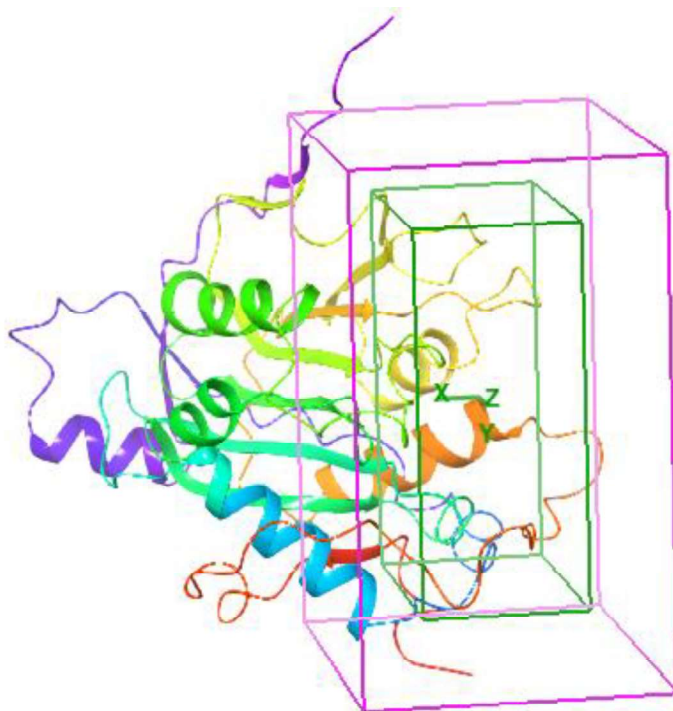


Figure 2.13: The structure of 6W4H integrated with grid box on Schrödinger Maestro. Green box represents docking box (center_x = 76.0000, center_y = 25.0000 and center_z = 16.0000; size_x = 15, size_y = 35 and size_z = 25). Pink box represents the length of ligands to be docked (<13 Å).

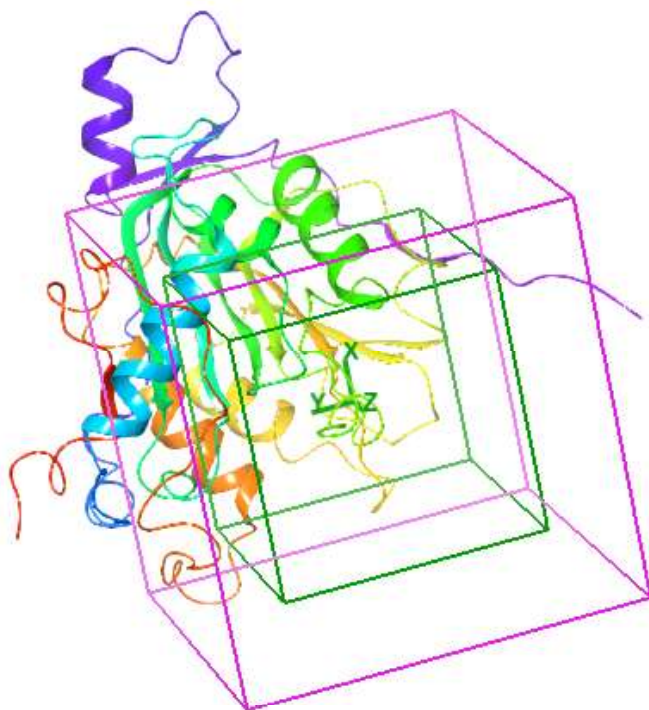


Figure 2.14: The structure of 6W4H integrated with grid box on Schrödinger Maestro. Green box represents docking box (center_x = 83.0000, center_y = 15.0000 and center_z = 27.0000; size_x = 25, size_y = 25 and size_z = 25). Pink box represents the length of ligands to be docked (<13 Å).

2.6. Fingerprint Clustering using Schrödinger CANVAS

Top 200 ligands for the docking of Spike protein were clustered by using CANVAS software of Schrödinger Suite. In CANVAS software, top 200 ligands were imported as input for “fingerprint clustering”. In “Applications” section, “Binary Fingerprints from Structures” was selected and “Dendritic fingerprinting” was run along with other settings at default. FP (fingerprinting) Dendritic scores were ranked and formed into a separate column. Then, in “Applications” section, hierarchical clustering was applied. For this application, Dendritic FP score column was used as fingerprinting column and “Tanimoto similarity” was used as “Metric” parameter. Among final clusters generated from this hierarchical clustering, clustering cut-off “Merging Distance” was set to be optimum value (generally, between 0.5 and 0.7) to cluster chemically and structurally more similar ligands together and increase the accuracy of clustering. Each cluster was analyzed separately to verify the fact that similar ligands were gathered into the same cluster and vastly different ones were clustered to different clusters, indicating clustering accuracy. After fingerprinting clustering, a single ligand with the highest docking score and the lowest state penalty (higher binding affinity, greater stability and more chemically favorable) was selected from each cluster to prevent the resemblance of ligands which was used for the next step and increase the ligand variety.

The same analysis with the same settings were performed for top 200 ligands which were docked on RBD interaction site of ACE2, Nsp10 interaction site of Nsp16 and SAM binding site of Nsp16.

2.7. ADME/T Analyses using Schrödinger QikProp

Selected ligands, that can bind to ACE2 interaction site of 6VSB_1_1_1 and represent each cluster, was used for ADME/T (absorption, distribution, metabolism, and excretion/toxicity) analysis. They were screened for their pharmaceutical properties by using QikProp module of Meastro to define optimum ligands for the use. Screened ligands were scored and ranked based on these values. The most optimum 100 ligands in terms of drug-likeness and safety were selected based on the pharmaceutical properties explained in Table 2.2, to be used for following molecular dynamic simulations. Prioritized properties were stars, RuleOfFive, RuleOfThree, rtvFG and

Metab in order as they are most critical ones to reflect the drug-likeness and toxicity of a particular ligands. Ligands with the least violations from these properties can be assessed as potent inhibitors with greater pharmaceutical capabilities and lower risk of adverse reactivity/toxicity.

The same ADME analysis was done for the clustered ligands which are belonged to RBD interaction site of 6VW1, Nsp10 interaction site of 6W4H and SAM binding site of 6W4H.

Table 2.2: Important pharmaceutical properties of ligands for ADME/T analysis (retrieved from QikProp 4.4 User Manual).

Pharmaceutical Property	Description	Optimum Range
stars	Drug-likeness (Number of properties outside of 95% range of known drugs)	0-5 (lower is preferred, especially 0-1)
rtvFG	Reactive functional group number (low toxicity in low values)	0-2 (lower is preferred, especially 0-1)
mol_MW	Molecular weight	130.0-725.0
QPlogS and CIQPlogS	Aqueous solubility	-6.5-0.5
QPPCaco	Gut-blood permeability	poor for <25 great for >500
QplogBB	Brain-blood permeability	-3.0-1.2
Metab	Possible metabolic reaction numbers	1-8 (lower values like ~1-5 are preferred)
HumanOralAbsorption	Oral absorption value of drug	1 (low), 2 (medium), 3 (high) with high is preferred
PercentHumanOralAbsorption	Oral absorption rate of drug	poor for <25% high for >80% (preferred)
RuleOfFive	Violations from Lipinski's rule of five representing properties of drug-likeness	not more than 4 (0-1 preferred)
RuleOfThree	Violations from Jorgensen's rule of five representing properties of drug-likeness	not more than 4 (0-1 preferred)

2.8. Analyses of Interaction between Target Proteins and Ligands

10 ligands among most drug-like, lower toxicity, low energy and high binding-affinity 100 ligands related to each target was selected to analyze molecular interactions with their respective targets. Protein-ligand interactions were assessed by “Ligand Interaction Diagram” module of Maestro Software. Hydrogen bonds and hydrophobic interactions which were formed between ligand and binding site were identified. They were also compared to residues from literature which are critical to S protein RBD-ACE2, Nsp16-Nsp10 and Nsp16-SAM interactions.

2.9. Molecular Dynamic Simulations using Schrödinger Desmond

Previously selected best 10 ligands for each target (S protein RBD, human ACE2, Nsp10 interaction site of Nsp16 and SAM binding site of Nsp16) was used for MD simulations. In Schrödinger Desmond Software, “System Builder” was used to generate solvent box for each bound ligand-target protein complex. Water model was set to be TIP3P, orthorhombic box was used and the system was neutralized by adding given ions. The system was minimized and solvation box was generated. In “Desmond Molecular Dynamics” panel, simulation time was set to be 100 ns and temperature was set to be 300 K. No restraint was added with the exception of RBD (intersection of backbone and residues between Asn 334 and Gly 526) as only RBD part was taken from entire Spike protein structure. RMSD were calculated OSPL_2005 force field. After MD simulations, atom trajectory demonstrating changes in RMSD values of protein alpha carbons, side chains and the ligand over 100 ns was generated for each complex. Changes in protein RMSD values were analyzed to identify conformational changes of protein and its stability during the simulation. Changes in ligand RMSD values were analyzed to identify ligand stability and its retention within binding pocket during the simulation. Lastly, the first and the last frame of protein-ligand complexes were superimposed to show displacements and the state of stability. After these analyses, ligands with the best RMSD values were identified.



3. RESULTS

3.1. Small-molecule Inhibitor Prediction for SARS-CoV-2 Spike RBD

After initial docking of 400.000 ligands, ligands were scored and ranked based on their docking binding energy scores as seen in Figure 3.1. AutoDock Vina docking scores of ligands showed a docking score range from -9,20 kcal/mol to -2,90 kcal/mol. Docking scores were populated around approximately -6,21 kcal/mol. Additionally, the top 50.000 ligands showed a docking score range from -9,20 kcal/mol to -7,00 kcal/mol. These values indicate that the initial screen of randomized 400.000 ligand library gave a large number of ligands with great docking score and high binding affinity to ACE2 binding site of Spike RBD.

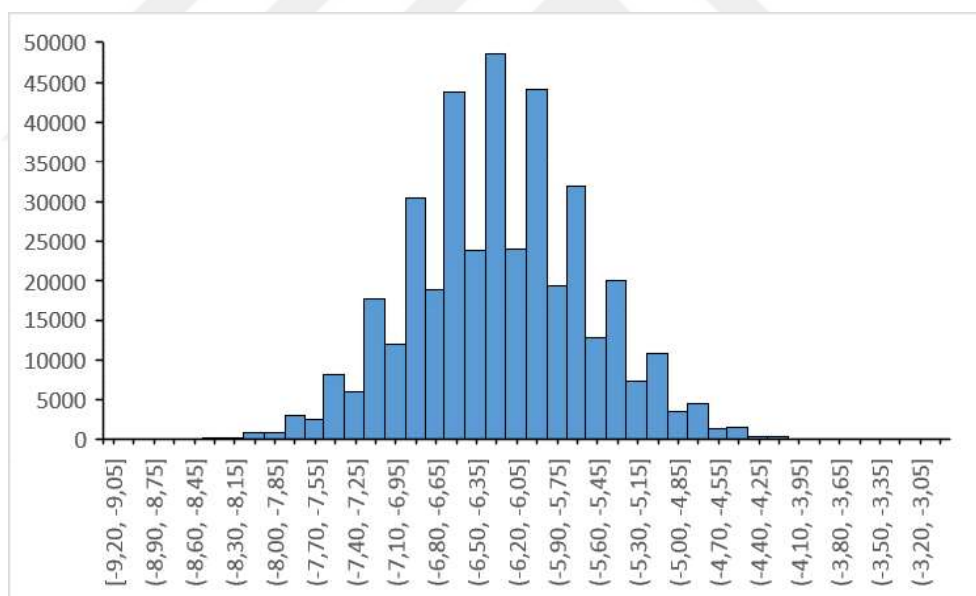


Figure 3.1: Distribution of AutoDock Vina docking scores (x-axis) of 400.000 ligands (y-axis) against ACE2 binding site on Spike RBD.

After scoring and sorting, second run of AutoDock Vina with top 50.000 ligands showed a ligand docking score range from -9,20 kcal/mol to -4,70 kcal/mol. Docking scores were populated around -7,25 kcal/mol which is better than the initial run of ligand samples as seen in Figure 3.2. This indicates that a second iteration of docking improved the docking results.

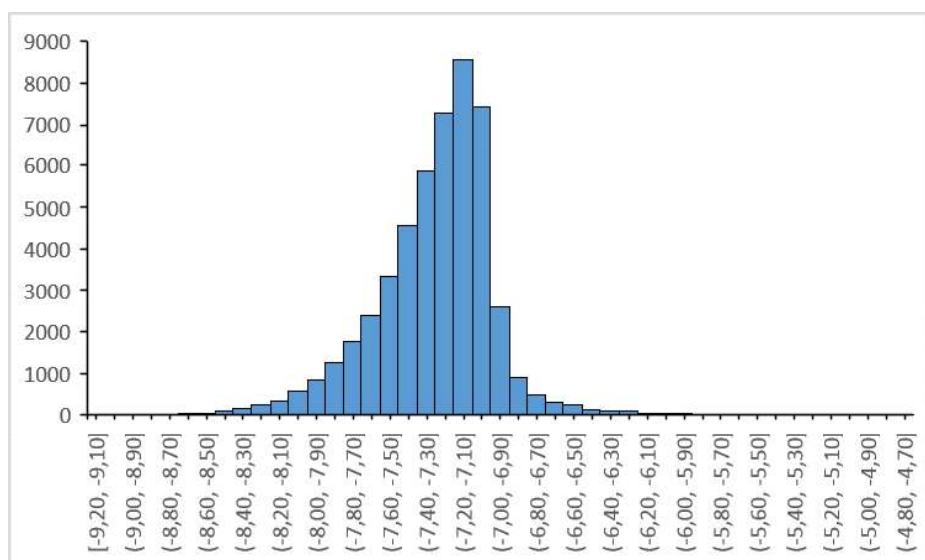


Figure 3.2: Distribution of AutoDock Vina docking scores (x-axis) of 50.000 ligands (y-axis) against ACE2 binding site on Spike RBD.

Moreover, top 5.000 ligands demonstrated docking range from -9,20 kcal/mol to -7,70 kcal/mol as seen in Figure 3.3. The highest scored ligands were also almost the same with the top results of initial docking run, indicating that top-docked ligands can always bind to target site with high affinity and there is high consistency between docking analyses.

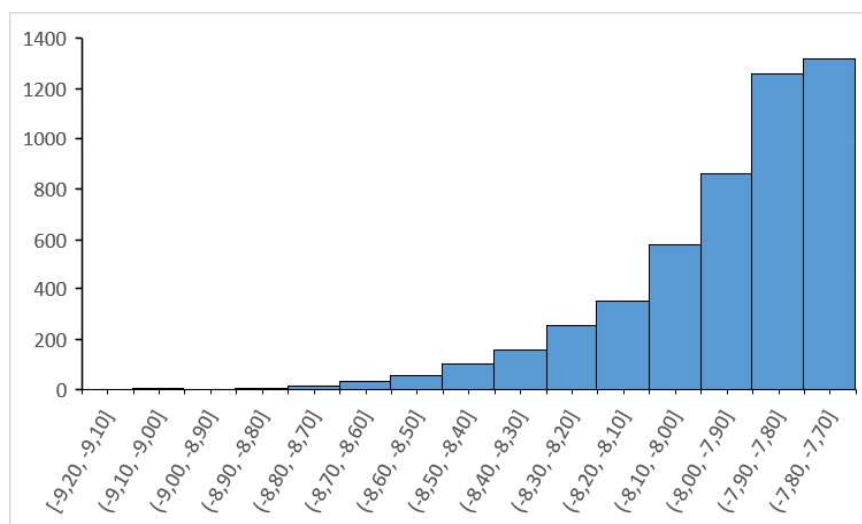


Figure 3.3: Distribution of AutoDock Vina docking scores (x-axis) of 5.000 ligands (y-axis) against ACE2 binding site on Spike RBD.

Top 5.000 ligands with the highest docking score were prepared by LigPrep module of Maestro. Total 7.378 stereoisomers of top 5.000 ligands were generated. The most favorable and biochemically sensible conformation with low energy were identified for each ligand based on their low state penalty and ionization penalty. Almost all of

the most favorable stereoisomers of 5.000 ligands showed ionization and state penalty closer to 0, indicating that ligands used for Glide XP were chemically stable and sensible; and they represented the most probable conformation of a particular ligand being closer to real biological system.

After Glide XP docking of 5.000 ligands, ligands were scored and ranked based on their Glide XP binding energy. Docking score was taken into consideration instead of Glide score as it also includes Epik penalty values. Thus, Epik penalty-added final docking scores accurately reflected true docking score of ligands by excluding energetically unfavourable states. As seen in Figure 3.4, Glide docking scores of small-molecule ligands were in the range from -6,848 kcal/mol to 1,898 kcal/mol. Additionally, the top 200 ligands showed a docking score range from -6,848 kcal/mol to -4,221 kcal/mol. Since Glide XP used a greatly extensive, different scoring function and Epik penalties were added to exclude energetically unlikely states, docking scores were lower compared to AutoDock Vina docking scores as expected. In Glide XP calculations, many researches considered docking scores less than -5,5 kcal/mol good and the ones less than -8 kcal/mol great. Thus, it could be stated that majority of top ligands docked here had good docking scores and could bind the target protein with a high affinity. Moreover, the ligands with highest docking score were also mostly similar to the ones with highest AutoDock Vina docking scores. This finding further verifies that there is high consistency between each docking steps and the screened top 200 ligands could represent potential small-molecule inhibitors with high binding affinity to ACE2 binding site of Spike RBD.

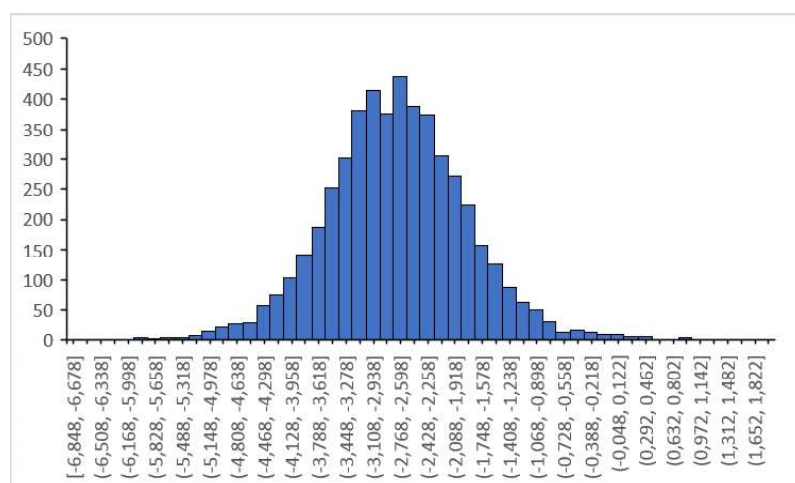


Figure 3.4: Distribution of docking scores (x-axis) of 5.000 ligands (y-axis) against ACE2 binding site on Spike RBD from Glide XP docking simulation.

After performing the dendritic fingerprinting hierarchical clustering, cut-off “Merging Distance” was set to be 0.7 and 164 different clusters were generated as seen in Figure A.1 in Appendix A. As a result of analysis of each cluster based on the structural and chemical similarities of cluster members, it was concluded that each cluster was highly unique as members of clusters are similar within clusters and different between clusters. This can be seen in Figure 3.5 which demonstrates the in-cluster similarity and between-clusters differences for 3 randomly selected clusters.

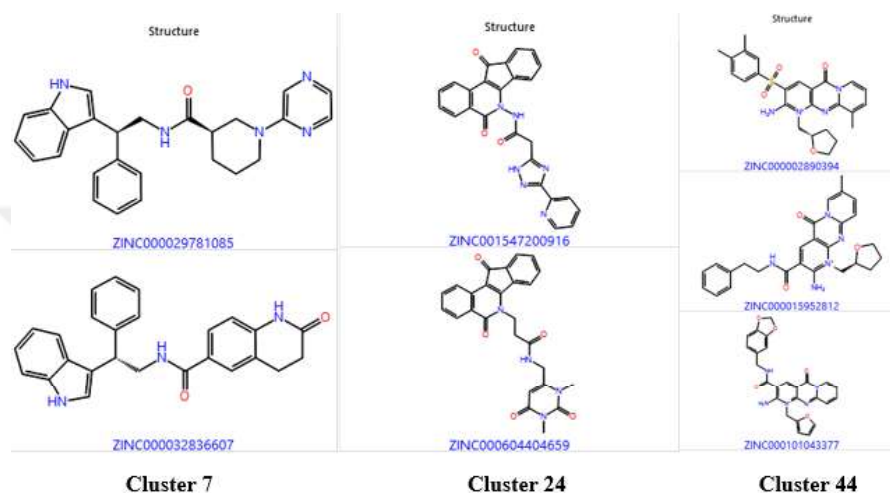


Figure 3.5: Three clusters from dendritic fingerprinting hierarchical clustering of 200 ligands against ACE2 binding site on Spike RBD, identified by CANVAS Software.

For each cluster, the member with the greatest docking score and lowest state penalty was selected as the representative of its respective cluster. These 164 ligands were then screened for ADME/T analysis to assess their pharmaceutical properties and toxicity. 100 molecules with the most optimum ADME/T properties were identified. Among these most drug-like, lower toxicity, low energy and high binding-affinity 100 ligands, top-scored 10 ligands were selected for further steps. Their state penalty, docking scores and ADME/T values can be seen in Table 3.1. All values were in overall great range and this indicates that these ligands can show proper pharmaceutical properties and this procedure might predict drug-like small-molecule ligands against ACE2 binding site on Spike RBD.

Table 3.1: Low-energy and high binding-affinity of 10 ligands with the lowest state penalty, the highest docking scores and optimum ADME/T values against ACE2 binding site of Spike RBD.

Important Values	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00
	0067920	0011870	0009264	0009828	0009939	1772594	0014228	0009134	0008772	0001360				
	242	981	459	805	150	593	954	748	453	058				
State Penalty	0.0862	0.0025	0.3679	0.0000	0.1001	0.0176	0.0000	0.0000	0.0000	0.0440				
Final AutoDock Vina														
Score (kcal/mol)	-8.400	-7.700	-7.700	-7.700	-7.700	-7.800	-7.800	-8.100	-7.800	-7.800				
Glide XP Docking														
Score (kcal/mol)	-6.848	-6.525	-5.938	-5.908	-5.897	-5.366	-5.230	-5.173	-5.148	-5.062				
stars	0	0	0	0	0	0	0	0	0	0				
rtvFG	0	0	0	1	0	0	0	0	0	0				
mol_MW	375.469	477.992	347.419	421.455	419.503	463.483	367.404	441.442	403.480	432.453				
QPlogS	-4.513	-4.766	-4.167	-4.893	-5.649	-5.113	-5.168	-5.506	-5.667	-5.083				
CIQPlogS	-4.485	-5.601	-4.931	-5.281	-5.839	-5.322	-4.579	-5.826	-5.643	-6.125				

Table 3.1 (continued): Low-energy and high binding-affinity of 10 ligands with the lowest state penalty, the highest docking scores and optimum ADME/T values against ACE2 binding site of Spike RBD.

Important Values	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00
	0067920	0011870	0009264	0009828	000939	000939	1772594	0014228	0009134	0008772	0001360						
	242	981	459	805	150	150	593	954	748	453	058						
QPPCaco	489.225	155.294	326.972	1,778.54	658.989	285.119	238.570	329.526	687.383	504.527							
QplogBB	-0.072	-0.936	-1.169	-0.389	-0.949	0.040	-1.446	-1.325	-0.772	-0.949							
Metab	7	4	4	5	5	1	5	2	5	4							
HumanOralAbsorp tion	3	3	3	3	3	3	3	3	3	3							
PercentHumanOral Absorption	100.000	87.746	87.572	100.000	100.000	92.737	83.772	90.334	100.000	100.000							
RuleOffFive	0	0	0	0	0	0	0	0	0	0							
RuleOfThree	1	0	0	0	0	0	0	0	0	0							

Top-scored ligands were analyzed for their interaction with Spike RBD by Ligand Interaction Diagram module of Maestro Software. Ligands were observed to be properly docked onto ACE2 binding site of Spike RBD (Figure 3.6).

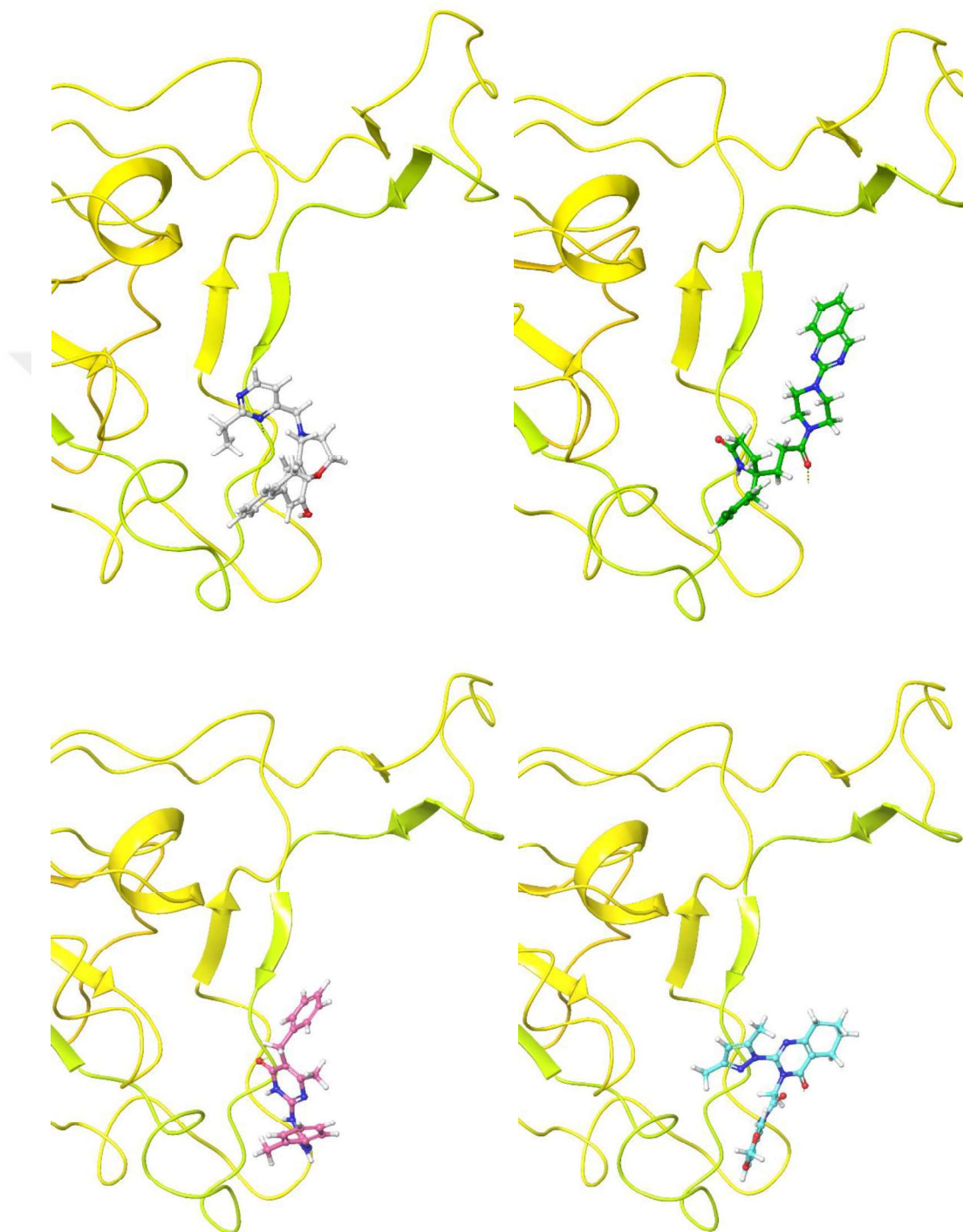


Figure 3.6: Top-scored 10 ligands in complex with Spike RBD, visualized on Maestro. From left to right: ZINC000067920242, ZINC000011870981, ZINC000009264459, ZINC000009828805, ZINC000009939150, ZINC001772594593, ZINC000014228954, ZINC000009134748, ZINC000008772453, ZINC000001360058.

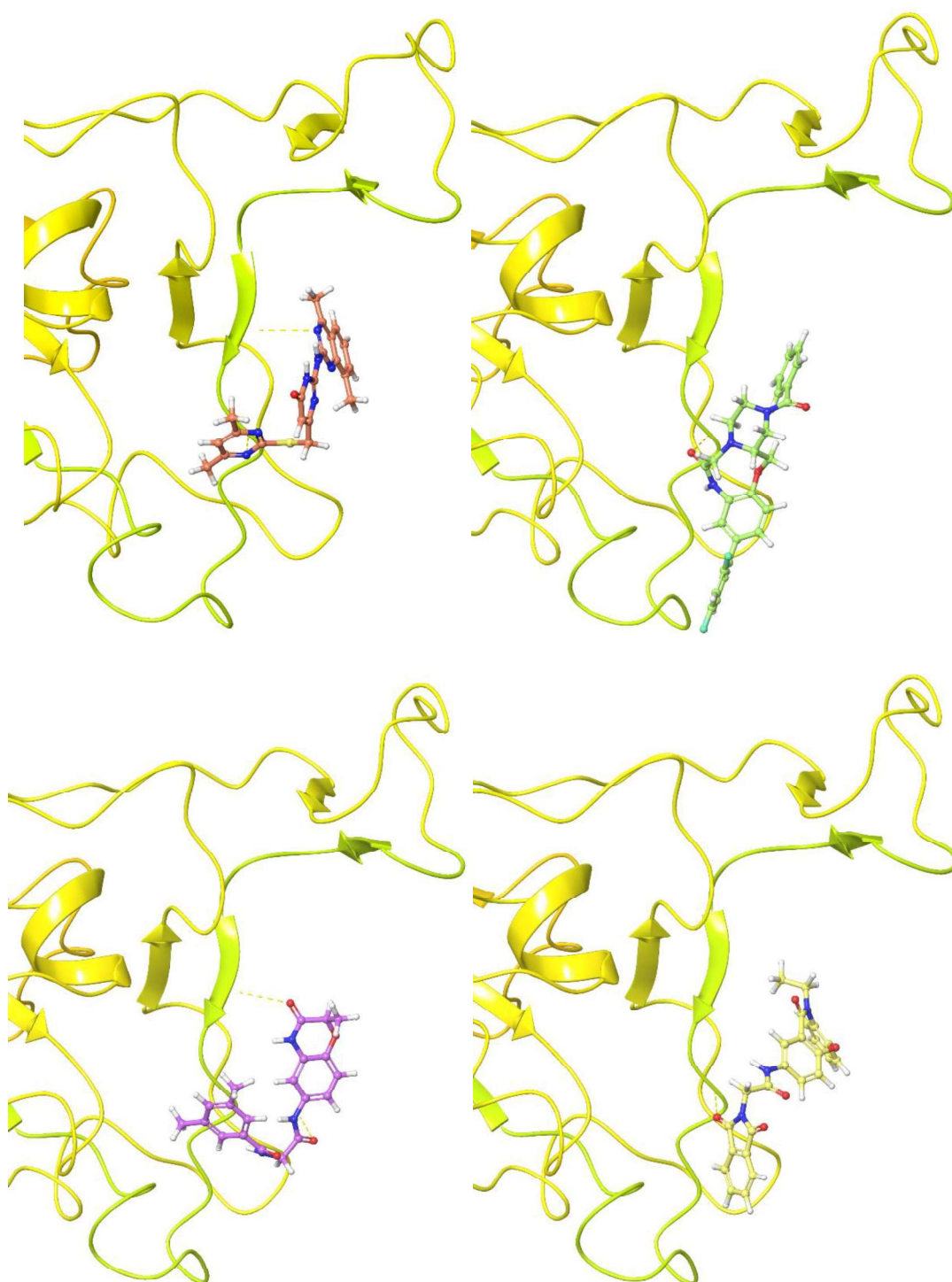


Figure 3.6 (continued): Top-scored 10 ligands in complex with Spike RBD, visualized on Maestro. From left to right: ZINC000067920242, ZINC000011870981, ZINC000009264459, ZINC000009828805, ZINC000009939150, ZINC001772594593, ZINC000014228954, ZINC000009134748, ZINC000008772453, ZINC000001360058.

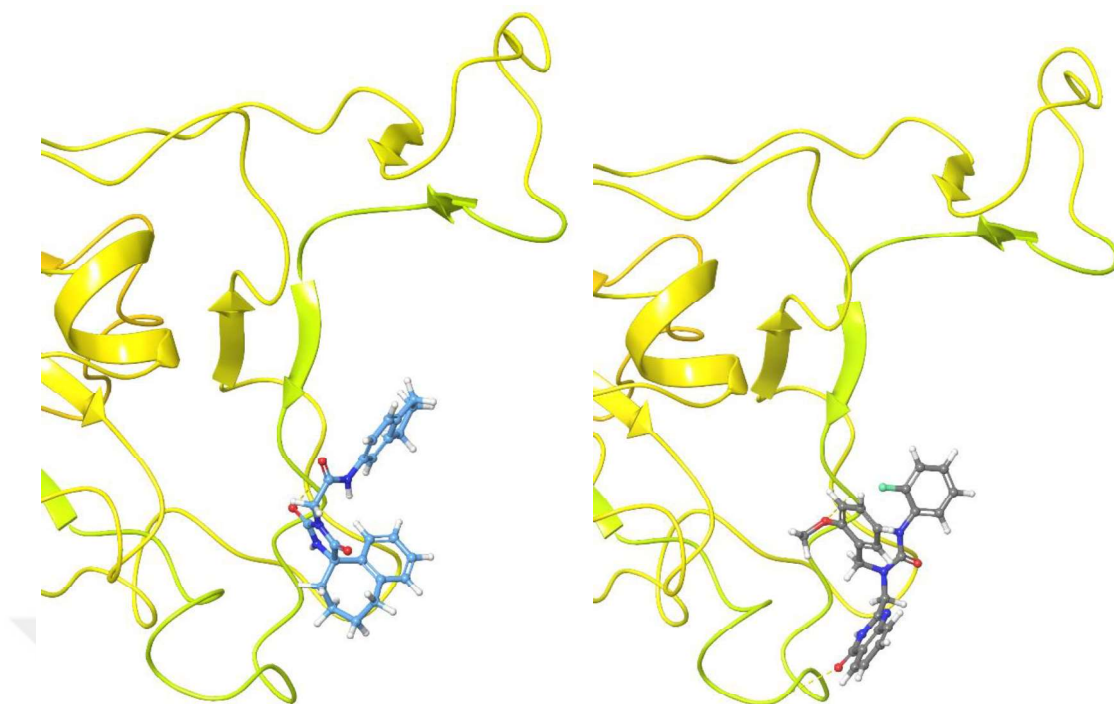


Figure 3.6 (continued): Top-scored 10 ligands in complex with Spike RBD, visualized on Maestro. From left to right: ZINC000067920242, ZINC000011870981, ZINC000009264459, ZINC000009828805, ZINC000009939150, ZINC001772594593, ZINC000014228954, ZINC000009134748, ZINC000008772453, ZINC000001360058.

Moreover, a high number of hydrophobic interactions and hydrogen bonding was identified (Figure 3.7, Table 3.2), further verifying that these top-docked ligands can bind to Spike RBD with high affinity and can block binding of other molecules. Major residues making hydrophobic interactions observed in ligands were Tyr 449, Leu 452, Tyr 453, Leu 455, Phe 490, Leu 492, Tyr 495, Phe 497 and Tyr 505. Major residues making hydrogen bonds observed in ligands were Tyr 449, Gln 493, Ser 494, Tyr 495, Gly 496, Phe 497, Gln 498, Asn 501 and Gly 502. Most of these residues were also the same with or next to critical Spike RBD residues required for Spike RBD and ACE2 interaction such as Gly 446, Tyr 449, Tyr 453, Tyr 489, Gln 493, Gln 498, Thr 500, Gly 502 and Tyr 505. This highly suggests that the use of these ligands can block Spike RBD-ACE2 interaction as they can antagonistically overlap and bind to the same or close residues on binding interface with high affinity.

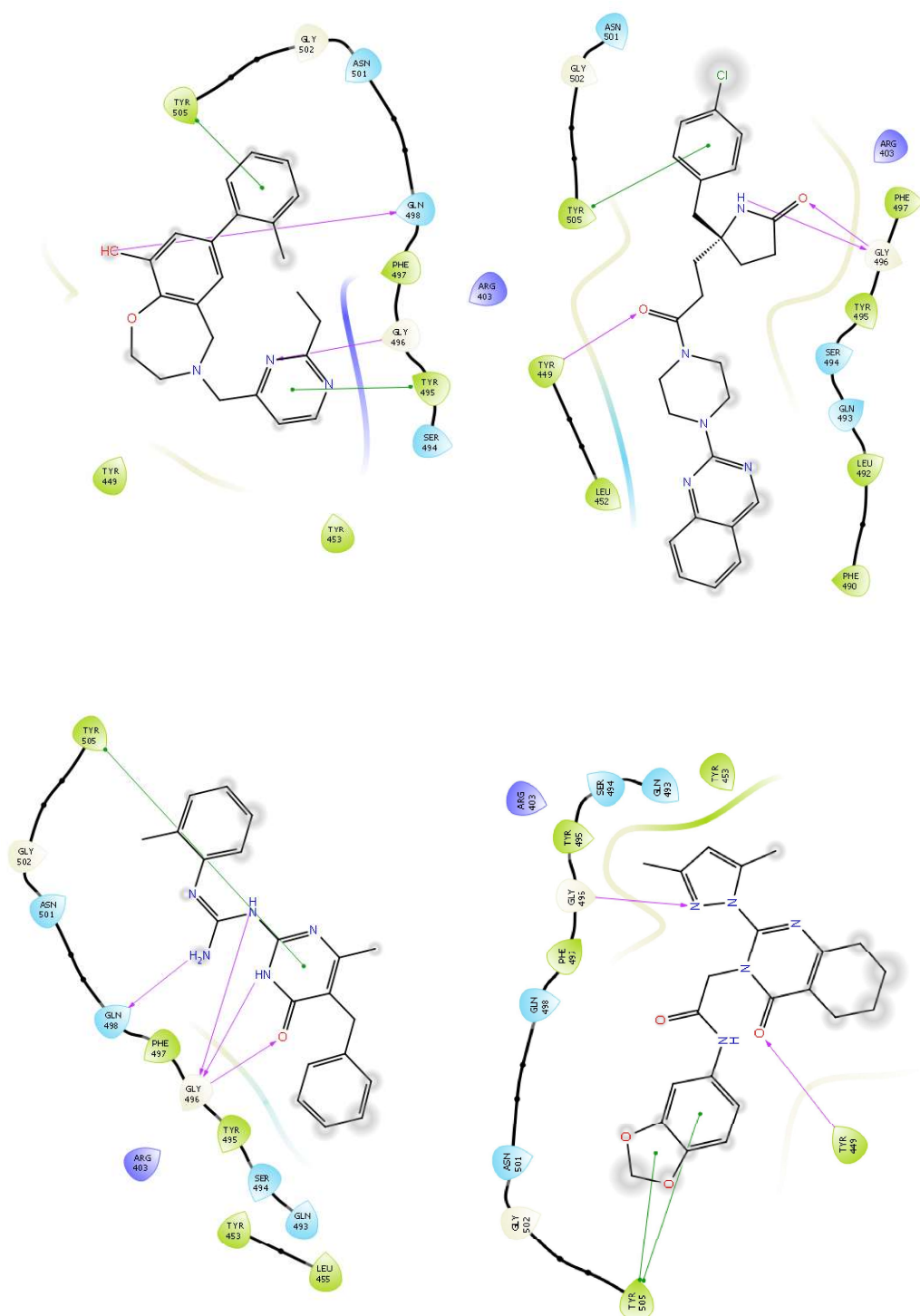


Figure 3.7: Ligand interaction diagrams of top-scored 10 ligands with Spike RBD, visualized by Schrödinger Maestro. Red residue: Negatively charged, Purple residue: Positively charged, White residue: Glycine, Yellow residue: Hydrophobic, Blue residue: Polar, Purple arrow: H-bond, Orange arrow: Halogen Bond, Purple line: Metal coordination, Green line: Pi-Pi stacking, Grey circle: Solvent exposure. From left to right: ZINC000067920242, ZINC000011870981, ZINC000009264459, ZINC000009828805, ZINC000009939150, ZINC001772594593, ZINC000014228954, ZINC000009134748, ZINC000008772453, ZINC000001360058.

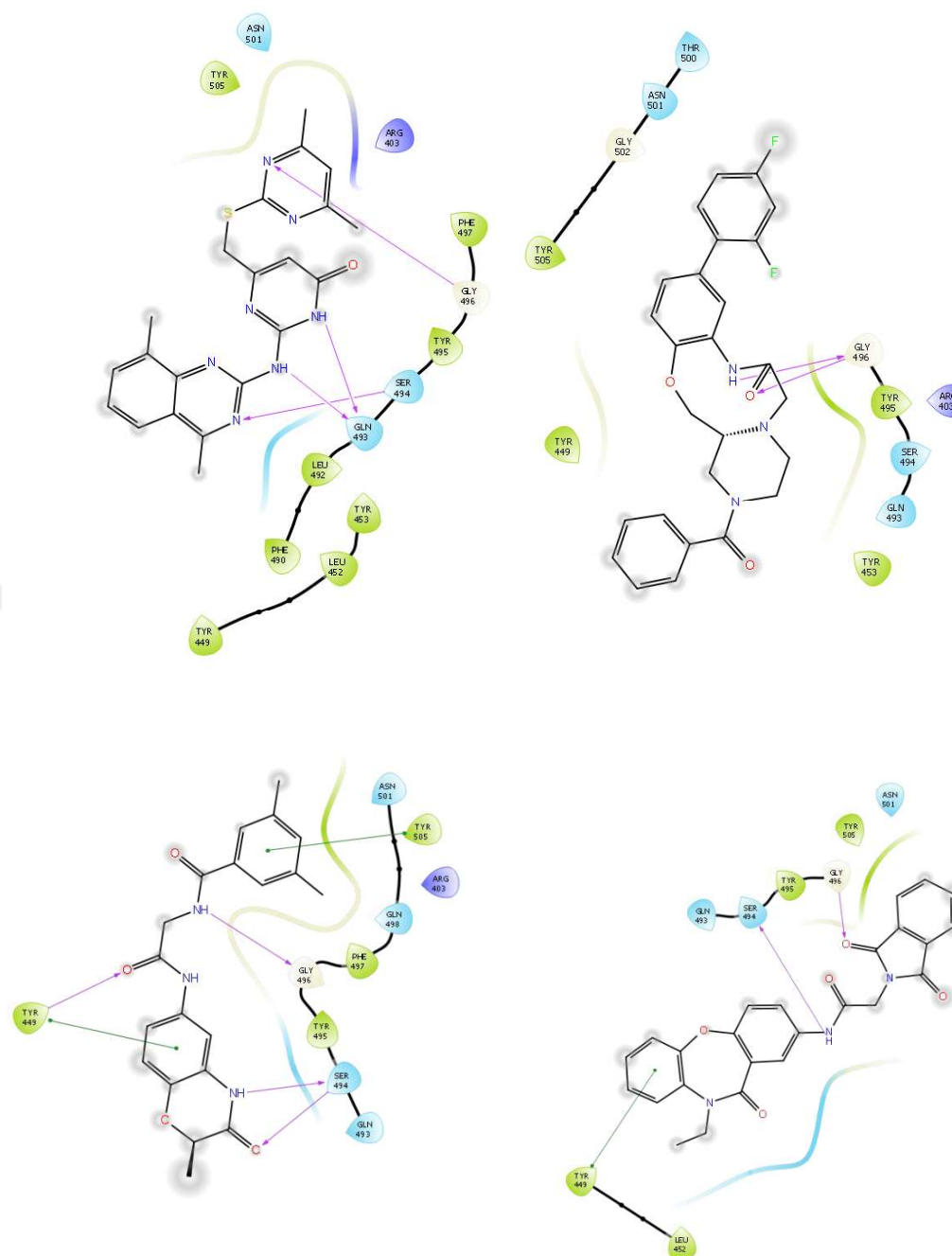


Figure 3.7 (continued): Ligand interaction diagrams of top-scored 10 ligands with Spike RBD, visualized by Schrödinger Maestro. Red residue: Negatively charged, Purple residue: Positively charged, White residue: Glycine, Yellow residue: Hydrophobic, Blue residue: Polar, Purple arrow: H-bond, Orange arrow: Halogen Bond, Purple line: Metal coordination, Green line: Pi-Pi stacking, Grey circle: Solvent exposure. From left to right: ZINC000067920242, ZINC000011870981, ZINC000009264459, ZINC000009828805, ZINC000009939150, ZINC001772594593, ZINC000014228954, ZINC000009134748, ZINC000008772453, ZINC000001360058.

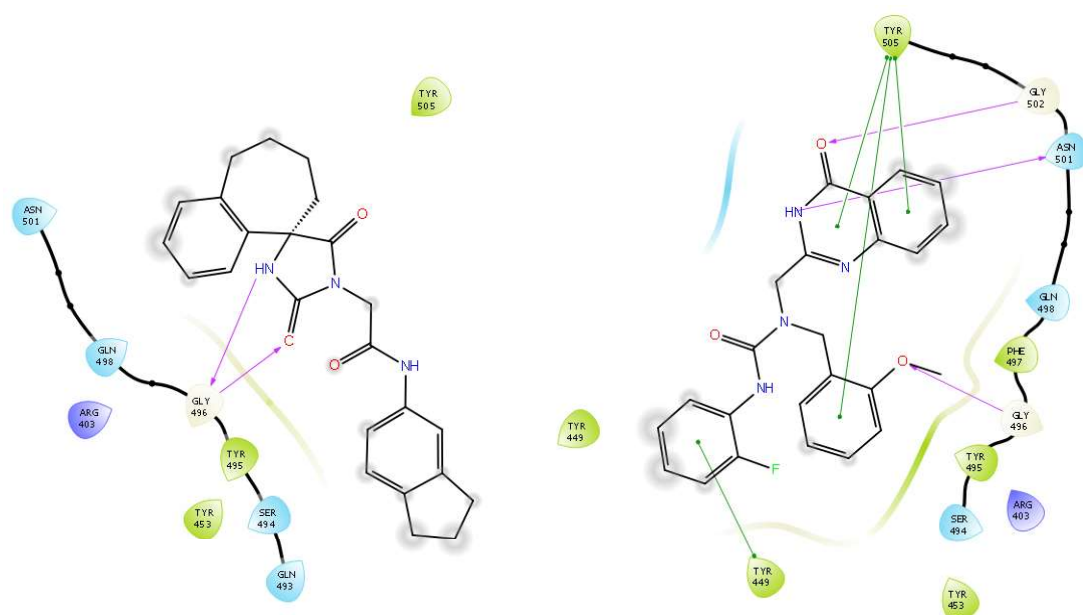


Figure 3.7 (continued): Ligand interaction diagrams of top-scored 10 ligands with Spike RBD, visualized by Schrödinger Maestro. Red residue: Negatively charged, Purple residue: Positively charged, White residue: Glycine, Yellow residue: Hydrophobic, Blue residue: Polar, Purple arrow: H-bond, Orange arrow: Halogen Bond, Purple line: Metal coordination, Green line: Pi-Pi stacking, Grey circle: Solvent exposure. From left to right: ZINC000067920242, ZINC000011870981, ZINC000009264459, ZINC000009828805, ZINC000009939150, ZINC001772594593, ZINC000014228954, ZINC000009134748, ZINC000008772453, ZINC000001360058.

Table 3.2: Protein-ligand interaction profile of Spike RBD and top-scored 10 small-molecule ligands.

Small-molecule Ligand	Hydrophobic Interactions	Hydrogen Bonds
ZINC000067920242	Tyr 449, Tyr 453, Tyr 495, Phe 497, Tyr 505	Tyr 495, Phe 497
ZINC000011870981	Tyr 449, Leu 452, Phe 490, Leu 492, Tyr 495, Phe 497	Tyr 449, Gly 496
ZINC000009264459	Tyr 453, Leu 455, Tyr 495, Phe 497, Tyr 505	Gly 496, Gln 498
ZINC000009828805	Tyr 449, Tyr 453, Tyr 495, Phe 497, Tyr 505	Tyr 449, Gly 496
ZINC000009939150	Tyr 449, Leu 452, Tyr 453, Phe 490, Leu 492, Tyr 495, Phe 497, Tyr 505	Gln 493, Ser 494, Gly 496

Table 3.2 (continued): Protein-ligand interaction profile of Spike RBD and top-scored 10 small-molecule ligands.

Small-molecule Ligand	Hydrophobic Interactions	Hydrogen Bonds
ZINC000067920242	Tyr 449, Tyr 453, Tyr 495, Phe 497, Tyr 505	Tyr 495, Phe 497
ZINC000014228954	Tyr 449, Tyr 495, Phe 497, Tyr 505	Tyr 449, Ser 494, Gly 496
ZINC000009134748	Tyr 449, Leu 452, Tyr 495, Tyr 505	Ser 494, Gly 496
ZINC000008772453	Tyr 449, Tyr 453, Tyr 495, Tyr 505	Gly 496
ZINC000001360058	Tyr 449, Tyr 453, Tyr 495, Phe 497, Tyr 505	Gly 496, Asn 501, Gly 502

3.2. Small-molecule Inhibitor Prediction for Human ACE2 Receptor

After initial docking of 400.000 ligands, ligands were scored and ranked based on their docking binding energy scores as seen in Figure 3.8. AutoDock Vina docking scores of ligands showed a docking score range from -9,10 kcal/mol to -3,10 kcal/mol. Docking scores were populated around approximately -6,08 kcal/mol. Additionally, the top 50.000 ligands showed a docking score range from -9,10 kcal/mol to -6,90 kcal/mol. These values indicate that the initial screen of randomized 400.000 ligand library gave a large number of ligands with great docking score and high binding affinity to Spike RBD binding site of human ACE2 receptor.

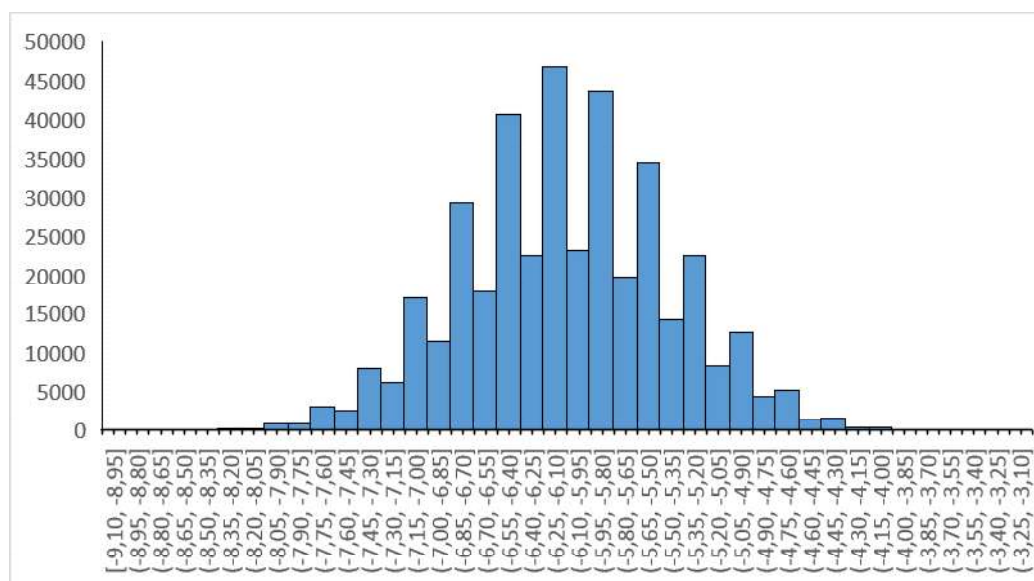


Figure 3.8: Distribution of AutoDock Vina docking scores (x-axis) of 400.000 ligands (y-axis) against Spike RBD binding site of human ACE2.

After scoring and sorting, second run of AutoDock Vina with top 50.000 ligands showed a ligand docking score range from -9,10 kcal/mol to -4,20 kcal/mol. Docking scores were populated around -6,99 kcal/mol which is better than the initial run of ligand samples as seen in Figure 3.9. This indicates that a second iteration of docking improved the docking results.

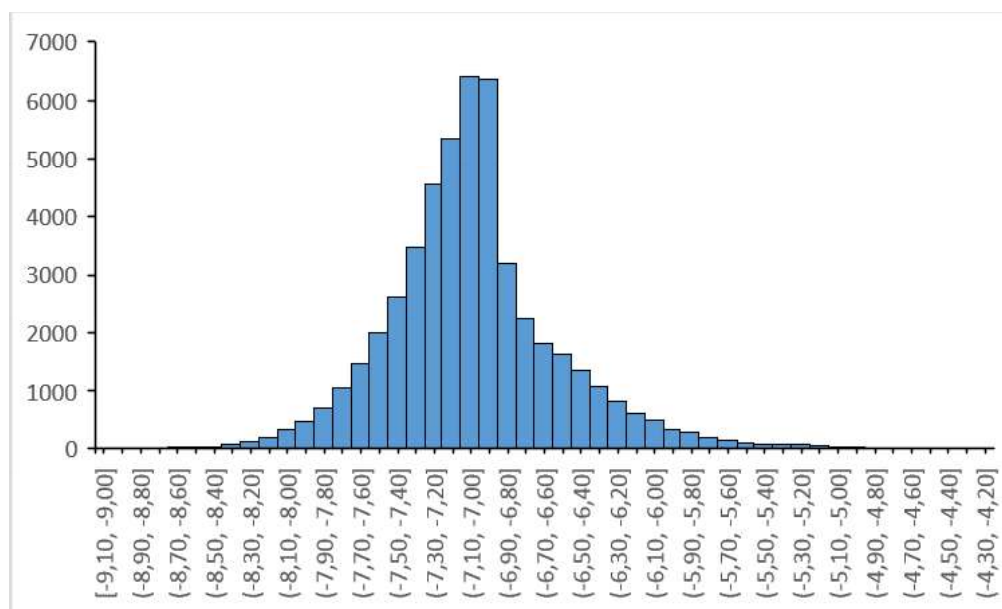


Figure 3.9: Distribution of AutoDock Vina docking scores (x-axis) of 50.000 ligands (y-axis) against Spike RBD binding site of human ACE2.

Moreover, top 5.000 ligands demonstrated docking range from -9,10 kcal/mol to -7,50 kcal/mol as seen in Figure 3.10. The highest scored ligands were also almost the same with the top results of initial docking run, indicating that top-docked ligands can always bind to target site with high affinity and there is high consistency between docking analyses.

Top 5.000 ligands with the highest docking score were prepared by LigPrep module of Maestro. Total 7.419 stereoisomers of top 5.000 ligands were generated. The most favorable and biochemically sensible conformation with low energy were identified for each ligand based on their low state penalty and ionization penalty. Almost all of the most favorable stereoisomers of 5.000 ligands showed ionization and state penalty closer to 0, indicating that ligands used for Glide XP were chemically stable and sensible; and they represented the most probable conformation of a particular ligand being closer to real biological system.

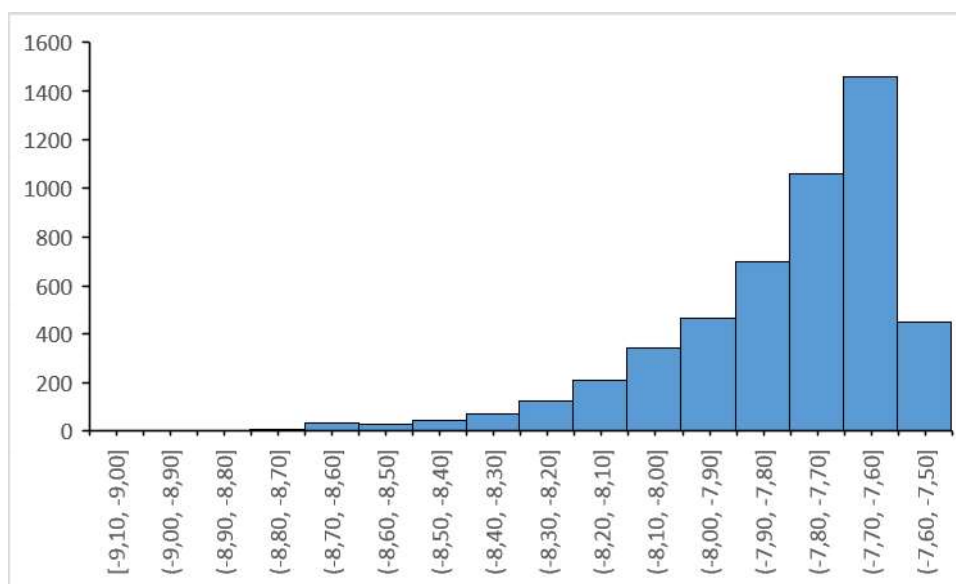


Figure 3.10: Distribution of AutoDock Vina docking scores (x-axis) of 5.000 ligands (y-axis) against Spike RBD binding site of human ACE2.

After Glide XP docking of 5.000 ligands, ligands were scored and ranked based on their Glide XP binding energy. Docking score was taken into consideration instead of Glide score as it also includes Epik penalty values. Thus, Epik penalty-added final docking scores accurately reflected true docking score of ligands by excluding energetically unfavourable states. As seen in Figure 3.11, Glide docking scores of small-molecule ligands were in the range from -6,410 kcal/mol to 0,882 kcal/mol. Additionally, the top 200 ligands showed a docking score range from -6,410 kcal/mol to -4,383 kcal/mol. Since Glide XP used a greatly extensive, different scoring function and Epik penalties were added to exclude energetically unlikely states, docking scores were lower compared to AutoDock Vina docking scores as expected. In Glide XP calculations, many researches considered docking scores less than -5,5 kcal/mol good and the ones less than -8 kcal/mol great. Thus, it could be stated that majority of top ligands docked here had good docking scores and could bind the target protein with a high affinity. Thus, it could be stated that majority of top ligands docked here had great docking scores and could bind the target protein with a high affinity. Moreover, the ligands with highest docking score were also mostly similar to the ones with highest AutoDock Vina docking scores. This finding further verifies that there is high consistency between each docking steps and the screened top 200 ligands could represent potential small-molecule inhibitors with high binding affinity to Spike RBD binding site of human ACE2.

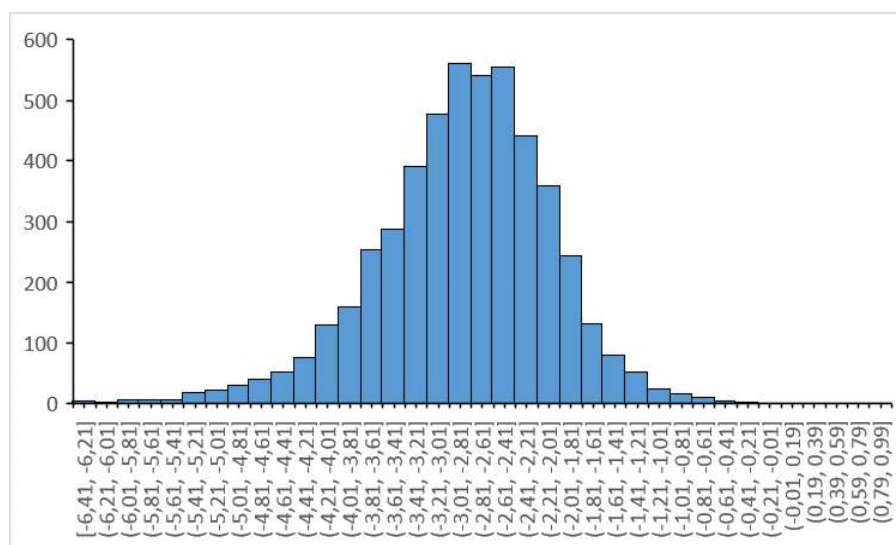


Figure 3.11: Distribution of docking scores (x-axis) of 5.000 ligands y-axis) against Spike RBD binding site of human ACE2 from Glide XP docking simulation.

After performing the dendritic fingerprinting hierarchical clustering, cut-off “Merging Distance” was set to be 0.7 and 166 different clusters were generated as seen in Figure A.2 in Appendix A. As a result of analysis of each cluster based on the structural and chemical similarities of cluster members, it was concluded that each cluster was highly unique as members of clusters are similar within clusters and different between clusters. This can be seen in Figure 3.12 which demonstrates the in-cluster similarity and between-clusters differences for 3 randomly selected clusters.

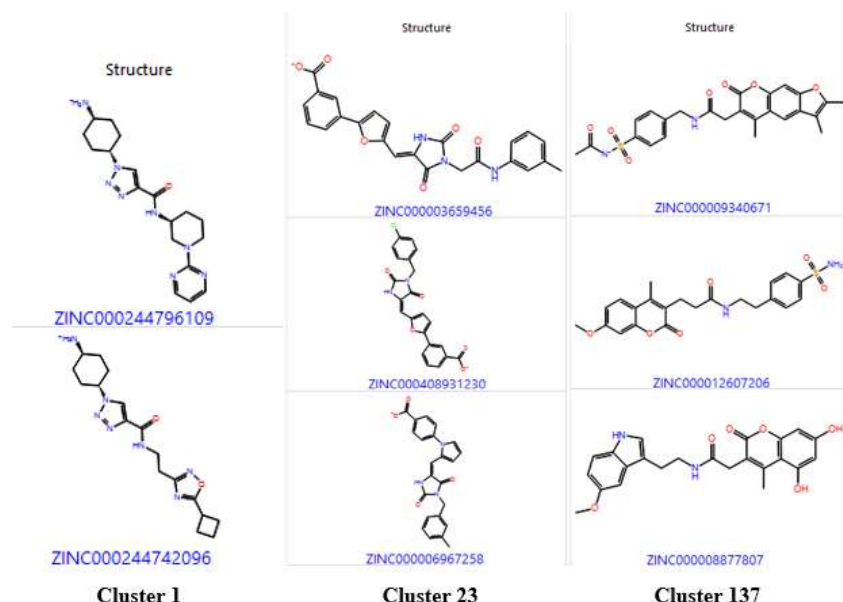


Figure 3.12: Three clusters from dendritic fingerprinting hierarchical clustering of 200 ligands against Spike RBD binding site of human ACE2, identified by CANVAS Software.

For each cluster, the member with the greatest docking score and lowest state penalty was selected as the representative of its respective cluster. These 166 ligands were then screened for ADME/T analysis to assess their pharmaceutical properties and toxicity. 100 molecules with the most optimum ADME/T properties were identified. Among these most drug-like, lower toxicity, low energy and high binding-affinity 100 ligands, top-scored 10 ligands were selected for further steps. Their state penalty, docking scores and ADME/T values can be seen in Table 3.3. All values were in overall great range and this indicates that these ligands can show proper pharmaceutical properties and this procedure might predict drug-like small-molecule ligands against ACE2 binding site on Spike RBD.

Table 3.3: Low-energy and high binding-affinity of 10 ligands with the lowest state penalty, the highest docking scores and optimum ADME/T values against RBD binding site of human ACE2.

Important Values	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00
	0014879	822	0000813	934	0.0000	0.3547	0.0485	0.2767	0.1932	0.0485	0.0983	0.0670	0.0564						
State Penalty	0.0234	0.0000	0.3547	0.0485	0.2767	0.1932	0.0485	0.0983	0.0670	0.0564									
Final AutoDock																			
Vina Score (kcal/mol)	-7.700	-7.500	-7.600	-7.700	-7.900	-7.700	-7.600	-7.900	-7.700	-7.600	-7.900	-7.700	-7.600	-7.600	-7.700	-7.900	-7.700	-7.600	-7.600
Glide XP Docking																			
Score (kcal/mol)	-6.410	-6.398	-6.324	-6.268	-6.150	-6.138	-6.054	-5.971	-5.850	-5.711									
stars	0	1	1	0	0	0	0	0	0	0	0	1	0						
rtvFG	0	1	0	0	0	0	0	0	0	0	0	0	1						
mol_MW	462.566	374.395	412.536	428.530	407.410	409.484	441.572	431.420	468.554	367.450									
QPlogS	-5.736	-6.069	-6.657	-5.857	-6.456	-4.181	-5.956	-5.060	-5.613	-3.874									
CIQPlogS	-5.408	-5.393	-5.580	-5.714	-6.102	-4.275	-5.496	-4.641	-6.423	-3.028									

Table 3.3 (continued): Low-energy and high binding-affinity of 10 ligands with the lowest state penalty, the highest docking scores and optimum ADME/T values against RBD binding site of human ACE2.

Important Values	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00
	0014879	0000813	0019742	0222656	0090607	0019022	0020149	0013561	0057712	0534636							
	822	934	660	494	226	189	756	951	634	983							
QPPCaco	370.183	1,451.50	402.612	955.896	1,032.60	164.759	1,077.47	84.372	281.980	405.430							
		3			0		6										
QplogBB	-0.186	-0.660	-0.306	0.195	-0.421	-0.754	0.305	-0.595	-0.388	-0.083							
Metab	5	3	5	5	3	7	5	3	6	2							
HumanOralAbsorption	3	3	1	3	1	2	3	3	3	3							
PercentHumanOralAbsorption	89.994	100.000	100.000	100.000	100.000	83.244	100.000	78.437	100.000	87.553							
RuleOfFive	1	0	0	1	0	0	1	0	0	0							
RuleOfThree	1	1	1	1	1	1	1	0	0	0							

Top-scored ligands were analyzed for their interaction with ACE2 by Ligand Interaction Diagram module of Maestro Software. Ligands were observed to be properly docked onto Spike RBD binding site of ACE2 (Figure 3.13).

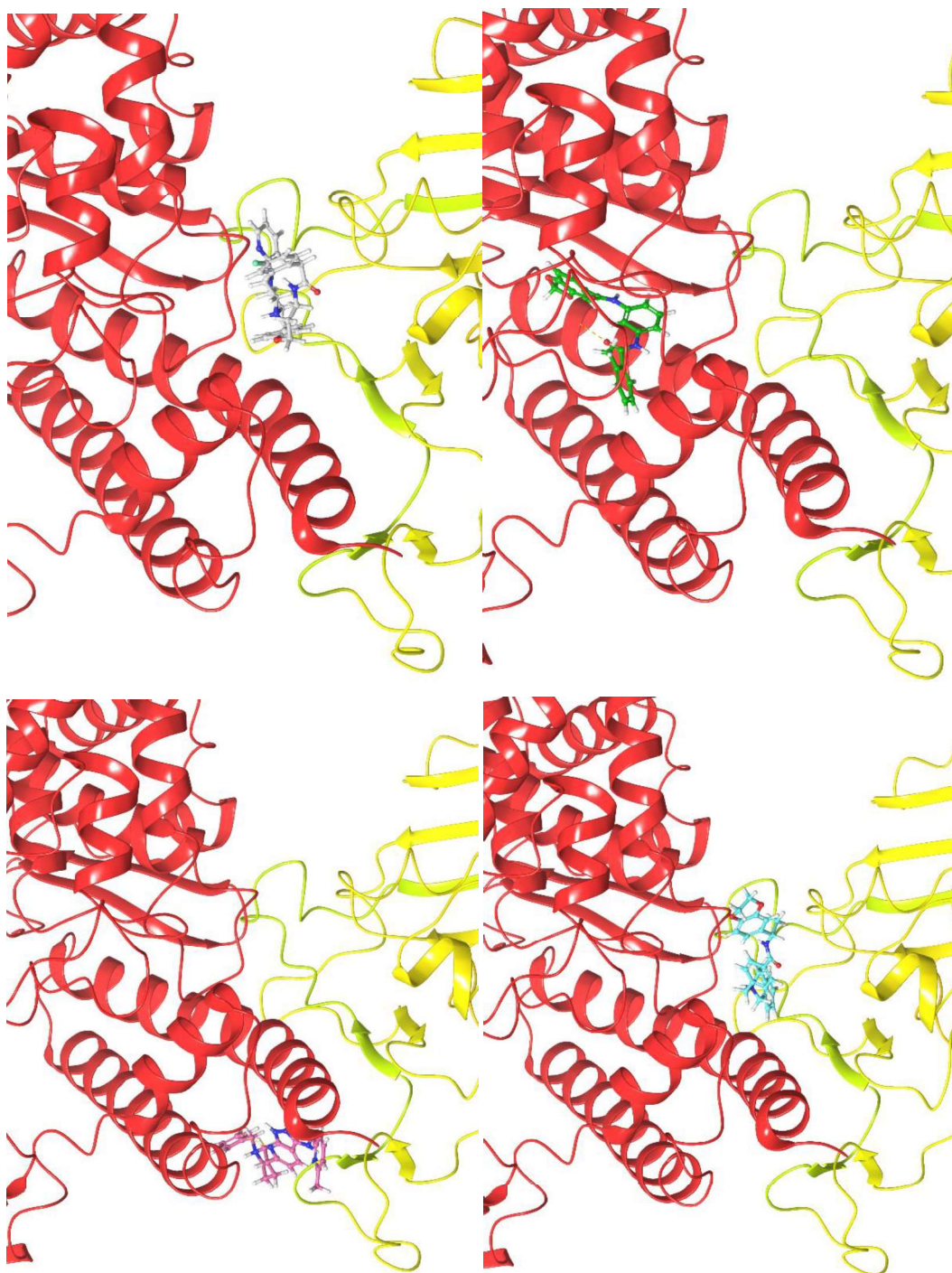


Figure 3.13: Top-scored 10 ligands in complex with human ACE2, visualized on Maestro. Red: ACE2. Yellow: Spike RBD. Note that only ACE2 was used in Glide XP. From left to right: ZINC000014879822, ZINC000000813934, ZINC000019742660, ZINC000222656494, ZINC000090607226, ZINC000019022189, ZINC000020149756, ZINC000013561951, ZINC000057712634, ZINC000534636983.

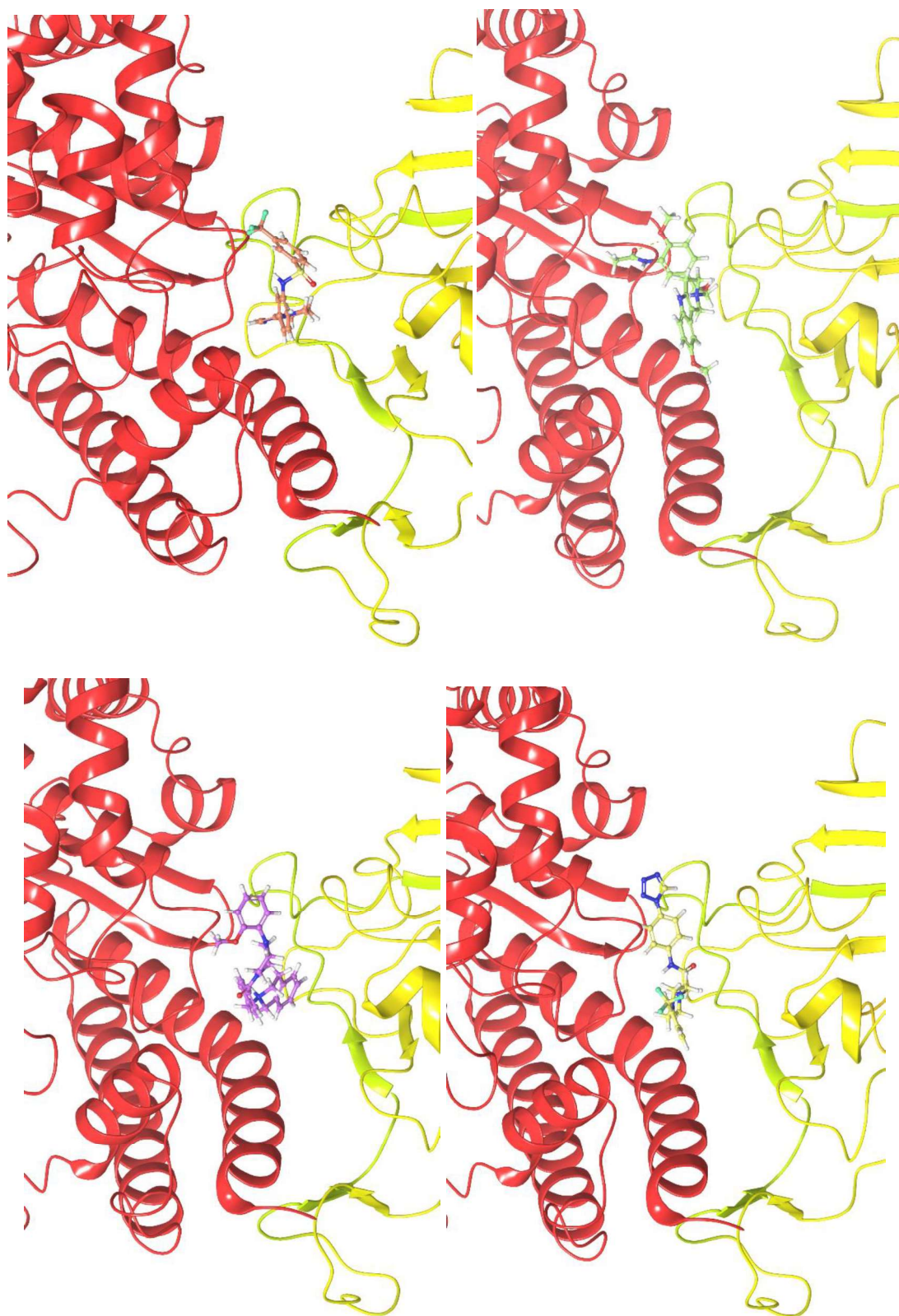


Figure 3.13 (continued): Top-scored 10 ligands in complex with human ACE2, visualized on Maestro. Red: ACE2. Yellow: Spike RBD. Note that only ACE2 was used in Glide XP. From left to right: ZINC000014879822, ZINC000000813934, ZINC000019742660, ZINC000222656494, ZINC000090607226, ZINC000019022189, ZINC000020149756, ZINC000013561951, ZINC000057712634, ZINC000534636983.

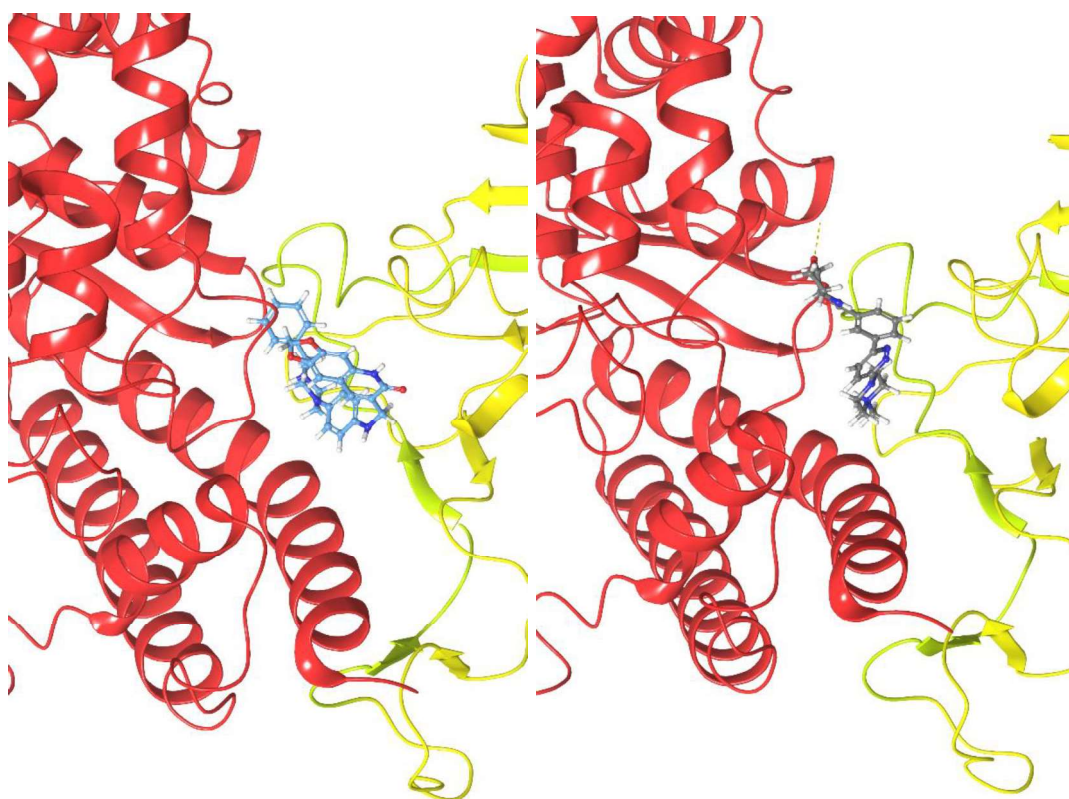


Figure 3.13 (continued): Top-scored 10 ligands in complex with human ACE2, visualized on Maestro. Red: ACE2. Yellow: Spike RBD. Note that only ACE2 was used in Glide XP. From left to right: ZINC000014879822, ZINC000000813934, ZINC000019742660, ZINC000222656494, ZINC000090607226, ZINC000019022189, ZINC000020149756, ZINC000013561951, ZINC000057712634, ZINC000534636983.

Moreover, a high number of hydrophobic interactions and hydrogen bonding was identified (Figure 3.14, Table 3.4), further verifying that these top-docked ligands can bind to ACE2 with high affinity and can block binding of other molecules. Major residues making hydrophobic interactions observed in ligands were Phe 28, Phe 32, Leu 39, Phe 40, Trp 69, Ala 71, Phe 72, Leu 73, Leu 79, Tyr 83, Ala 99, Ala 348, Trp 349, Tyr 385, Ala 386, Ala 387, Pro 389, Phe 390 and Leu 391. Major residues making hydrogen bonds observed in ligands were Glu 37, Glu 75, Gln 76, Asp 350, Gly 354, Ala 386, Ala 387, Phe 390, Asn 394 and Arg 559. Most of these residues were also the same with or next to critical ACE2 residues required for Spike RBD and ACE2 interaction such as Thr 27, Lys 31, Glu 37, Asp 38, Tyr 41, Tyr 83, Lys 353, Gly 354, Asp 355, Ala 387 and Arg 393. This highly suggests that the use of these ligands can block Spike RBD-ACE2 interaction as they can antagonistically overlap and bind to the same or close residues on binding interface with high affinity.

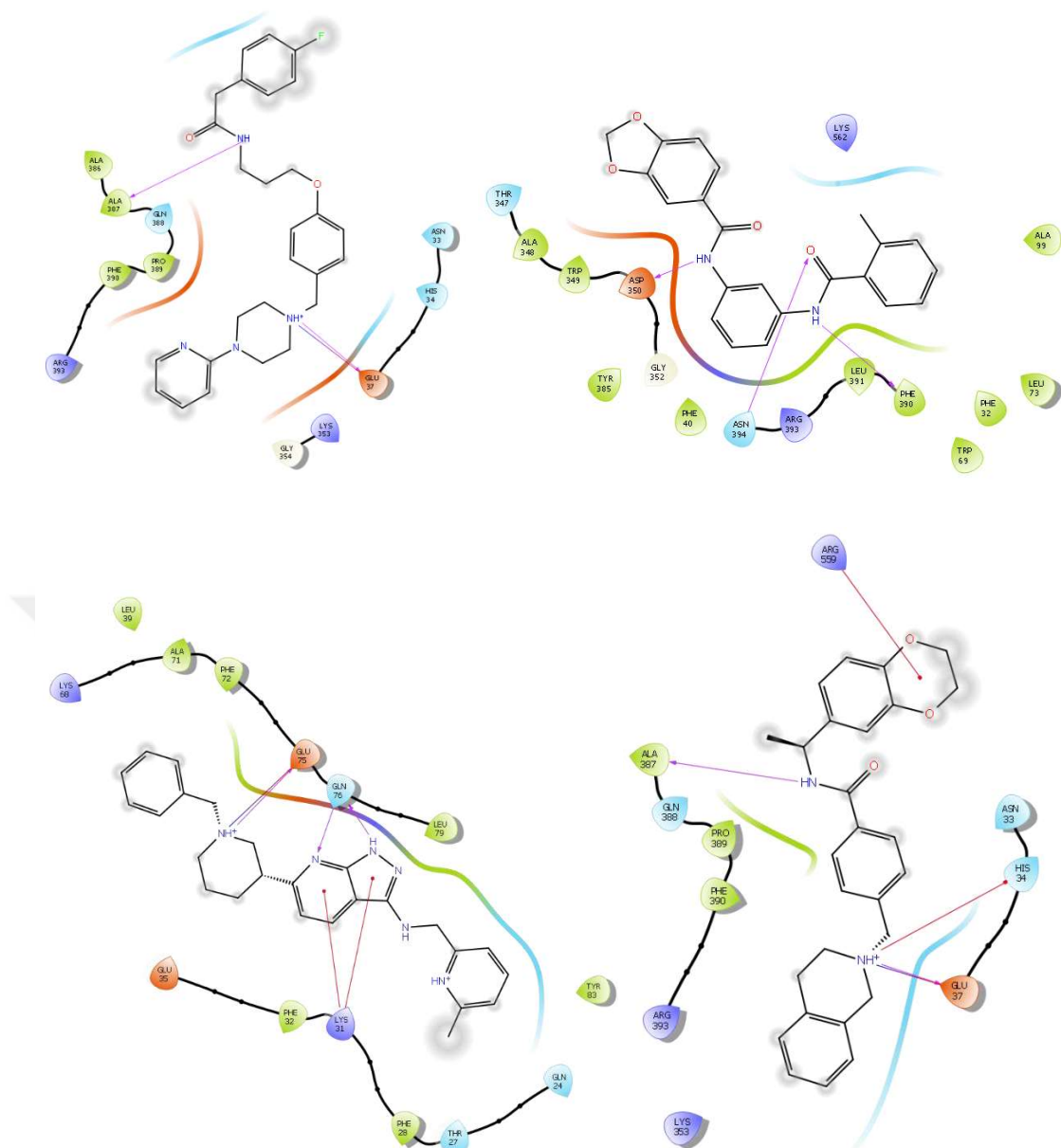


Figure 3.14: Ligand interaction diagrams of top-scored 10 ligands with human ACE2, visualized by Schrödinger Meastro. Red residue: Negatively charged, Purple residue: Positively charged, White residue: Glycine, Yellow residue: Hydrophobic, Blue residue: Polar, Purple arrow: H-bond, Orange arrow: Halogen Bond, Purple line: Metal coordination, Green line: Pi-Pi stacking, Grey circle: Solvent exposure. From left to right: ZINC000014879822, ZINC000000813934, ZINC000019742660, ZINC000222656494, ZINC000090607226, ZINC000019022189, ZINC000020149756, ZINC000013561951, ZINC000057712634, ZINC000534636983.

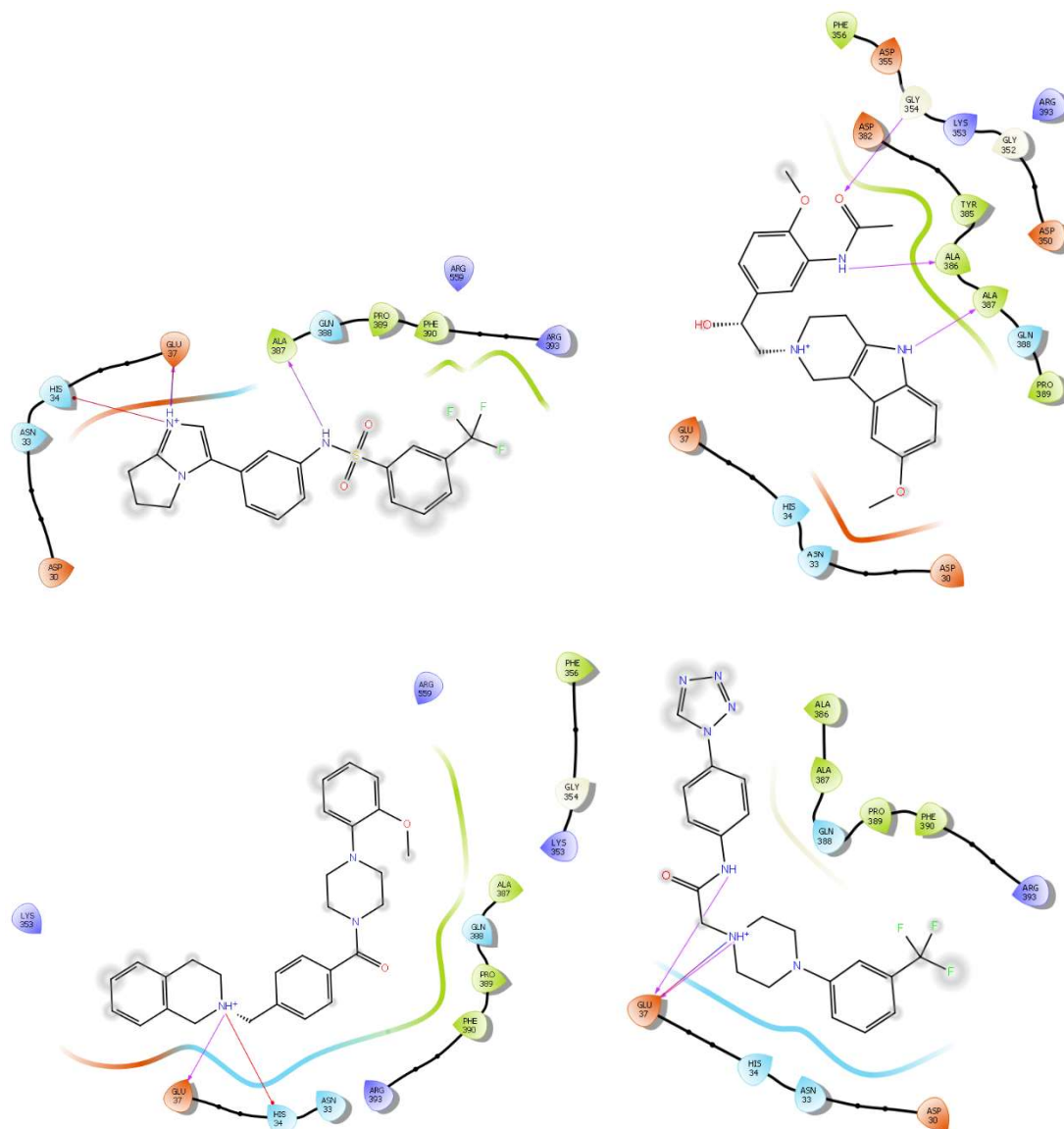


Figure 3.14 (continued): Ligand interaction diagrams of top-scored 10 ligands with human ACE2, visualized by Schrödinger Maestro. Red residue: Negatively charged, Purple residue: Positively charged, White residue: Glycine, Yellow residue: Hydrophobic, Blue residue: Polar, Purple arrow: H-bond, Orange arrow: Halogen Bond, Purple line: Metal coordination, Green line: Pi-Pi stacking, Grey circle: Solvent exposure. From left to right: ZINC000014879822, ZINC000000813934, ZINC000019742660, ZINC000222656494, ZINC000090607226, ZINC000019022189, ZINC000020149756, ZINC000013561951, ZINC000057712634, ZINC000534636983.

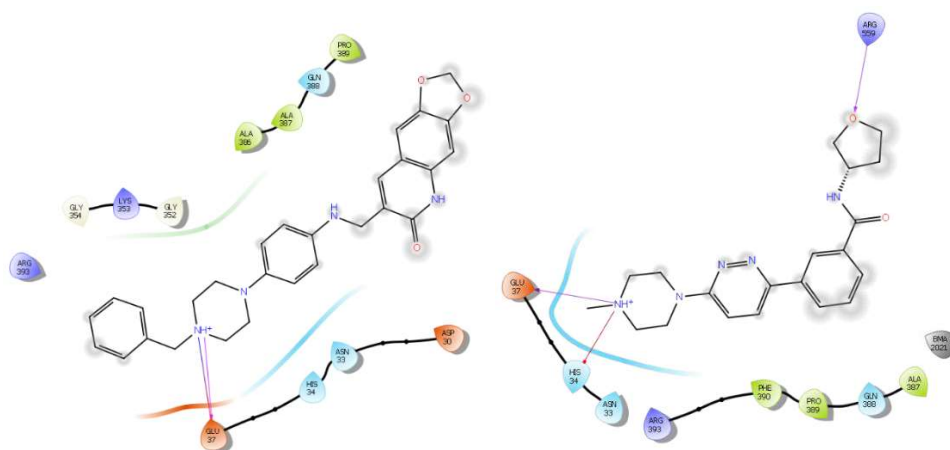


Figure 3.14 (continued): Ligand interaction diagrams of top-scored 10 ligands with human ACE2, visualized by Schrödinger Meastro. Red residue: Negatively charged, Purple residue: Positively charged, White residue: Glycine, Yellow residue: Hydrophobic, Blue residue: Polar, Purple arrow: H-bond, Orange arrow: Halogen Bond, Purple line: Metal coordination, Green line: Pi-Pi stacking, Grey circle: Solvent exposure. From left to right: ZINC000014879822, ZINC000000813934, ZINC000019742660, ZINC000222656494, ZINC000090607226, ZINC000019022189, ZINC000020149756, ZINC000013561951, ZINC000057712634, ZINC000534636983.

Table 3.4: Protein-ligand interaction profile of ACE2 and top-scored 10 small-molecule ligands.

Small-molecule Ligand	Hydrophobic Interactions	Hydrogen Bonds
ZINC000014879822	Ala 386, Ala 387, Pro 389, Phe 390	Glu 37, Ala 387
ZINC000000813934	Phe 32, Phe 40, Trp 69, Leu 73, Ala 99, Ala 348, Trp 349, Tyr 385, Phe 390, Leu 391	Asp 350, Phe 390, Asn 394
ZINC000019742660	Phe 28, Phe 32, Leu 39, Ala 71, Phe 72, Leu 79, Tyr 83	Glu 75, Gln 76
ZINC000222656494	Ala 387, Phe 390, Pro 389	Glu 37, Ala 387
ZINC000090607226	Ala 387, Pro 389, Phe 390	Glu 37, Ala 387
ZINC000019022189	Phe 356, Tyr 385, Ala 386, Ala 387, Pro 389	Gly 354, Ala 386, Ala 387
ZINC000020149756	Ala 387, Pro 389, Phe 390	Glu 37
ZINC000013561951	Phe 356, Ala 386, Ala 387, Pro 389, Phe 390	Glu 37
ZINC000057712634	Ala 386, Ala 387, Pro 389	Glu 37
ZINC000534636983	Ala 386, Ala 387, Pro 389	Glu 37, Arg 559

3.3. Small-molecule Inhibitor Prediction for Nsp16-Nsp10 Interaction

After initial docking of 400.000 ligands, ligands were scored and ranked based on their docking binding energy scores as seen in Figure 3.15. AutoDock Vina docking scores of ligands showed a docking score range from -10 kcal/mol to -3,30 kcal/mol. Docking scores were populated around approximately -6,59 kcal/mol. Additionally, the top 50.000 ligands showed a docking score range from -10.00 kcal/mol to -7,40 kcal/mol. These values indicate that the initial screen of randomized 400.000 ligand library gave a large number of ligands with great docking score and high binding affinity to Nsp10 binding site of Nsp16.

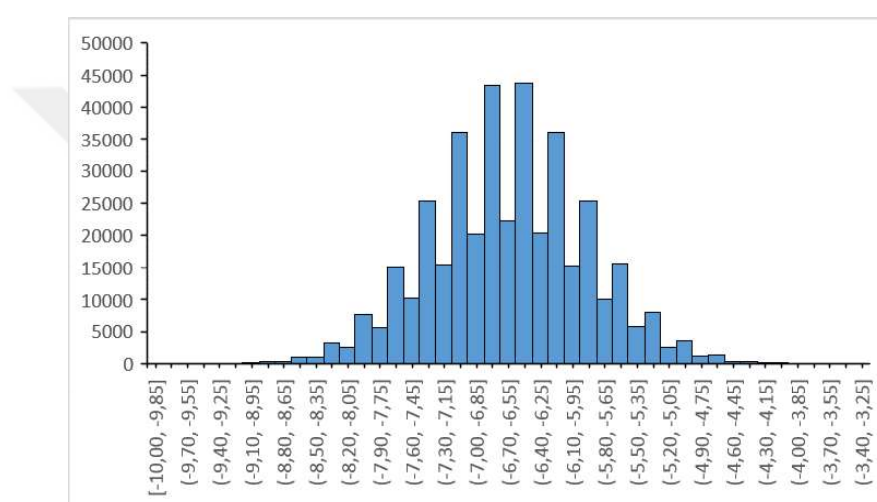


Figure 3.15: Distribution of AutoDock Vina docking scores (x-axis) of 400,000 ligands (y-axis) against Nsp10 binding site of Nsp16.

After scoring and sorting, second run of AutoDock Vina with top 50.000 ligands showed a ligand docking score range from -9,80 kcal/mol to -5,60 kcal/mol. Docking scores were populated around -7,75 kcal/mol which is better than the initial run of ligand samples as seen in Figure 3.16. This indicates that a second iteration of docking improved the docking results.

Moreover, top 5.000 ligands demonstrated docking range from -9,80 kcal/mol to -8,20 kcal/mol as seen in Figure 3.17. The highest scored ligands were also almost the same with the top results of initial docking run, indicating that top-docked ligands can always bind to target site with high affinity and there is high consistency between docking analyses.

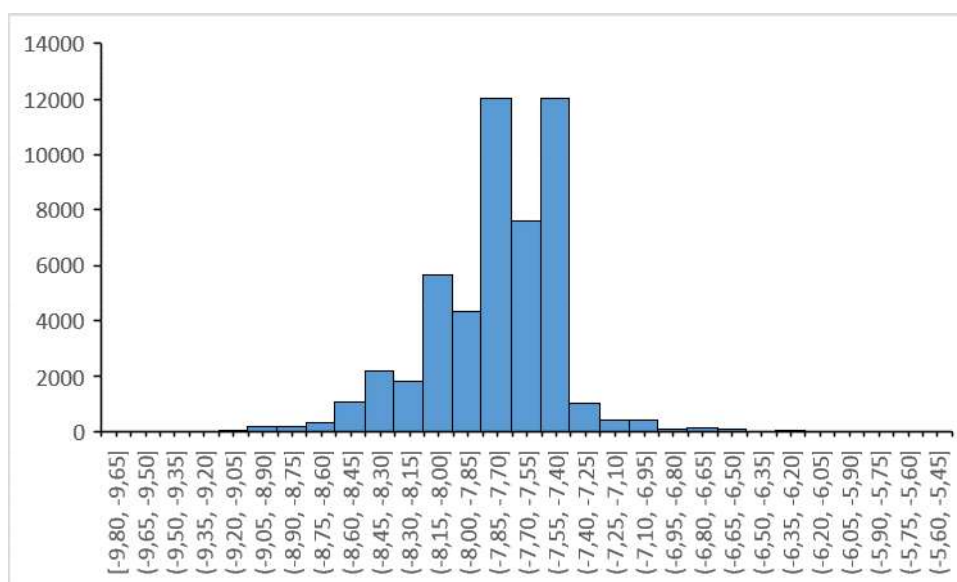


Figure 3.16: Distribution of AutoDock Vina docking scores (x-axis) of 50.000 ligands (y-axis) against Nsp10 binding site of Nsp16.

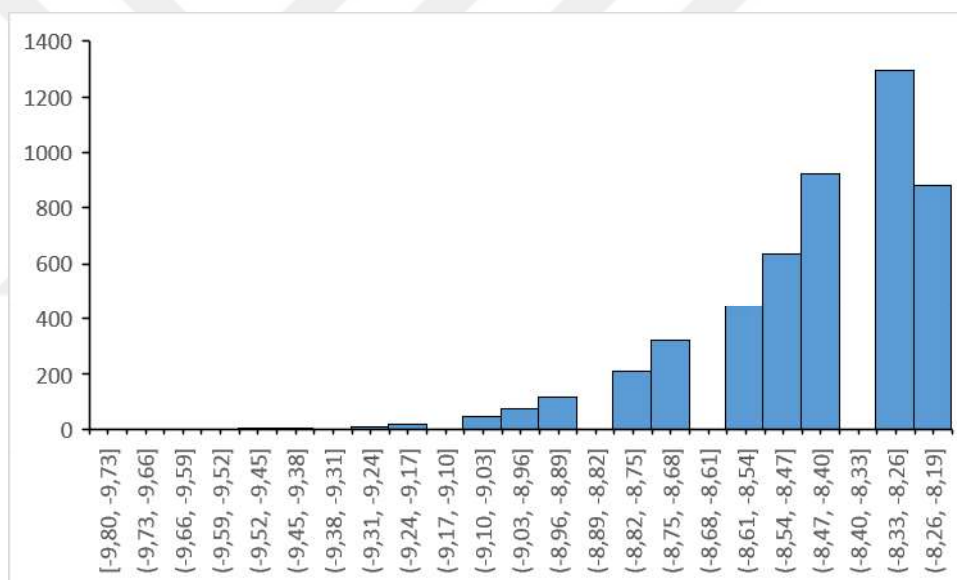


Figure 3.17: Distribution of AutoDock Vina docking scores (x-axis) of 5.000 ligands (y-axis) against Nsp10 binding site of human Nsp16.

Top 5.000 ligands with the highest docking score were prepared by LigPrep module of Maestro. Total 7.590 stereoisomers of top 5.000 ligands were generated. The most favorable and biochemically sensible conformation with low energy were identified for each ligand based on their low state penalty and ionization penalty. Almost all of the most favorable stereoisomers of 5.000 ligands showed ionization and state penalty closer to 0, indicating that ligands used for Glide XP were chemically stable and sensible; and they represented the most probable conformation of a particular ligand being closer to real biological system.

After Glide XP docking of 5.000 ligands, ligands were scored and ranked based on their Glide XP binding energy. Docking score was taken into consideration instead of Glide score as it also includes Epik penalty values. Thus, Epik penalty-added final docking scores accurately reflected true docking score of ligands by excluding energetically unfavourable states. As seen in Figure 3.18, Glide docking scores of small-molecule ligands were in the range from -8,479 kcal/mol to 1,912 kcal/mol. Additionally, the top 200 ligands showed a docking score range from -8,479 kcal/mol to -5,387 kcal/mol. Since Glide XP used a greatly extensive, different scoring function and Epik penalties were added to exclude energetically unlikely states, docking scores were slightly lower compared to AutoDock Vina docking scores as expected. In Glide XP calculations, many researches considered docking scores less than -5,5 kcal/mol good and the ones less than -8 kcal/mol great. Thus, it could be stated that majority of top ligands docked here had great docking scores and could bind the target protein with a high affinity. Moreover, the ligands with highest docking score were also mostly similar to the ones with highest AutoDock Vina docking scores. This finding further verifies that there is high consistency between each docking steps and the screened top 200 ligands could represent potential small-molecule inhibitors with high binding affinity to Nsp10 binding site of Nsp16.

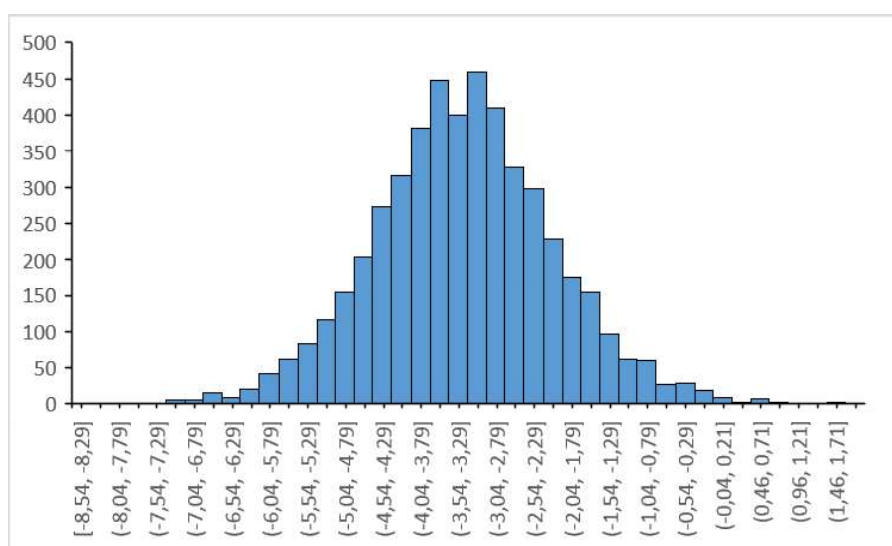


Figure 3.18: Distribution of docking scores (x-axis) of 5.000 ligands (y-axis) against Nsp10 binding site of Nsp16 from Glide XP docking simulation.

After performing the dendritic fingerprinting hierarchical clustering, cut-off “Merging Distance” was set to be 0.6 and 161 different clusters were generated as seen in Figure A.3 in Appendix A. As a result of analysis of each cluster based on the structural and

chemical similarities of cluster members, it was concluded that each cluster was highly unique as members of clusters are similar within clusters and different between clusters. This can be seen in Figure 3.19 which demonstrates the in-cluster similarity and between-clusters differences for 3 randomly selected clusters.

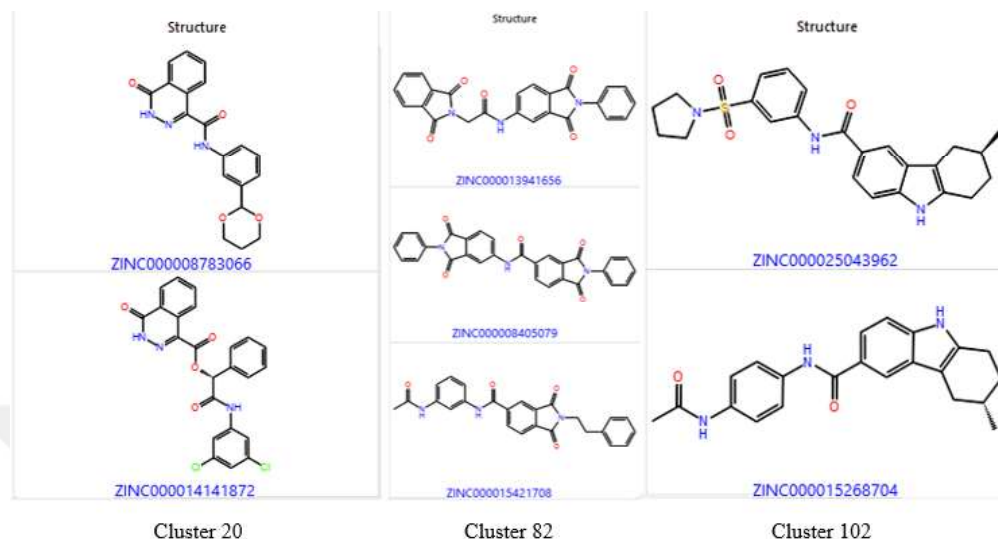


Figure 3.19: Three clusters from dendritic fingerprinting hierarchical clustering of 200 ligands against Nsp10 binding site of Nsp16, identified by CANVAS Software.

For each cluster, the member with the greatest docking score and lowest state penalty was selected as the representative of its respective cluster. These 161 ligands were then screened for ADME/T analysis to assess their pharmaceutical properties and toxicity. 100 molecules with the most optimum ADME/T properties were identified. Among these most drug-like, lower toxicity, low energy and high binding-affinity 100 ligands, top-scored 10 ligands were selected for further steps. Their state penalty, docking scores and ADME/T values can be seen in Table 3.5. All values were in overall great range and this indicates that these ligands can show proper pharmaceutical properties and this procedure might predict drug-like small-molecule ligands against Nsp10 binding site on Nsp16.

Table 3.5: Low-energy and high binding-affinity of 10 ligands with the lowest state penalty, the highest docking scores and optimum ADME/T values against Nsp10 binding site of Nsp16.

Important Values	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00
0009751	0002286	0009252	0019497	0009485	0055075	0000308	0019289	0044872	0001163										
488	111	898	136	446	822	303	953	049	520										
State Penalty	0.0000	0.1169	0.0000	0.0000	0.0000	0.0004	0.0000	0.2936	0.0284	0.4519									
Final AutoDock Vina																			
Score (kcal/mol)	-8.200	-8.300	-9.100	-9.800	-8.300	-8.500	-8.200	-8.300	-8.400	-8.200									
Glide XP Docking																			
Score (kcal/mol)	-7.087	-6.981	-6.857	-6.757	-6.642	-6.581	-6.480	-6.383	-6.315	-6.241									
stars	1	0	0	0	0	0	0	0	0	0									
rtvFG	0	0	0	0	0	1	0	0	0	0									
mol_MW	446.526	466.898	493.579	443.544	488.496	366.416	335.359	425.532	463.593	361.759									
QPlogS	-6.892	-5.367	-4.534	-5.741	-5.976	-4.957	-5.222	-4.383	-5.033	-5.964									

Table 3.5 (continued): Low-energy and high binding-affinity of 10 ligands with the lowest state penalty, the highest docking scores and optimum ADME/T values against Nsp10 binding site of Nsp16.

Important Values	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00
	0009751	0002286	0009252	0019497	0009485	0055075	000308	0019289	0044872	0001163							
	488	111	898	136	446	822	303	953	049	520							
CIQlogS	-6.313	-6.115	-5.449	-5.526	-6.524	-5.232	-5.360	-4.168	-5.077	-5.891							
QPPCaco	303.717	160.537	68.856	219.866	491.499	1032.39	5018.32	187.895	198.932	248.970							
QplogBB	-1.642	-1.561	-2.036	-1.220	-1.456	-0.419	0.131	-0.362	-0.806	-0.802							
Metab	2	2	2	4	6	4	2	6	5	2							
HumanOralAbsorption	1	3	3	3	3	3	3	3	3	3							
PercentHumanOralAbsorption	95.375	81.528	73.262	91.452	96.817	100.000	100.000	88.274	89.531	87.889							
RuleOffFive	0	0	0	0	0	0	0	0	0	0							
RuleOfThree	1	0	0	1	1	0	0	0	0	0							

Top-scored ligands were analyzed for their interaction with Nsp16 by Ligand Interaction Diagram module of Maestro Software. Ligands were observed to be properly docked onto Nsp10 binding site of Nsp16 (Figure 3.20).

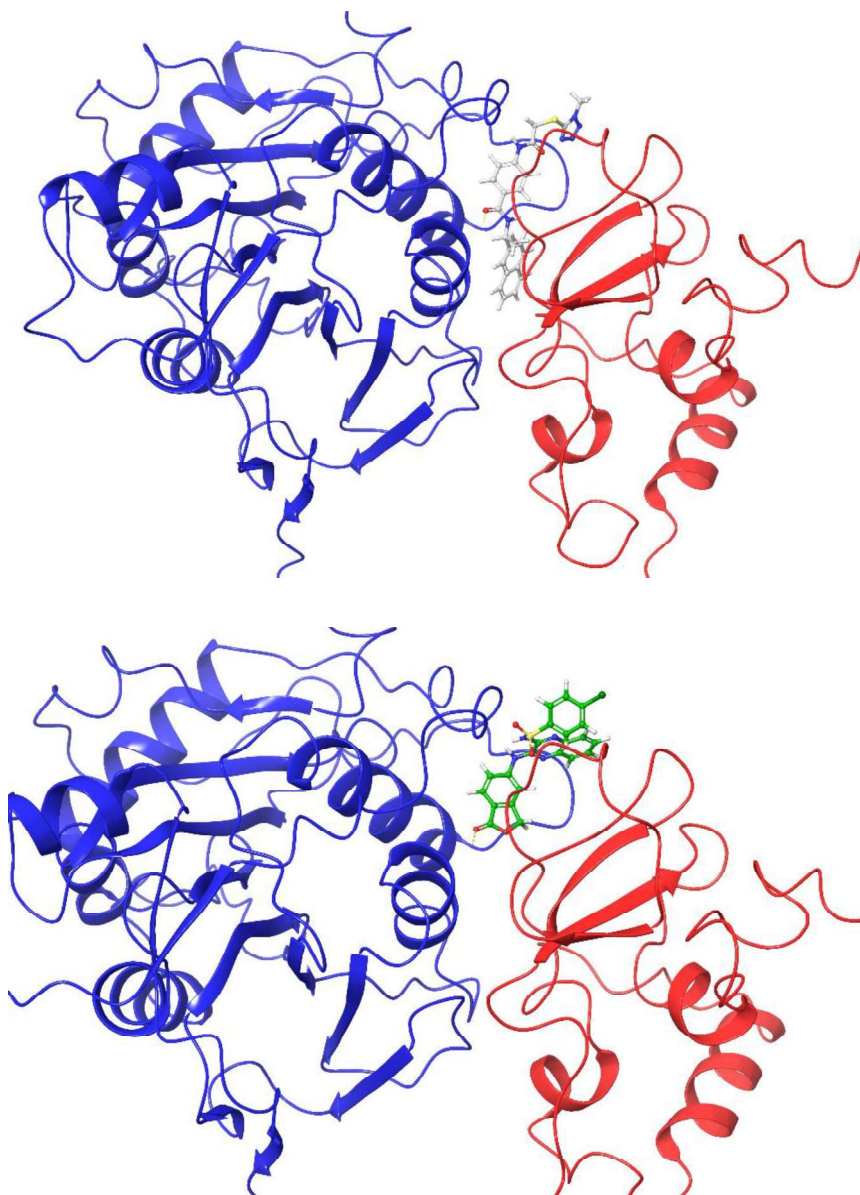


Figure 3.20: Top-scored 10 ligands within the Nsp16-Nsp10 binding interface, visualized on Maestro. Note that only Nsp16 was used in Glide XP. Purple: Nsp16, Red: Nsp 10. From top to bottom: ZINC000009751488, ZINC000002286111, ZINC000009252898, ZINC000019497136, ZINC000009485446, ZINC000055075822, ZINC000000308303, ZINC000019289953, ZINC000044872049, ZINC000001163520.

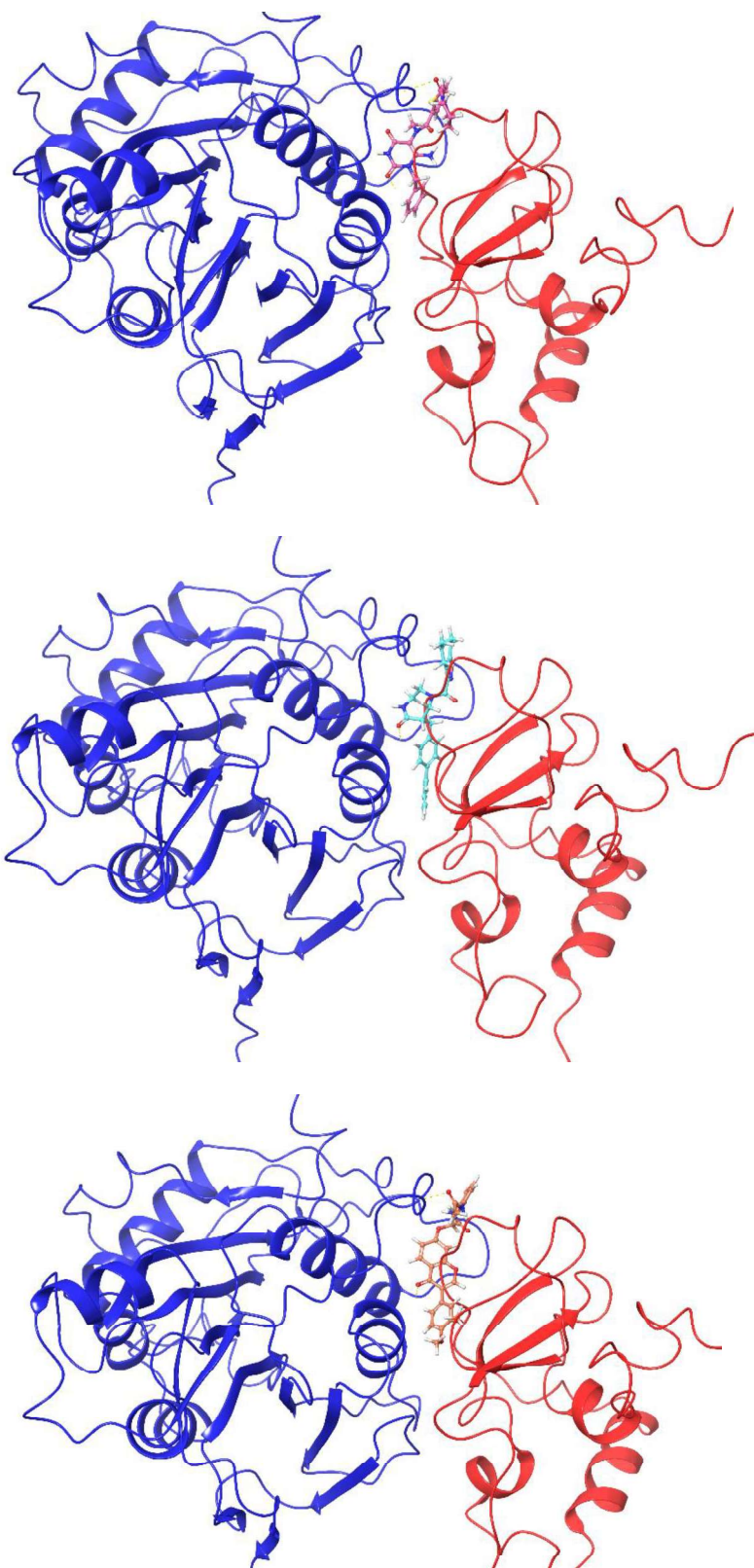


Figure 3.20 (continued): Top-scored 10 ligands within the Nsp16-Nsp10 binding interface, visualized on Maestro. Note that only Nsp16 was used in Glide XP. Purple: Nsp16, Red: Nsp 10. From top to bottom: ZINC000009751488, ZINC000002286111, ZINC000009252898, ZINC000019497136, ZINC000009485446, ZINC000055075822, ZINC000000308303, ZINC000019289953, ZINC000044872049, ZINC000001163520.

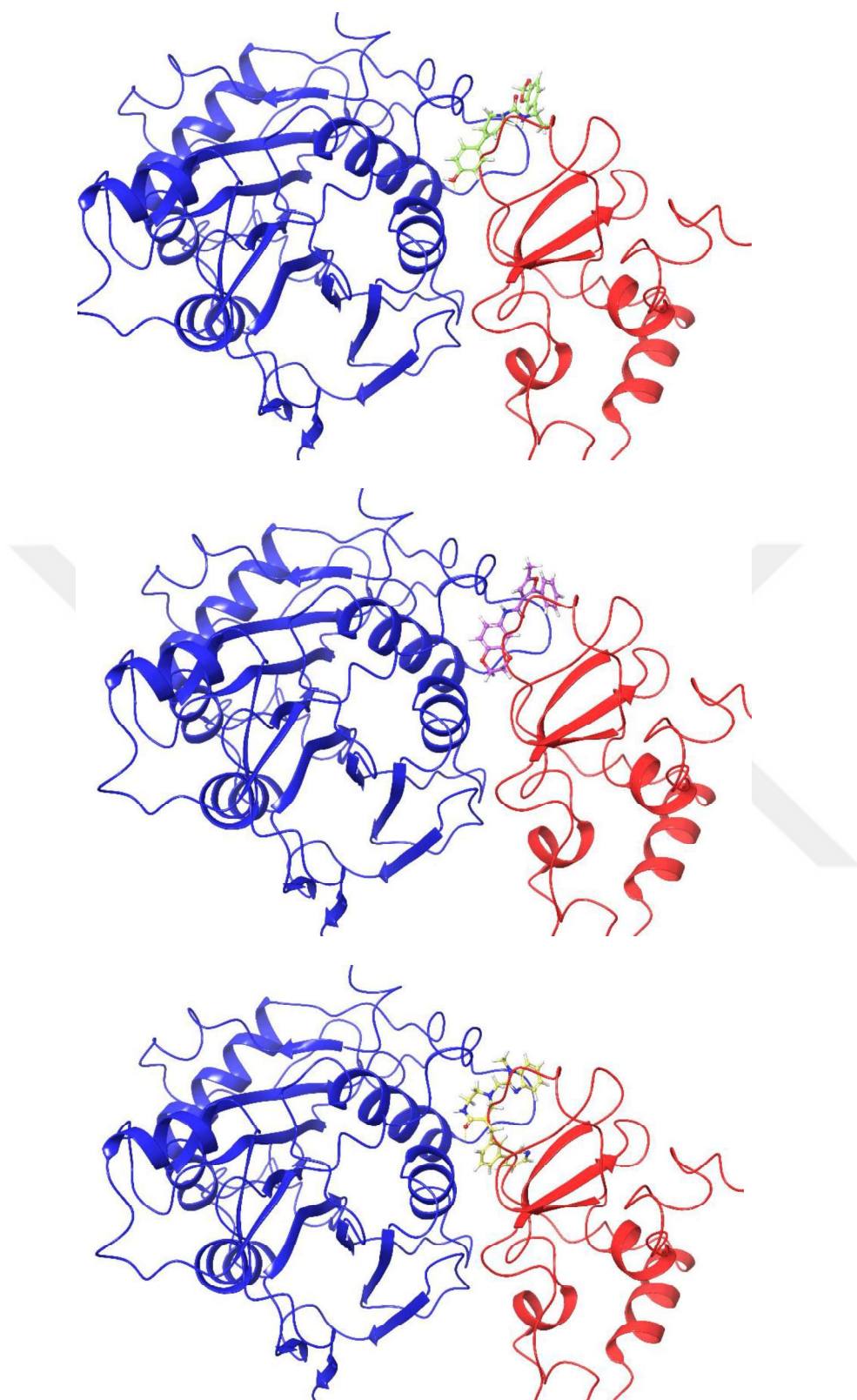


Figure 3.20 (continued): Top-scored 10 ligands within the Nsp16-Nsp10 binding interface, visualized on Maestro. Note that only Nsp16 was used in Glide XP. Purple: Nsp16, Red: Nsp 10. From top to bottom: ZINC000009751488, ZINC000002286111, ZINC000009252898, ZINC000019497136, ZINC000009485446, ZINC000055075822, ZINC000000308303, ZINC000019289953, ZINC000044872049, ZINC000001163520.

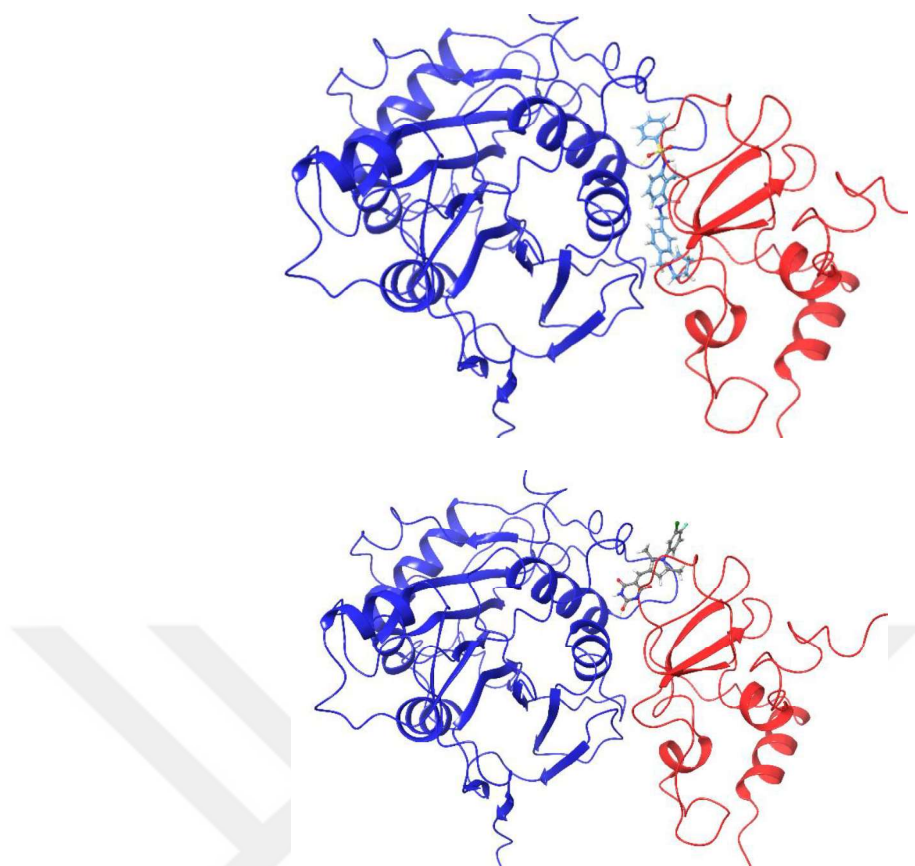


Figure 3.20 (continued): Top-scored 10 ligands within the Nsp16-Nsp10 binding interface, visualized on Maestro. Note that only Nsp16 was used in Glide XP. Purple: Nsp16, Red: Nsp 10. From top to bottom: ZINC000009751488, ZINC000002286111, ZINC000009252898, ZINC000019497136, ZINC000009485446, ZINC000055075822, ZINC000000308303, ZINC000019289953, ZINC000044872049, ZINC000001163520.

Moreover, a high number of hydrophobic interactions and hydrogen bonding was identified (Figure 3.21, Table 3.6), further verifying that these top-docked ligands can bind to Nsp10 binding site of Nsp16 with high affinity and can block binding of Nsp10. Major residues making hydrophobic interactions observed in ligands were Pro 6835, Ile 6838, Met 6839, Val 6842, Ala 6843, Val 6876, Ala 6877, Pro 6878, Ala 6881, Val 6882, Leu 7042, Phe 7043, Leu 7042 and Met 7045. Major residues making hydrogen bonds observed in ligands were Lys 6836, Thr 6846, Gln 6885, Asp 6904, Leu 7042, Leu 7045 and Ser 7046. Most of these residues were also the same with or next to critical residues required for Nsp16-Nsp10 interaction such as Pro 6835, Ile 6838, Met 6839, Val 6842, Thr 6846, Ala 6871, Val 6876, Pro 6878, Val 6882, Glu 6885, Leu 7042 and Met 7048. This highly suggests that the use of these ligands can block Nsp16-Nsp10 interaction as they can antagonistically overlap and bind to the same or close residues on binding interface with high affinity.

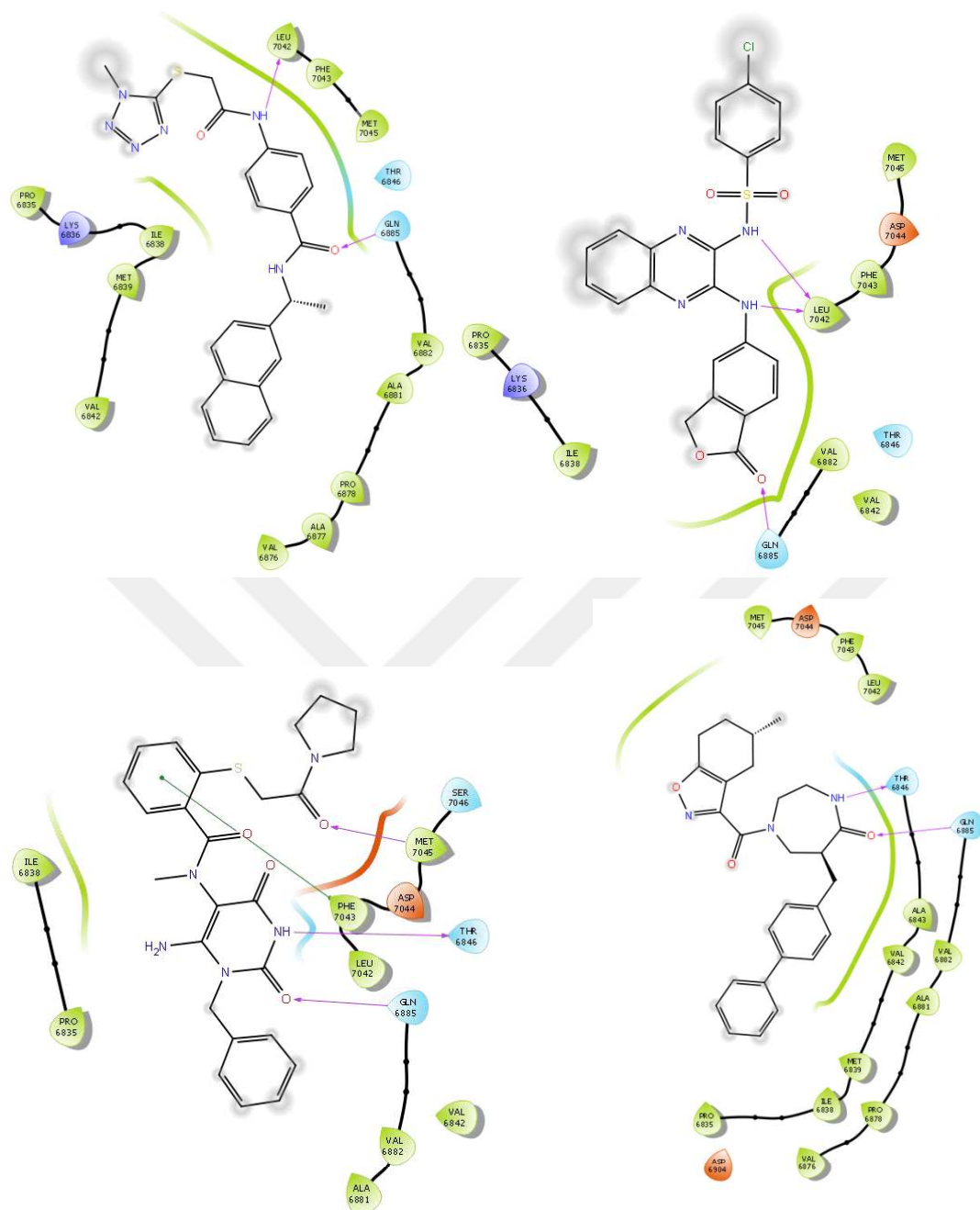


Figure 3.21: Ligand interaction diagrams of top-scored 10 ligands with Nsp16, visualized by Schrödinger Maestro. Red residue: Negatively charged, Purple residue: Positively charged, White residue: Glycine, Yellow residue: Hydrophobic, Blue residue: Polar, Purple arrow: H-bond, Orange arrow: Halogen Bond, Purple line: Metal coordination, Green line: Pi-Pi stacking, Grey circle: Solvent exposure. From left to right: ZINC000009751488, ZINC000002286111, ZINC000009252898, ZINC000019497136, ZINC000009485446, ZINC000055075822, ZINC000000308303, ZINC000019289953, ZINC000044872049, ZINC000001163520.

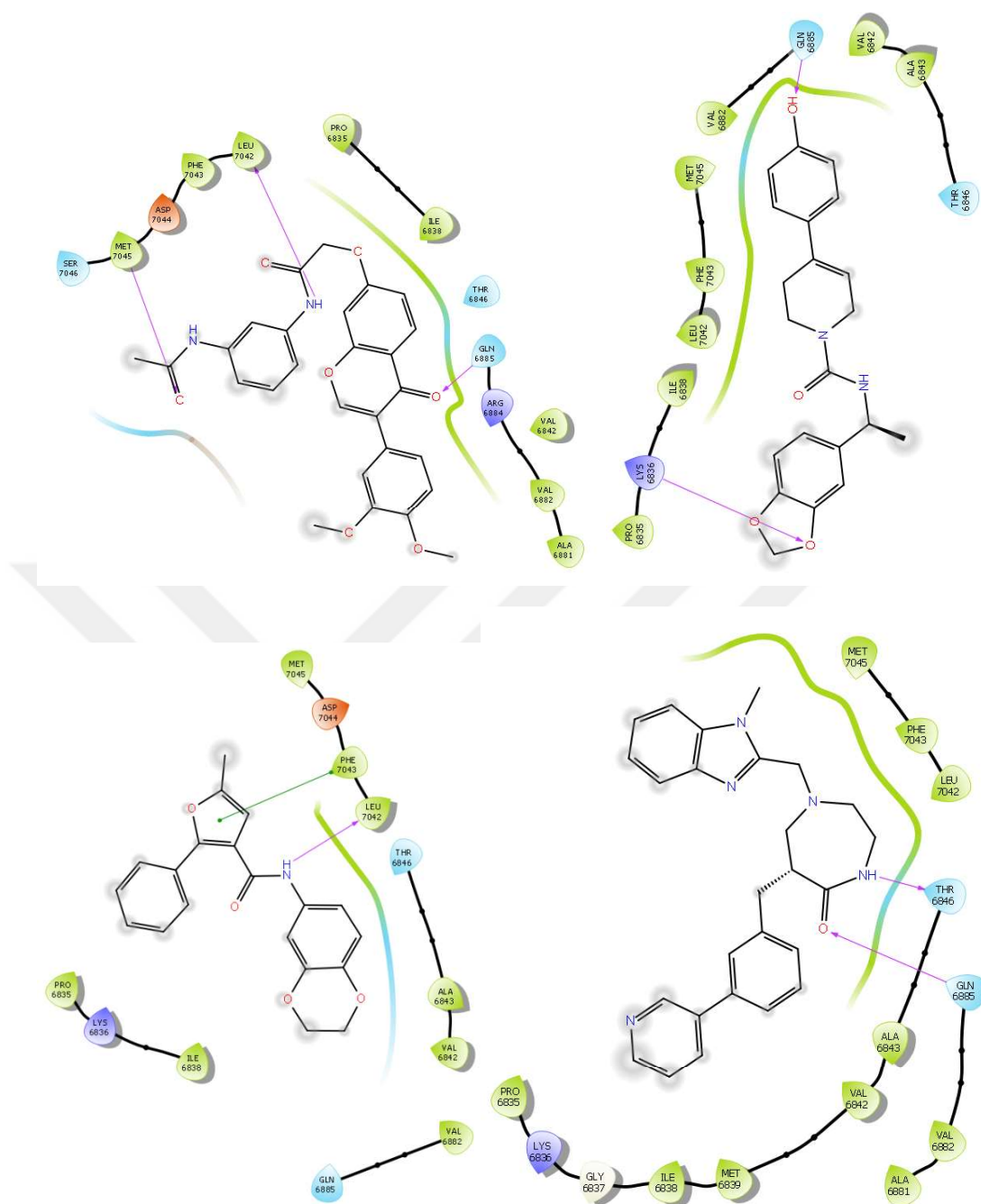


Figure 3.21 (continued): Ligand interaction diagrams of top-scored 10 ligands with Nsp16, visualized by Schrödinger Maestro. Red residue: Negatively charged, Purple residue: Positively charged, White residue: Glycine, Yellow residue: Hydrophobic, Blue residue: Polar, Purple arrow: H-bond, Orange arrow: Halogen Bond, Purple line: Metal coordination, Green line: Pi-Pi stacking, Grey circle: Solvent exposure. From left to right: ZINC000009751488, ZINC000002286111, ZINC000009252898, ZINC000019497136, ZINC000009485446, ZINC000055075822, ZINC000000308303, ZINC000019289953, ZINC000044872049, ZINC000001163520.

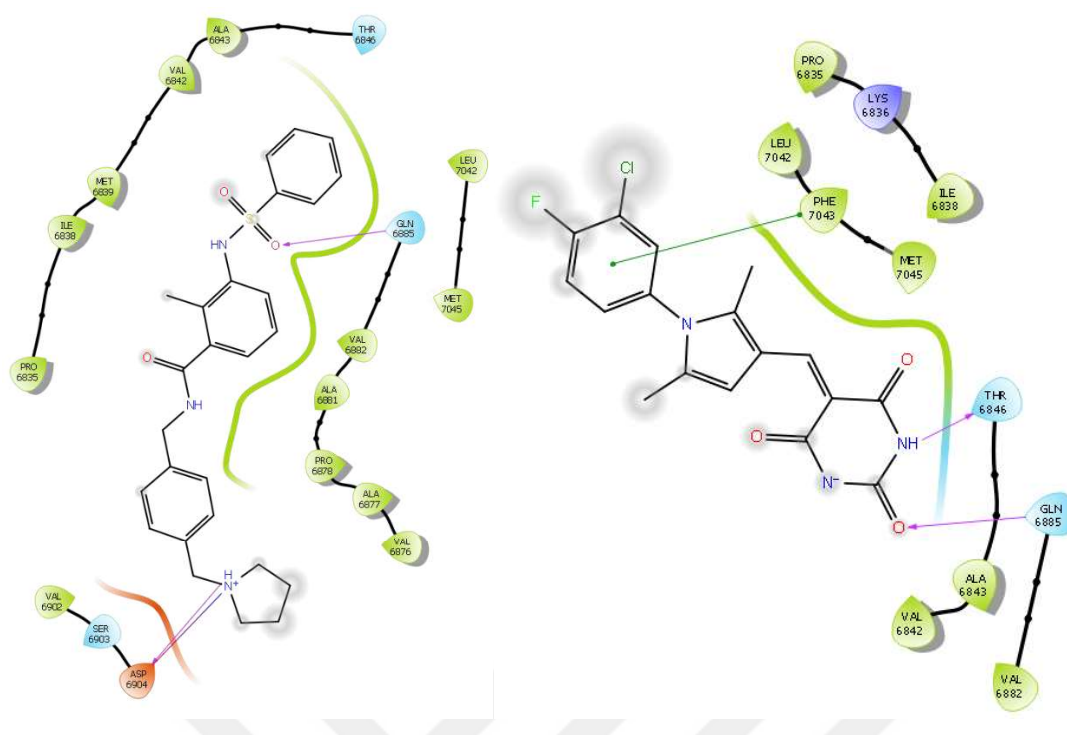


Figure 3.21 (continued): Ligand interaction diagrams of top-scored 10 ligands with Nsp16, visualized by Schrödinger Maestro. Red residue: Negatively charged, Purple residue: Positively charged, White residue: Glycine, Yellow residue: Hydrophobic, Blue residue: Polar, Purple arrow: H-bond, Orange arrow: Halogen Bond, Purple line: Metal coordination, Green line: Pi-Pi stacking, Grey circle: Solvent exposure. From left to right: ZINC000009751488, ZINC000002286111, ZINC000009252898, ZINC000019497136, ZINC000009485446, ZINC000055075822, ZINC000000308303, ZINC000019289953, ZINC000044872049, ZINC000001163520.

Table 3.6: Protein-ligand interaction profile of Nsp16 for its Nsp10 binding site and top-scored 10 small-molecule ligands.

Small-molecule Ligand	Hydrophobic Interactions	Hydrogen Bonds
ZINC000009751488	Pro 6835, Ile 6838, Met 6839, Val 6842, Val 6876, Ala 6877, Pro 6878, Ala 6881, Val 6882, Leu 7042, Phe 7043, Met 7045	Gln 6885, Leu 7042
ZINC000002286111	Pro 6835, Ile 6838, Val 6842, Val 6882, Leu 7042, Phe 7043, Met 7045	Gln 6885, Leu 7042
ZINC000009252898	Pro 6835, Ile 6838, Val 6842, Ala 6881, Val 6882, Leu 7042, Phe 7043, Met 7045	Thr 6846, Gln 6885, Ser 7046

Table 3.6 (continued): Protein-ligand interaction profile of Nsp16 for its Nsp10 binding site and top-scored 10 small-molecule ligands.

Small-molecule Ligand	Hydrophobic Interactions	Hydrogen Bonds
ZINC000019497136	Pro 6835, Ile 6838, Met 6839, Val 6842, Ala 6843, Val 6876, Pro 6878, Ala 6881, Val 6882, Leu 7042, Phe 7043, Met 7045	Thr 6846, Gln 6885
ZINC000009485446	Pro 6835, Ile 6838, Val 6842, Ala 6881, Val 6882, Leu 7042, Phe 7043, Met 7045	Gln 6885, Leu 7042, Met 7045
ZINC000055075822	Pro 6835, Ile 6838, Val 6842, Ala 6843, Val 6882, Leu 7042, Phe 7043, Met 7045	Lys 6836, Gln 6885
ZINC000000308303	Pro 6835, Ile 6838, Ala 6842, Ala 6843, Val 6882, Leu 7042, Phe 7043, Met 7045	Leu 7042
ZINC000019289953	Pro 6835, Ile 6838, Met 6839, Val 6842, Ala 6843, Ala 6881, Val 6882, Leu 7042, Phe 7043, Met 7045	Thr 6846, Gln 6885
ZINC000044872049	Pro 6835, Ile 6838, Met 6839, Val 6842, Ala 6843, Val 6876, Ala 6877, Pro 6878, Ala 6881, Val 6882, Val 6902, Leu 7042, Met 705	Gln 6885, Asp 6904
ZINC000001163520	Val 6842, Ala 683, Pro 6835, Ile 6838, Val 6882, Leu 7042, Phe 7043, Met 7045	Thr 6846, Gln 6885

3.4. Small-molecule Inhibitor Prediction for Nsp16-SAM Interaction

After initial docking of 400.000 ligands, ligands were scored and ranked based on their docking binding energy scores as seen in Figure 3.22. AutoDock Vina docking scores of ligands showed a docking score range from -12,20 kcal/mol to -3,60 kcal/mol. Docking scores were populated around approximately -8,02 kcal/mol. Additionally, the top 50.000 ligands showed a docking score range from -12,20 kcal/mol to -9,90 kcal/mol. These values indicate that the initial screen of randomized 400.000 ligand

library gave a large number of ligands with great docking score and high binding affinity to SAM binding pocket of Nsp16.

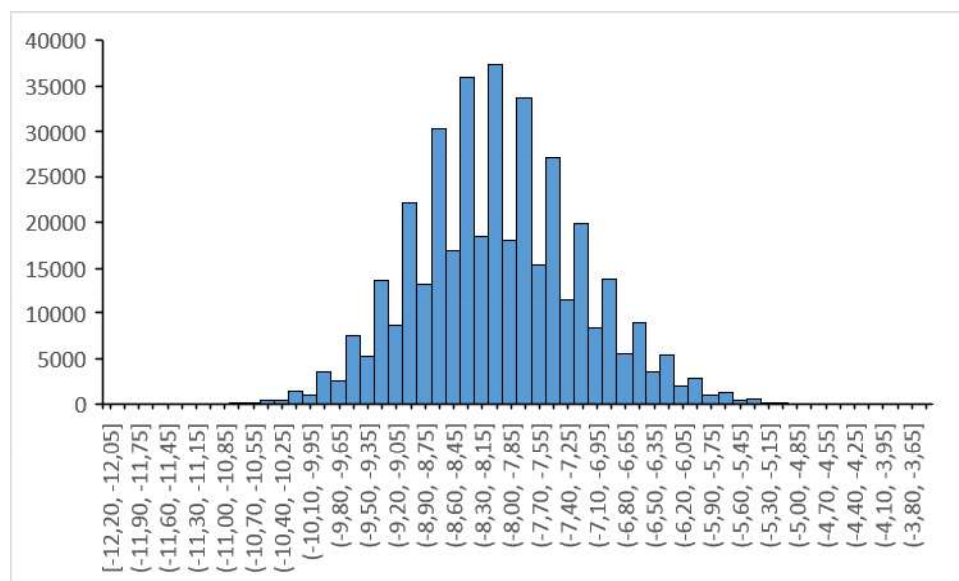


Figure 3.22: Distribution of AutoDock Vina docking scores (x-axis) of 400.000 ligands (y-axis) against SAM binding site of Nsp16.

After scoring and sorting, second run of AutoDock Vina with top 50,000 ligands showed a ligand docking score range from -12.20 kcal/mol to -7.10 kcal/mol. Docking scores were populated around -9.38 kcal/mol which is better than the initial run of ligand samples as seen in Figure 3.23. This indicates that a second iteration of docking improved the docking results.

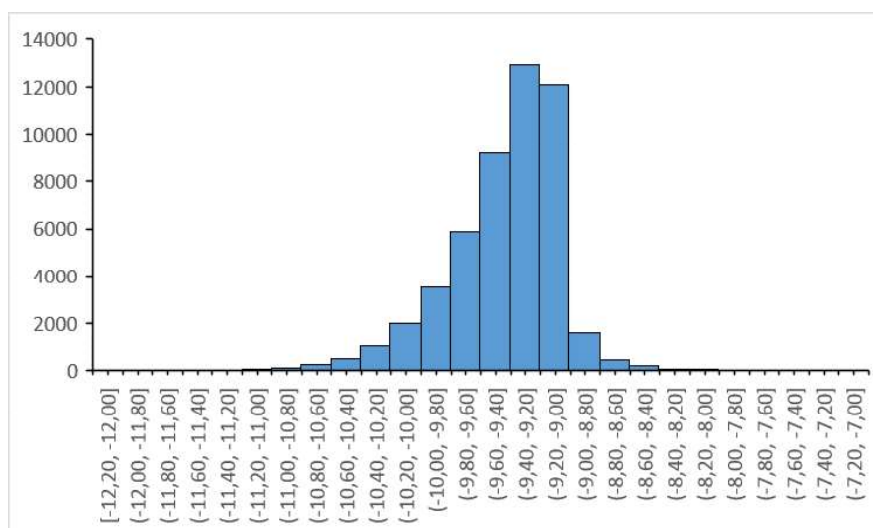


Figure 3.23: Distribution of AutoDock Vina docking scores (x-axis) of 50.000 ligands (y-axis) against SAM binding site of Nsp16.

Moreover, top 5.000 ligands demonstrated docking range from -12,20 kcal/mol to -9,90 kcal/mol as seen in Figure 3.24. The highest scored ligands were also almost the same with the top results of initial docking run, indicating that top-docked ligands can always bind to target site with high affinity and there is high consistency between docking analyses.

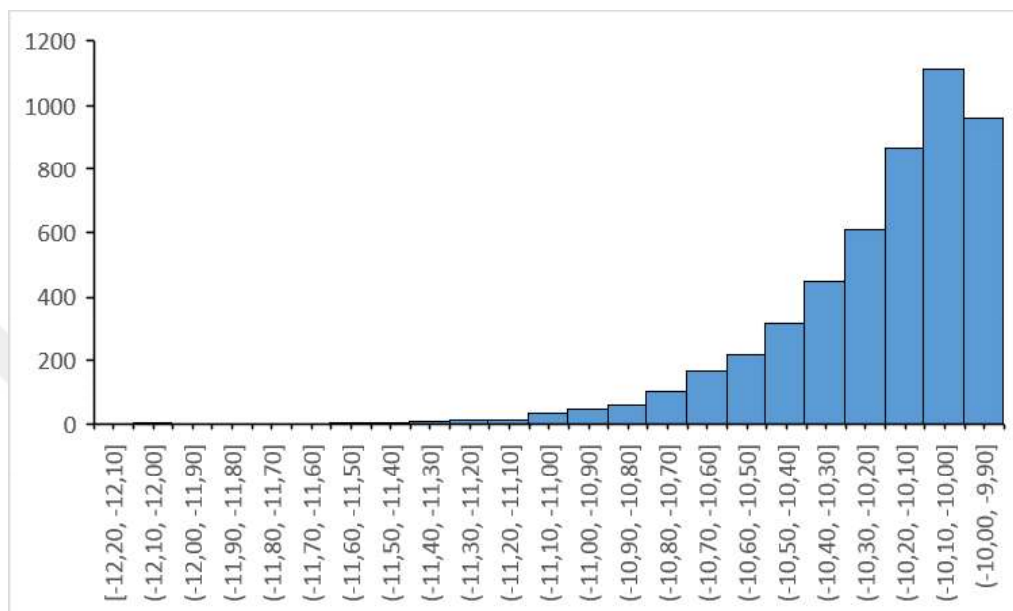


Figure 3.24: Distribution of AutoDock Vina docking scores (x-axis) of 5.000 ligands (y-axis) against SAM binding site of human Nsp16.

Top 5.000 ligands with the highest docking score were prepared by LigPrep module of Maestro. Total 7.897 stereoisomers of top 5.000 ligands were generated. The most favorable and biochemically sensible conformation with low energy were identified for each ligand based on their low state penalty and ionization penalty. Almost all of the most favorable stereoisomers of 5.000 ligands showed ionization and state penalty closer to 0, indicating that ligands used for Glide XP were chemically stable and sensible; and they represented the most probable conformation of a particular ligand being closer to real biological system.

After Glide XP docking of 5.000 ligands, ligands were scored and ranked based on their Glide XP binding energy. Docking score was taken into consideration instead of Glide score as it also includes Epik penalty values. Thus, Epik penalty-added final docking scores accurately reflected true docking score of ligands by excluding energetically unfavourable states. As seen in Figure 3.25, Glide docking scores of small-molecule ligands were in the range from -10,253 kcal/mol to 2,908 kcal/mol.

Additionally, the top 200 ligands showed a docking score range from -10,253 kcal/mol to -6,290 kcal/mol. Since Glide XP used a greatly extensive, different scoring function and Epik penalties were added to exclude energetically unlikely states, docking scores were slightly lower compared to AutoDock Vina docking scores as expected. In Glide XP calculations, many researches considered docking scores less than -5,5 kcal/mol good and the ones less than -8 kcal/mol great. Thus, it could be stated that majority of top ligands docked here had highly great docking scores and could bind the target protein with a high affinity. Moreover, the ligands with highest docking score were also mostly similar to the ones with highest AutoDock Vina docking scores. This finding further verifies that there is high consistency between each docking steps and the screened top 200 ligands could represent potential small-molecule inhibitors with high binding affinity to SAM binding pocket of Nsp16.

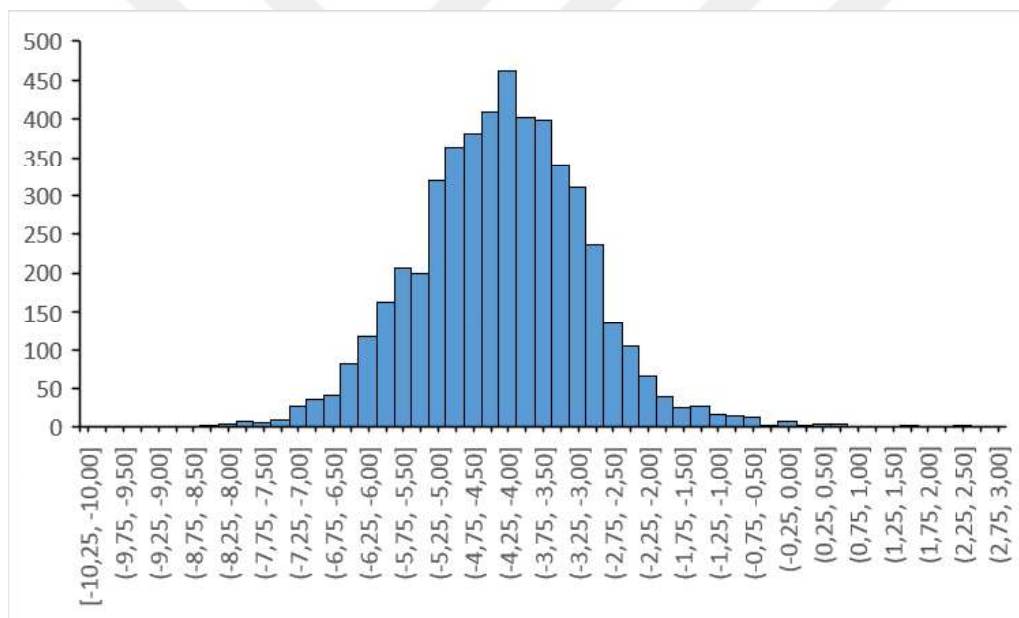


Figure 3.25: Distribution of docking scores (x-axis) of 5.000 ligands (y-axis) against SAM binding site of Nsp16 from Glide XP docking simulation.

After performing the dendritic fingerprinting hierarchical clustering, cut-off “Merging Distance” was set to be 0.65 and 164 different clusters were generated as seen in Figure A.4 in Appendix A. As a result of analysis of each cluster based on the structural and chemical similarities of cluster members, it was concluded that each cluster was highly unique as members of clusters are similar within clusters and different between clusters. This can be seen in Figure 3.26 which demonstrates the in-cluster similarity and between-clusters differences for 3 randomly selected clusters.

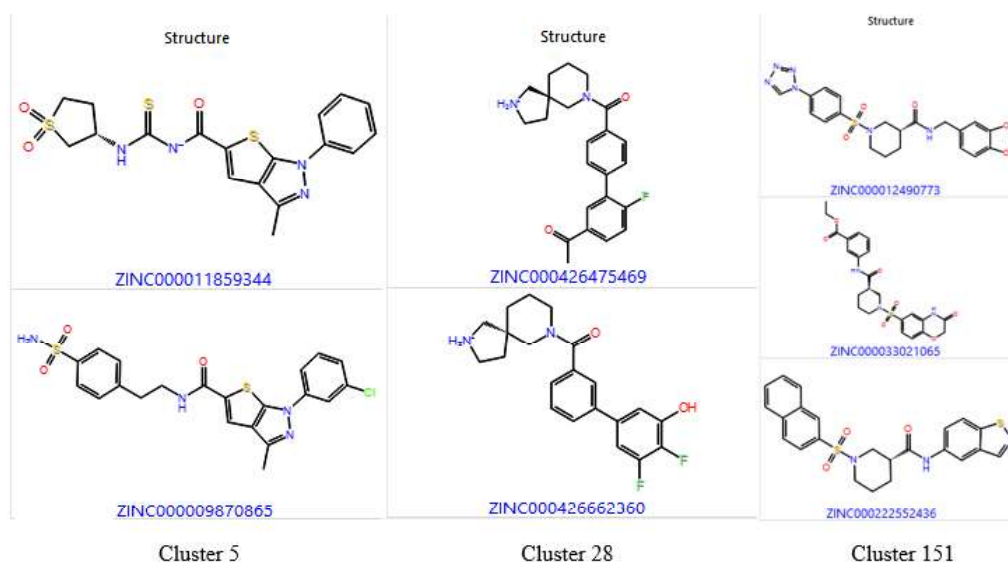


Figure 3.26: Three clusters from dendritic fingerprinting hierarchical clustering of 200 ligands against SAM binding site of Nsp16, identified by CANVAS Software.

For each cluster, the member with the greatest docking score and lowest state penalty was selected as the representative of its respective cluster. These 164 ligands were then screened for ADME/T analysis to assess their pharmaceutical properties and toxicity. 100 molecules with the most optimum ADME/T properties were identified. Among these most drug-like, lower toxicity, low energy and high binding-affinity 100 ligands, top-scored 10 ligands were selected for further steps. Their state penalty, docking scores and ADME/T values can be seen in Table 3.7. All values were in overall great range and this indicates that these ligands can show proper pharmaceutical properties and this procedure might predict drug-like small-molecule ligands against SAM binding pocket on Nsp16.

Table 3.7: Low-energy and high binding-affinity of 10 ligands with the lowest state penalty, the highest docking scores and optimum ADME/T values against SAM binding site of Nsp16.

Important Values	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00	ZINC00
0019373	581	984	0.0000	0.4292	0.0044	0.0024	0.0005	0.0348	0.0000	0.4332	0.0329	0012408	0042384	0014979	433	0012408	0042384	0014979	433
State Penalty	0.0286	0.0000	0.4292	0.0044	0.0024	0.0005	0.0348	0.0000	0.4332	0.0329									
Final AutoDock																			
Vina Score (kcal/mol)	-10.100	-10.000	-9.900	-10.100	-10.100	-10.000	-10.100	-10.200	-10.200	-10.000	-11.000								
Glide XP Docking																			
Score (kcal/mol)	-10.253	-8.716	-8.488	-8.337	-8.176	-8.163	-8.139	-7.838	-7.832	-7.616									
stars	1	0	1	0	0	0	0	0	0	0	0								
rtvFG	0	0	0	1	0	1	0	0	1	0	0								
mol_MW	410.474	376.417	486.610	433.891	489.573	374.454	370.478	452.939	437.454	496.608									
QPlogS	-4.937	-3.796	-3.921	-4.166	-5.029	-3.325	-4.775	-6.346	-3.640	-5.329									
CIQPlogS	-5.191	-3.961	-4.704	-5.348	-5.384	-3.921	-4.925	-5.972	-4.040	-4.986									

Table 3.7 (continued): Low-energy and high binding-affinity of 10 ligands with the lowest state penalty, the highest docking scores and optimum ADME/T values against SAM binding site of Nsp16.

Important Values	ZINC00 0019373 581	ZINC00 0096137 984	ZINC00 0022289 086	ZINC00 0072395 414	ZINC00 0257313 608	ZINC00 0215598 573	ZINC00 0012222 910	ZINC00 0012408 152	ZINC00 0042384 493	ZINC00 0014979 433
QPPCaco	241.509	28.574	188.402	813.402	287.993	595.621	917.807	153.466	170.167	156.319
QplogBB	-0.483	-2.206	0.014	0.355	-1.435	-0.533	0.637	-1.484	-0.625	-0.385
Metab	5	7	6	6	5	6	6	4	2	7
HumanOralAbsorption	3	2	3	3	3	3	3	1	3	3
PercentHumanOralAbsorption	91.857	58.062	92.813	100.000	90.014	88.398	100.000	88.895	78.582	87.866
RuleOfFive	0	0	0	0	0	0	0	0	0	0
RuleOfThree	0	1	0	0	0	0	0	1	0	1

Top-scored ligands were analyzed for their interaction with Nsp16 by Ligand Interaction Diagram module of Maestro Software. Ligands were observed to be properly docked onto binding site (Figure 3.27).

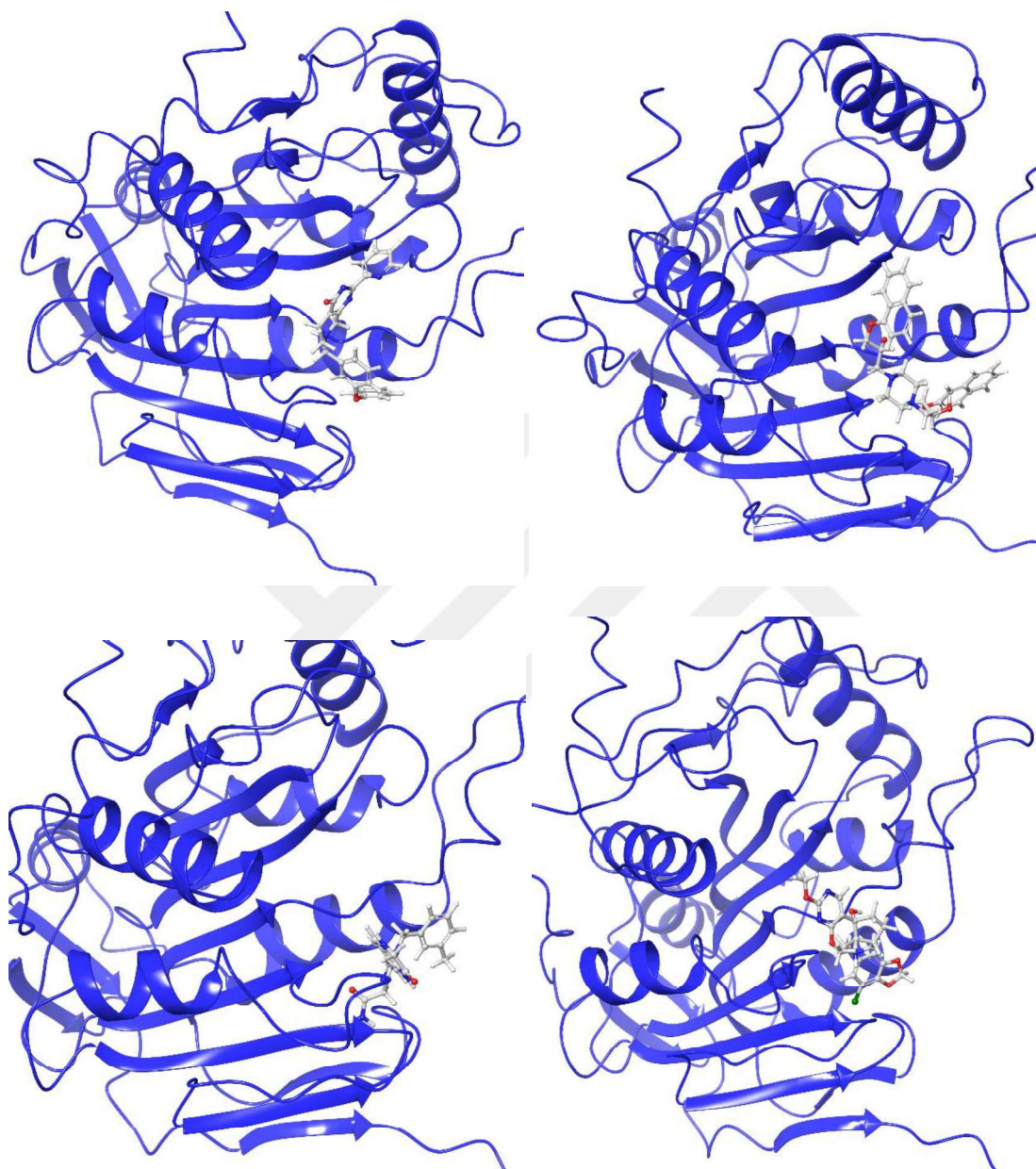


Figure 3.27: Top-scored 10 ligands within SAM binding pocket of Nsp16, visualized on Maestro. From left to right: ZINC000019373581, ZINC000096137984, ZINC000022289086, ZINC000072395414, ZINC000257313608, ZINC000215598573, ZINC000012222910, ZINC000012408152, ZINC000042384493, ZINC000014979433.

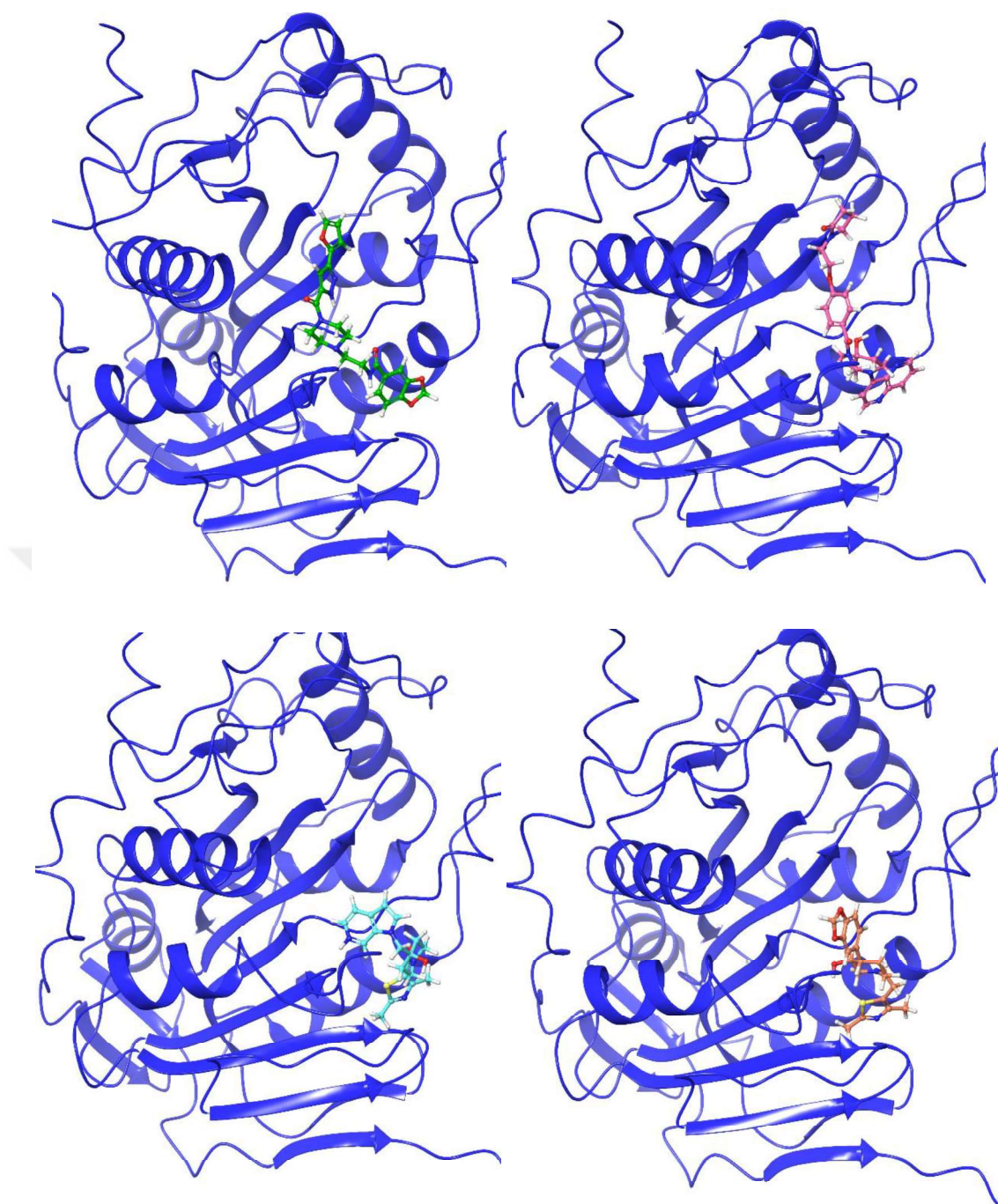


Figure 3.27 (continued): Top-scored 10 ligands within SAM binding pocket of Nsp16, visualized on Maestro. From left to right: ZINC000019373581, ZINC000096137984, ZINC000022289086, ZINC000072395414, ZINC000257313608, ZINC000215598573, ZINC000012222910, ZINC000012408152, ZINC000042384493, ZINC000014979433.

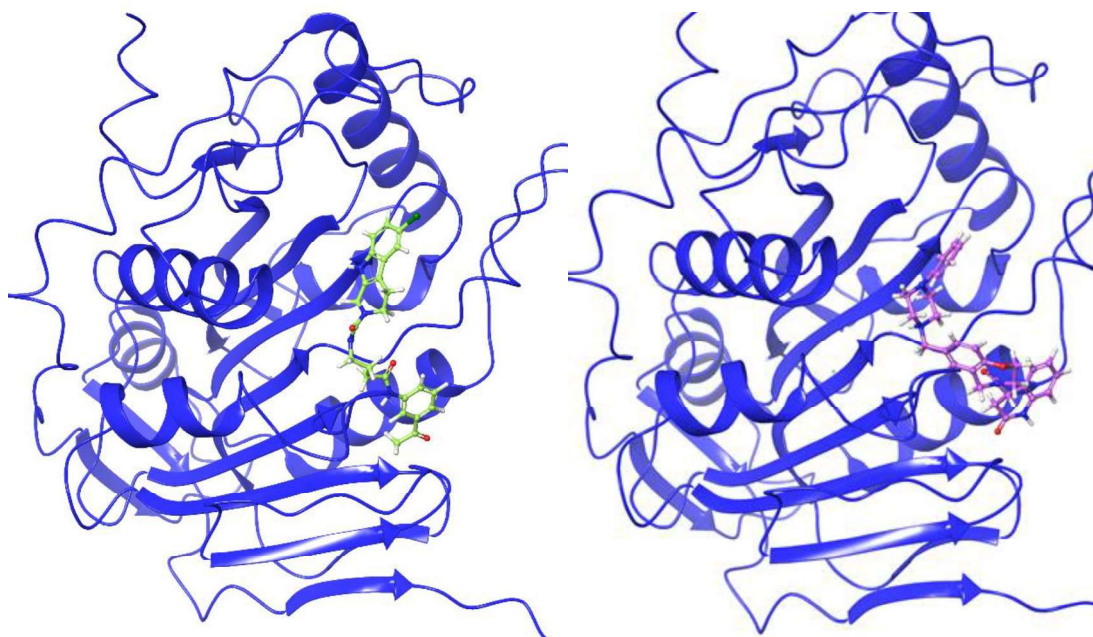


Figure 3.27 (continued): Top-scored 10 ligands within SAM binding pocket of Nsp16, visualized on Maestro. From left to right: ZINC000019373581, ZINC000096137984, ZINC000022289086, ZINC000072395414, ZINC000257313608, ZINC000215598573, ZINC000012222910, ZINC000012408152, ZINC000042384493, ZINC000014979433.

Moreover, a high number of hydrophobic interactions and hydrogen bonding was identified (Figure 3.28, Table 3.8), further verifying that these top-docked ligands can bind to SAM binding site of Nsp16 with high affinity and can block binding of SAM. Major residues making hydrophobic interactions observed in ligands were Tyr 6845, Phe 6868, Ala 6870, Leu 6898, Cys 6913, Ala 6914, Val 6916, Met 6929, Tyr 6930, Pro 6932 and Phe 6947. Major residues making hydrogen bonds observed in ligands were Asn 6841, Gly 6869, Ser 6896, Asp 6897, Leu 6898, Asn 6899, Gly 6911, Asp 6912, Cys 6913, Asp 6928, Tyr 6930, Asp 6931, Lys 6968 and Glu 7001. Most of these residues were also the same with or next to critical residues required for Nsp16-SAM interaction such as Asn 6841, Tyr 6845, Phe 6868, Gly 6869, Ser 6896, Asp 6897, Leu 6898, Asn 6899, Cys 6913, Asp 6928, Met 6929 and Tyr 6930. This highly suggests that the use of these ligands can block Nsp16-SAM interaction as they can antagonistically overlap and bind to the same or close residues on binding interface with high affinity.

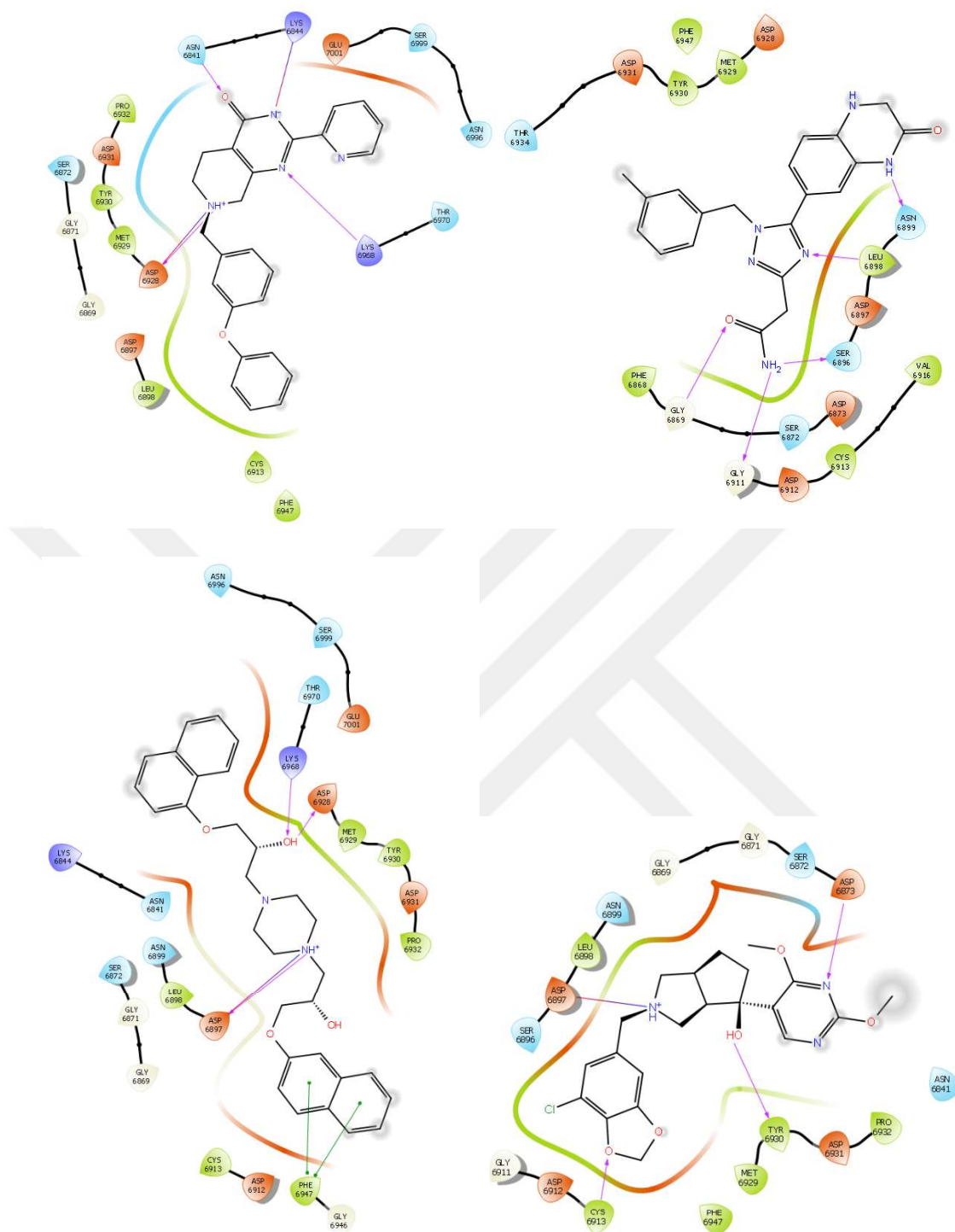


Figure 3.28: Ligand interaction diagrams of top-scored 10 ligands with Nsp16, visualized by Schrödinger Maestro. Red residue: Negatively charged, Purple residue: Positively charged, White residue: Glycine, Yellow residue: Hydrophobic, Blue residue: Polar, Purple arrow: H-bond, Orange arrow: Halogen Bond, Purple line: Metal coordination, Green line: Pi-Pi stacking, Grey circle: Solvent exposure. From left to right: ZINC000019373581, ZINC000096137984, ZINC000022289086, ZINC000072395414, ZINC000257313608, ZINC000215598573, ZINC000012222910, ZINC000012408152, ZINC000042384493, ZINC000014979433.

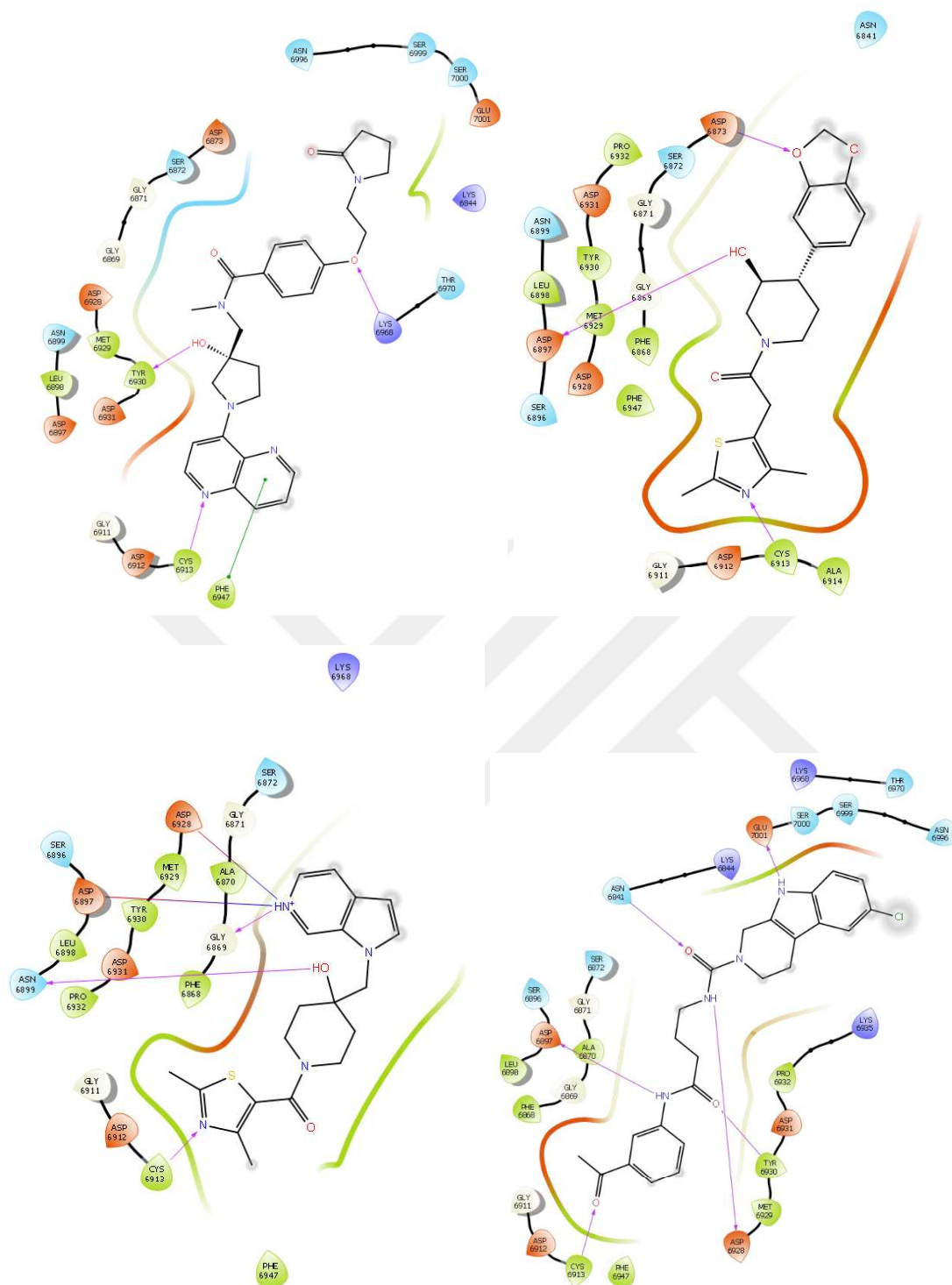


Figure 3.28 (continued): Ligand interaction diagrams of top-scored 10 ligands with Nsp16, visualized by Schrödinger Maestro. Red residue: Negatively charged, Purple residue: Positively charged, White residue: Glycine, Yellow residue: Hydrophobic, Blue residue: Polar, Purple arrow: H-bond, Orange arrow: Halogen Bond, Purple line: Metal coordination, Green line: Pi-Pi stacking, Grey circle: Solvent exposure. From left to right: ZINC000019373581, ZINC000096137984, ZINC000022289086, ZINC000072395414, ZINC000257313608, ZINC000215598573, ZINC000012222910, ZINC000012408152, ZINC000042384493, ZINC000014979433.

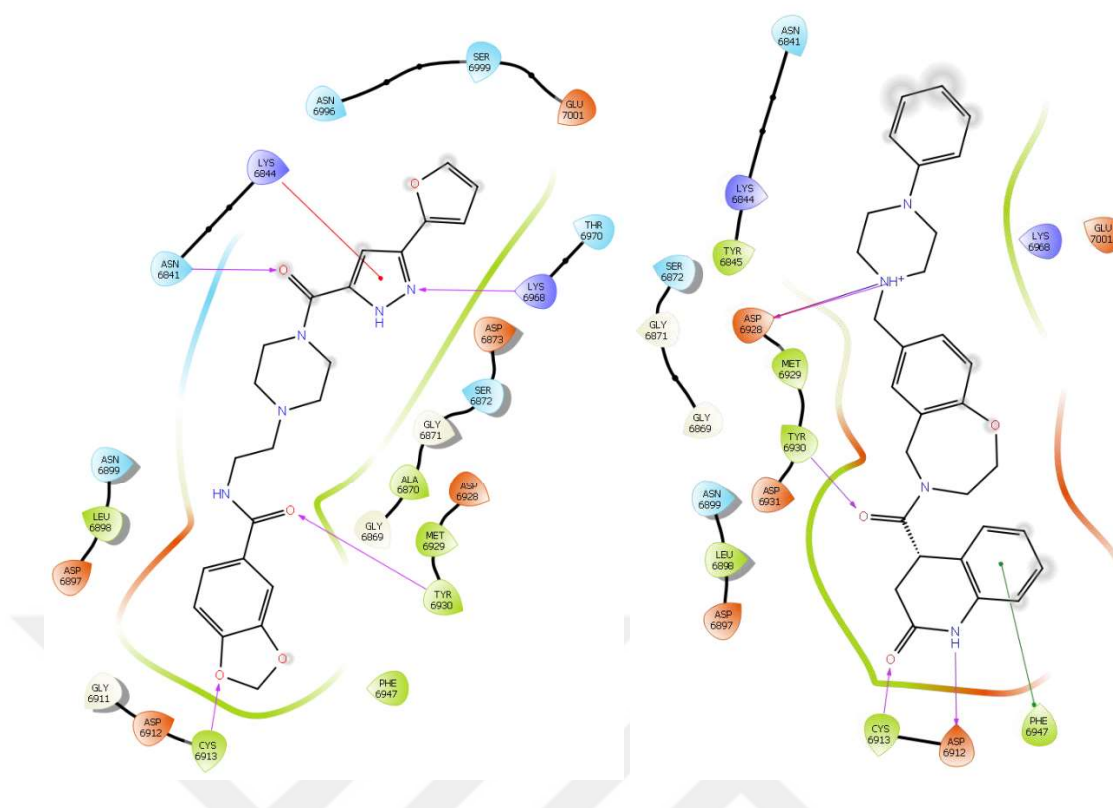


Figure 3.28 (continued): Ligand interaction diagrams of top-scored 10 ligands with Nsp16, visualized by Schrödinger Meastro. Red residue: Negatively charged, Purple residue: Positively charged, White residue: Glycine, Yellow residue: Hydrophobic, Blue residue: Polar, Purple arrow: H-bond, Orange arrow: Halogen Bond, Purple line: Metal coordination, Green line: Pi-Pi stacking, Grey circle: Solvent exposure. From left to right: ZINC000019373581, ZINC000096137984, ZINC000022289086, ZINC000072395414, ZINC000257313608, ZINC000215598573, ZINC000012222910, ZINC000012408152, ZINC000042384493, ZINC000014979433.

Table 3.8: Protein-ligand interaction profile of Nsp16 for its SAM binding site and top-scored 10 small-molecule ligands.

Small-molecule Ligand	Hydrophobic Interactions	Hydrogen Bonds
ZINC000019373581	Leu 6898, Cys 6913, Met 6929, Tyr 6930, Pro 6932, Phe 6947	Asn 6841, Asp 6928, Lys 6968
ZINC000096137984	Phe 6868, Leu 6898, Cys 6913, Val 6916, Met 6929, Tyr 6930, Phe 6947	Gly 6869, Ser 6896, Leu 6898, Asn 6899, Gly 6911
ZINC000022289086	Leu 6898, Cys 6913, Met 6929, Tyr 6930, Pro 6932, Phe 6947	Asp 6897, Asp 6928, Lys 6968

Table 3.8: Protein-ligand interaction profile of Nsp16 for its SAM binding site and top-scored 10 small-molecule ligands.

Small-molecule Ligand	Hydrophobic Interactions	Hydrogen Bonds
ZINC000072395414	Leu 6898, Cys 6913, Met 6929, Tyr 6930, Pro 6932, Phe 6947	Asp 6873, Asp 6897, Cys 6913, Tyr 6930
ZINC000257313608	Leu 6898, Cys 6913, Met 6929, Tyr 6930, Phe 6947	Cys 6913, Tyr 6930, Lys 6968
ZINC000215598573	Leu 6898, Phe 6868, Cys 6913, Ala 6914, Met 6929, Tyr 6930, Pro 6932, Phe 6947	Asp 6897, Cys 6913
ZINC000012222910	Phe 6868, Ala 6870, Leu 6898, Cys 6913, Met 6929, Tyr 6930, Pro 6932, Phe 6947	Asp 6897, Cys 6913, Asp 6928, Asp 6931
ZINC000012408152	Ala 6870, Phe 6868, Leu 6898, Cys 6913, Met 6929, Tyr 6930, Pro 6932, Phe 6947	Asn 6841, Asp 6897, Cys 6913, Asp 6928, Tyr 6930, Glu 7001
ZINC000042384493	Ala 6870, Leu 6898, Cys 6913, Met 6929, Tyr 6930, Phe 6947	Asn 6841, Cys 6913, Tyr 6930, Lys 6968
ZINC000014979433	Tyr 6845, Leu 6898, Cys 6913, Met 6929, Tyr 6930, Phe 6947	Asp 6928, Cys 6913, Asp 6912, Tyr 6930

Lastly, when compared to docking score of redocked SAM (-8,1 kcal/mol), top-scored 7 ligands (ZINC000019373581, ZINC000096137984, ZINC000022289086, ZINC000072395414, ZINC000257313608, ZINC000215598573 and ZINC000012222910) docked onto SAM binding pocket showed docking scores (-10,253 kcal/mol, -8,716 kcal/mol, -8,488 kcal/mol, -8,337 kcal/mol, -8,176 kcal/mol, -8,163 kcal/mol and -8,139 kcal/mol) greater than SAM. This indicates that these ligands can potentially prevent SAM binding and thus the methyltransferase activity of Nsp16 as they can bind to SAM binding site with higher binding affinity.

3.5. Molecular Dynamics Simulations

Atom displacement trajectories of top-scored ligands against Spike RBD, ACE2 and Nsp16 were analyzed. Left-y axis of each plot represents the RMSD deviation of protein based on its center alpha carbons and side chains with respect to backbone reference frame ($t=0$). Fluctuations between protein RMSD values at each frame indicates whether the protein can reach equilibrium through simulation. The state of equilibrium represents the rigidity and stability of protein within ligand-bound complex. Right-y axis of each plot represents the RMSD deviation of a particular ligand with respect to reference frame ligand ($t=0$). Fluctuations between ligand RMSD values at each frame indicates whether the ligand can reach equilibrium through simulation. The state of equilibrium indicates whether the ligand is stable within the complex and can properly remain within the binding pocket of target protein. RMSD deviations between 1.0-3.0 Å generally means that atom displacements are low and thus, overall stabilities of protein and ligands were preserved in bound complex. This also indicates that docked ligand can retain its position within binding pocket for a long-period of time to be considered as inhibitory molecule against target protein. Additionally, the last frame of each ligand was compared to first frame to demonstrate the conformational changes of the system.

According to trajectory of ligand ZINC000067920242 against Spike RBD (Fig 3.29), Spike RBD was found to be adequately stable over the course of simulation as it reached a slight equilibrium despite showing a momentary large fluctuation that can be attributed to slight conformational loop changes. It also showed RMSD fluctuation range less than ~3.0 Å which can be considered acceptable. The ligand showed great stability over 100 ns as it constantly remained in equilibrium and showed RMSD fluctuation range less than ~1.2 Å which is highly good. This result suggested that ligand can preserve its bound position within a stable complex to be used as potential inhibitory molecule against ACE2 binding. Compared frames showed low displacement, indicating ligand constantly remained within binding site (Figure B.1).

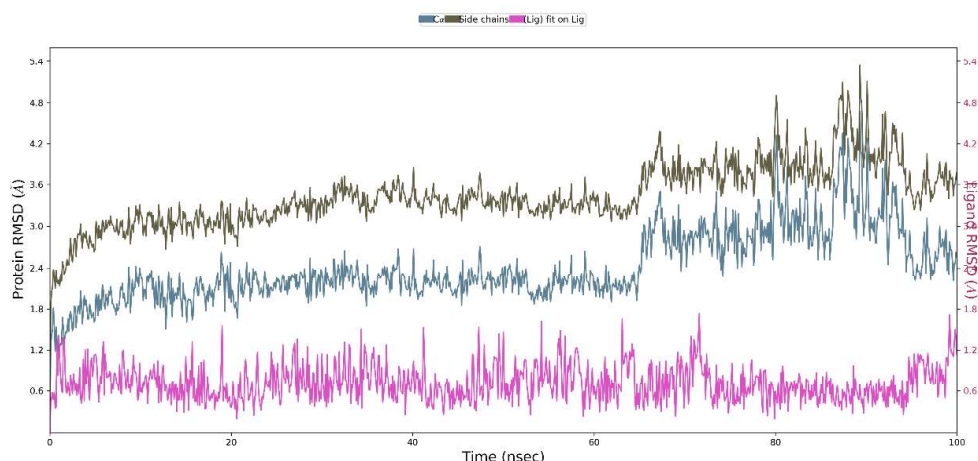


Fig 3.29: Changes in RMSD of Spike RBD C α and side chains (left y-axis) and ligand ZINC000067920242 (right y-axis) with respect to reference frame ($t=0$) over 100 ns atom displacement trajectory. The graph was generated by Maestro “Simulation Interaction Diagram” module.

According to trajectory of ligand ZINC000009828805 against Spike RBD (Fig 3.30), Spike RBD was found to be highly stable over the course of simulation as it mostly remained in equilibrium and showed RMSD fluctuation range less than ~ 2.5 Å which can be considered acceptable. The ligand showed great stability over 100 ns as it constantly remained in equilibrium despite slight atom deviations and showed RMSD fluctuation range less than ~ 1.5 Å. This result suggested that ligand can preserve its bound position within a stable complex to be used as potential inhibitory molecule against ACE2 binding. Compared frames also showed moderate displacement, indicating that ligand remained mostly within binding site (Figure B.2).

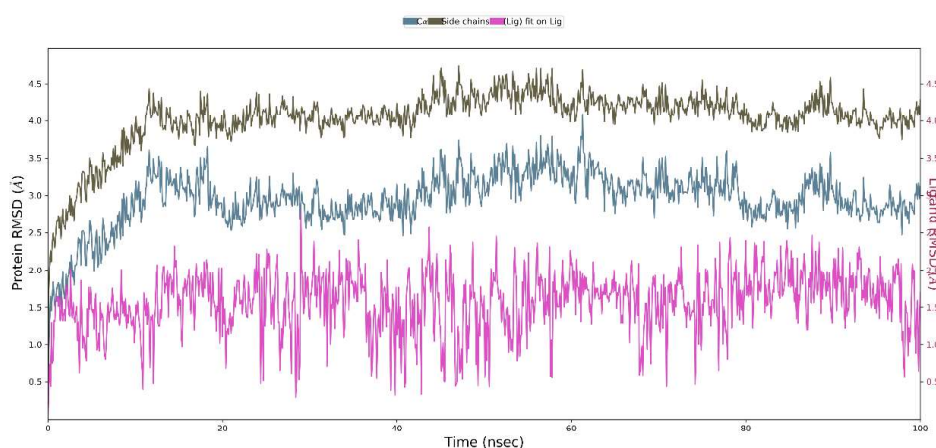


Fig 3.30: Changes in RMSD of Spike RBD C α and side chains (left y-axis) and ligand ZINC000009828805 (right y-axis) with respect to reference frame ($t=0$) over 100 ns atom displacement trajectory. The graph was generated by Maestro “Simulation Interaction Diagram” module.

According to trajectory of ligand ZINC000014228954 against Spike RBD (Fig 3.31), Spike RBD was found to be highly stable over the course of simulation as it constantly remained in equilibrium and showed RMSD fluctuation range less than ~ 2.0 Å which can be considered acceptable. The ligand showed great stability over 100 ns as it remained mostly in equilibrium state and showed RMSD fluctuation range less than ~ 2.0 Å. This result suggested that ligand can preserve its bound position within a stable complex to be used as potential inhibitory molecule against ACE2 binding. Compared frames also showed moderate displacement, indicating that ligand remained mostly within binding site (Figure B.3).

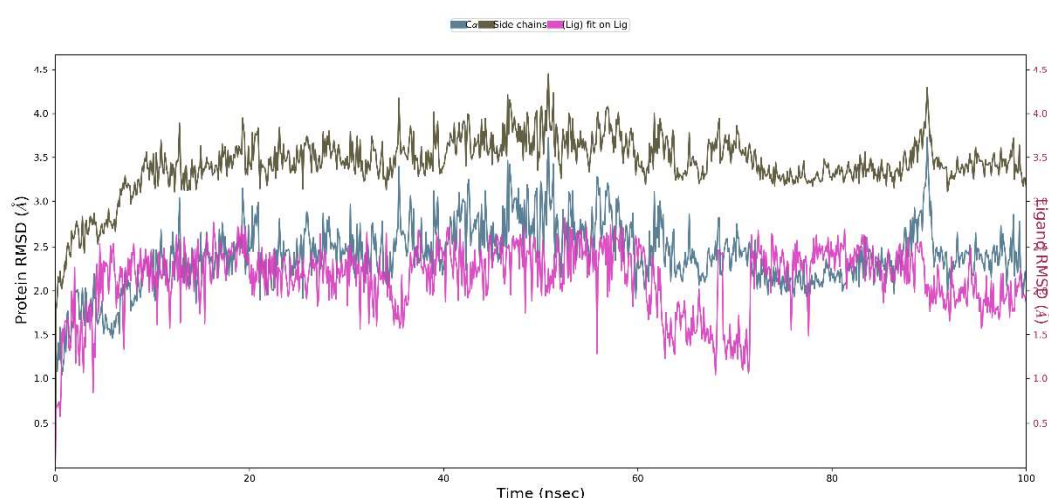


Fig 3.31: Changes in RMSD of Spike RBD C α and side chains (left y-axis) and ligand ZINC000014228954 (right y-axis) with respect to reference frame (t=0) over 100 ns atom displacement trajectory. The graph was generated by Maestro “Simulation Interaction Diagram” module.

According to trajectory of ligand ZINC000009134748 against Spike RBD (Fig 3.32), Spike RBD was found to be mostly stable over the course of simulation as it reached equilibrium and showed RMSD fluctuation range less than ~ 3.0 Å which can be considered acceptable. The ligand showed good stability over 100 ns as it remained mostly in equilibrium despite two single large momentary fluctuation and showed RMSD fluctuation range less than ~ 1.8 Å. This result suggested that ligand can preserve its bound position within a stable complex to be used as potential inhibitory molecule against ACE2 binding. Compared frames also showed low displacement, indicating that ligand constantly remained within binding site (Figure B.4).

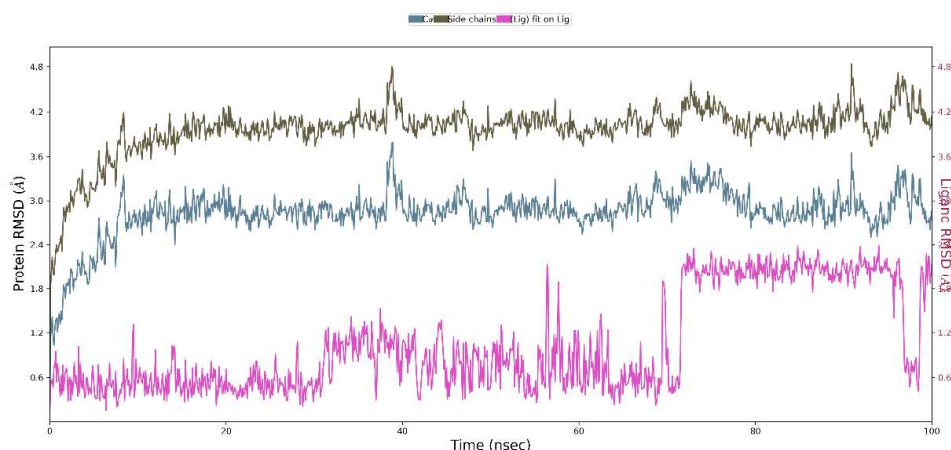


Fig 3.32: Changes in RMSD of Spike RBD C α and side chains (left y-axis) and ligand ZINC000009134748 (right y-axis) with respect to reference frame (t=0) over 100 ns atom displacement trajectory. The graph was generated by Maestro “Simulation Interaction Diagram” module.

According to trajectory of ligand ZINC000008772453 against Spike RBD (Fig 3.33), Spike RBD was found to be highly stable over the course of simulation as it reached a constant equilibrium and showed RMSD fluctuation range less than ~ 3.0 Å which can be considered acceptable. The ligand showed great stability over 100 ns as it constantly remained in equilibrium state and showed RMSD fluctuation range less than ~ 0.8 Å. This result suggested that ligand can preserve its bound position within a stable complex to be used as potential inhibitory molecule against ACE2 binding. Compared frames also showed moderate displacement, indicating that ligand remained mostly within binding site (Figure B.5).

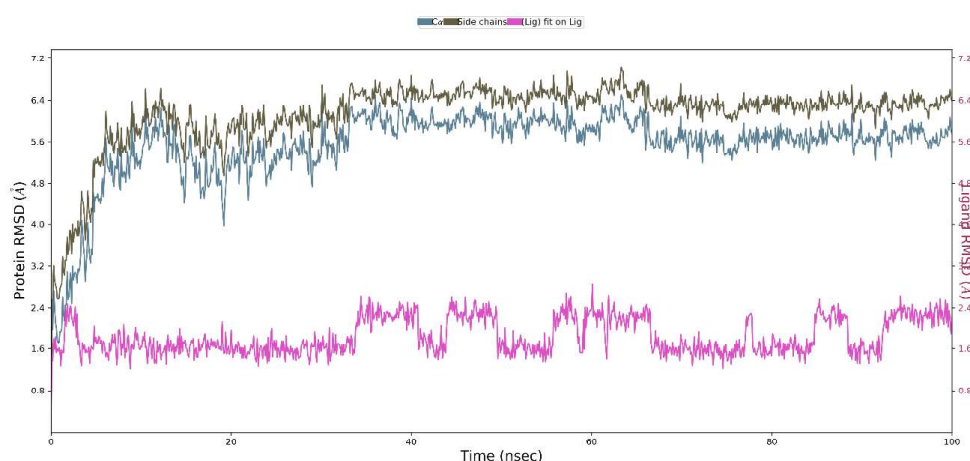


Fig 3.33: Changes in RMSD of Spike RBD C α and side chains (left y-axis) and ligand ZINC000008772453 (right y-axis) with respect to reference frame (t=0) over 100 ns atom displacement trajectory. The graph was generated by Maestro “Simulation Interaction Diagram” module.

According to trajectory of ligand ZINC000014879822 against ACE2 (Fig 3.34), ACE2 was found to be highly stable over the course of simulation as it reached a constant equilibrium and showed RMSD fluctuation range less than ~ 2.0 Å which can be considered acceptable. The ligand showed good stability over 100 ns as it reached a constant equilibrium despite its initial RMSD fluctuations and showed RMSD fluctuation range less than ~ 2.0 Å. This result suggested that ligand can preserve its bound position within a stable complex to be used as potential inhibitory molecule against RBD binding. Compared frames also showed low displacement, indicating that ligand constantly remained within binding site (Figure B.6).

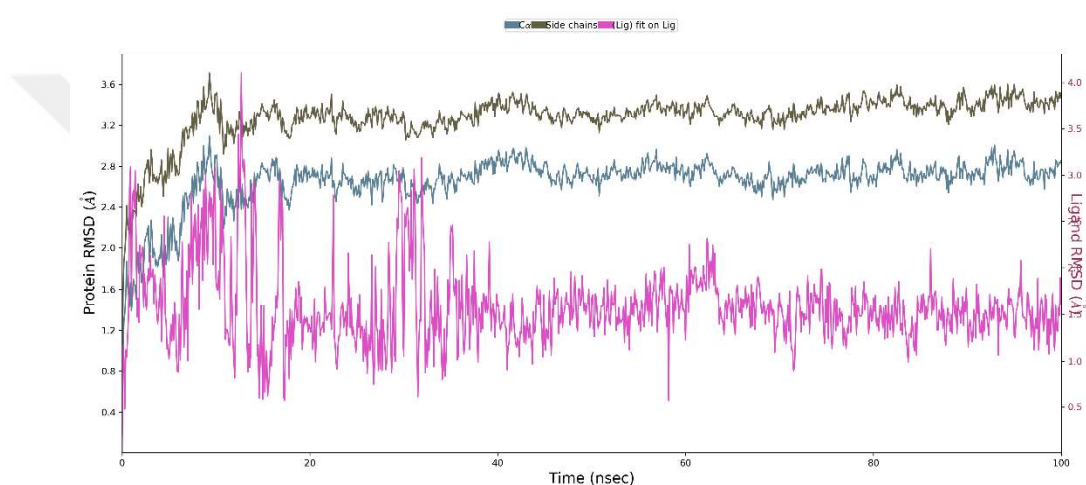


Fig 3.34: Changes in RMSD of ACE2 Cα and side chains (left y-axis) and ligand ZINC000014879822 (right y-axis) with respect to reference frame (t=0) over 100 ns atom displacement trajectory. The graph was generated by Maestro “Simulation Interaction Diagram” module.

According to trajectory of ligand ZINC000000813934 against ACE2 (Fig 3.35), ACE2 was found to be mostly stable over the course of simulation as it reached a slight equilibrium and showed RMSD fluctuation range less than ~ 2.8 Å which can be considered acceptable. The ligand showed great stability over 100 ns as it remained constantly in equilibrium state and showed RMSD fluctuation range less than ~ 1.2 Å. This result suggested that ligand can preserve its bound position within a stable complex to be used as potential inhibitory molecule against RBD binding. Compared frames also showed low displacement, indicating that ligand constantly remained within binding site. Since its placement within the deeper part of binding area, it can be considered that this ligand might inhibit RBD binding allosterically (Figure B.7).

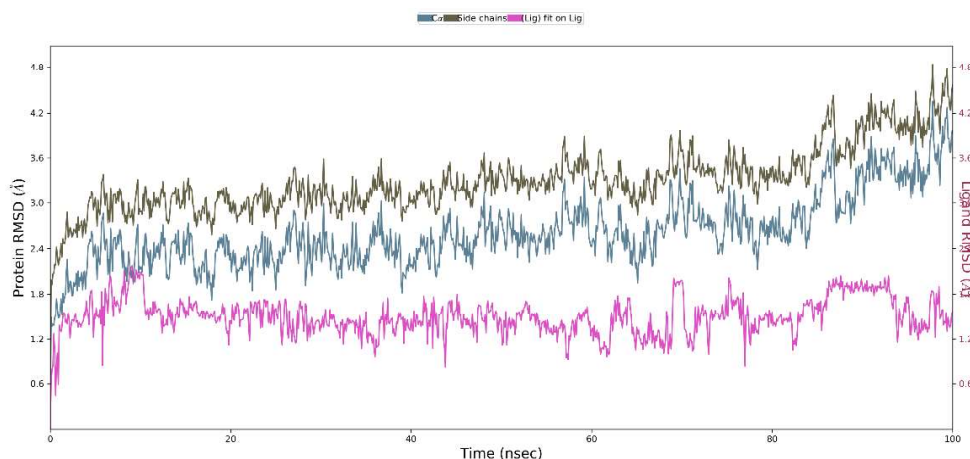


Fig 3.35: Changes in RMSD of ACE2 C α and side chains (left y-axis) and ligand ZINC00000813934 (right y-axis) with respect to reference frame (t=0) over 100 ns atom displacement trajectory. The graph was generated by Maestro “Simulation Interaction Diagram” module.

According to trajectory of ligand ZINC000019742660 against ACE2 (Fig 3.36), ACE2 was found to be mostly stable over the course of simulation as it reached a slight equilibrium and showed RMSD fluctuation range less than ~ 2.5 Å which can be considered acceptable. The ligand showed good stability over 100 ns as it could reach a slight equilibrium state despite two momentary fluctuations and showed RMSD fluctuation range less than ~ 3.0 Å. This result suggested that ligand can preserve its bound position within a stable complex to be used as potential inhibitory molecule against RBD binding. Compared frames also showed low displacement, indicating that ligand constantly remained within binding site (Figure B.8).

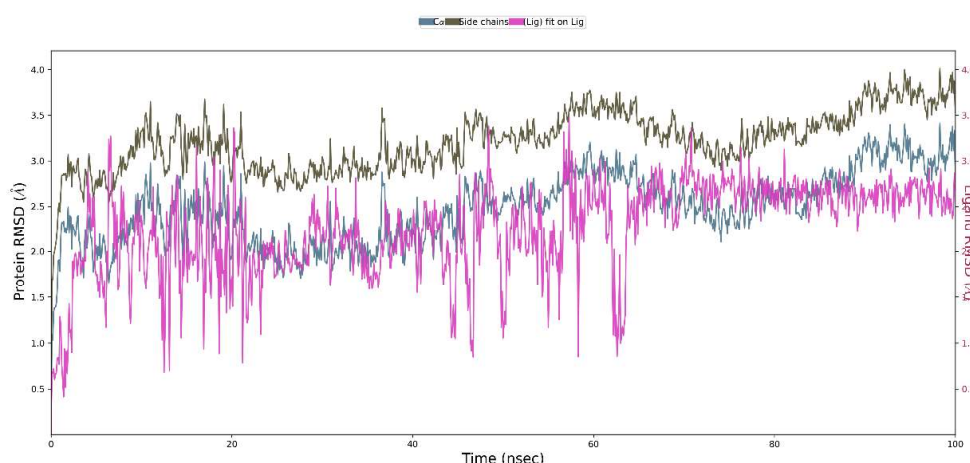


Fig 3.36: Changes in RMSD of ACE2 C α and side chains (left y-axis) and ligand ZINC000019742660 (right y-axis) with respect to reference frame (t=0) over 100 ns atom displacement trajectory. The graph was generated by Maestro “Simulation Interaction Diagram” module.

According to trajectory of ligand ZINC000019022189 against ACE2 (Fig 3.37), ACE2 was found to be mostly stable over the course of simulation as it reached a constant equilibrium and showed RMSD fluctuation range less than ~ 2.0 Å which can be considered acceptable. The ligand showed good stability over 100 ns as it remained mostly in equilibrium state despite a single large RMSD fluctuation and showed RMSD fluctuation range less than ~ 2.0 Å. This result suggested that ligand can preserve its bound position within a stable complex to be used as potential inhibitory molecule against RBD binding. Compared frames also showed moderate displacement, indicating that ligand remained mostly within binding site (Figure B.9).

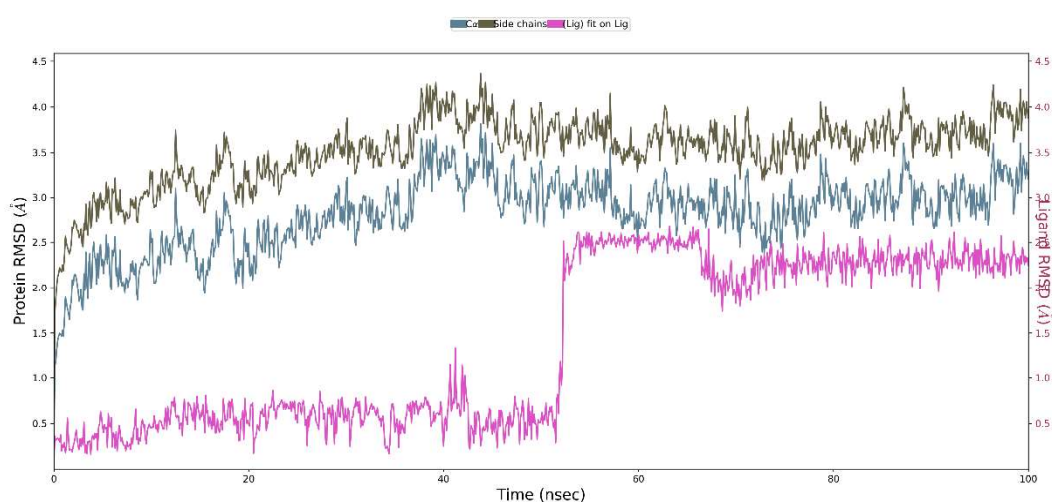


Fig 3.37: Changes in RMSD of ACE2 Cα and side chains (left y-axis) and ligand ZINC000019022189 (right y-axis) with respect to reference frame (t=0) over 100 ns atom displacement trajectory. The graph was generated by Maestro “Simulation Interaction Diagram” module.

According to trajectory of ligand ZINC000020149756 against ACE2 (Fig 3.38), ACE2 was found to be mostly stable over the course of simulation as it reached a slight equilibrium and showed RMSD fluctuation range less than ~ 2.0 Å which can be considered acceptable. The ligand showed great stability over 100 ns as it remained mostly in equilibrium state despite a single large RMSD fluctuation and showed RMSD fluctuation range less than ~ 1.8 Å. This result suggested that ligand can preserve its bound position within a stable complex to be used as potential inhibitory molecule against RBD binding. Compared frames also showed low displacement, indicating that ligand constantly remained within binding site (Figure B.10).

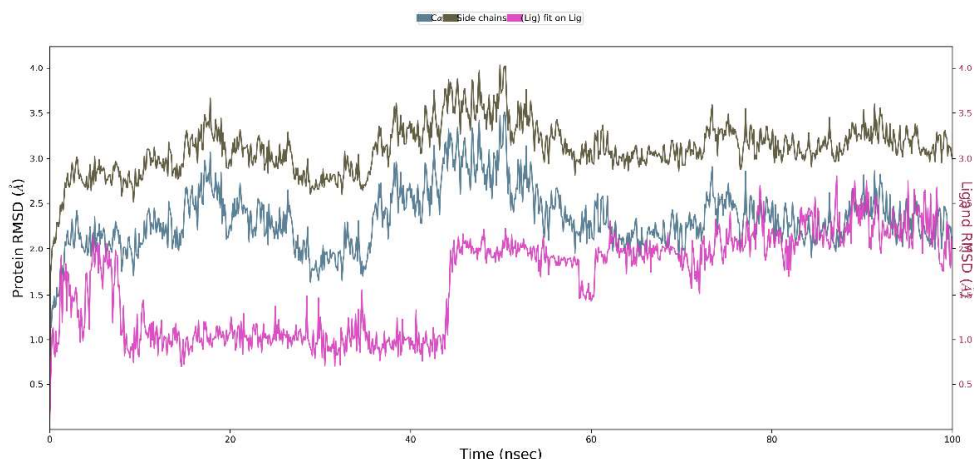


Fig 3.38: Changes in RMSD of ACE2 C α and side chains (left y-axis) and ligand ZINC000020149756 (right y-axis) with respect to reference frame (t=0) over 100 ns atom displacement trajectory. The graph was generated by Maestro “Simulation Interaction Diagram” module.

According to trajectory of ligand ZINC000009751488 against Nsp 10 binding site of Nsp16 (Fig 3.39), Nsp16 was found to be mostly stable over the course of simulation as it reached a slight equilibrium and showed RMSD fluctuation range less than ~ 2.0 Å which can be considered acceptable. The ligand showed great stability over 100 ns as it constantly remained in equilibrium state and showed RMSD fluctuation range less than ~ 2.0 Å. This result suggested that ligand can preserve its bound position within a stable complex to be used as potential inhibitory molecule against Nsp10 binding. Compared frames also showed low displacement, indicating that ligand constantly remained within binding site (Figure B.11).

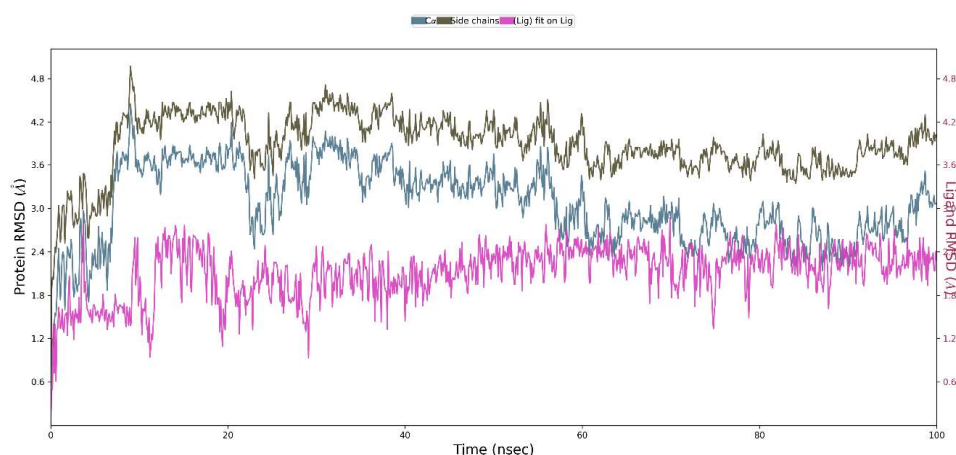


Fig 3.39: Changes in RMSD of Nsp16 C α and side chains (left y-axis) and ligand ZINC000009751488 (right y-axis) with respect to reference frame (t=0) over 100 ns atom displacement trajectory. The graph was generated by Maestro “Simulation Interaction Diagram” module.

According to trajectory of ligand ZINC000002286111 against Nsp 10 binding site of Nsp16 (Fig 3.40), Nsp16 was found to be highly stable over the course of simulation as it reached a constant equilibrium and showed RMSD fluctuation range less than ~ 2.5 Å which can be considered acceptable. The ligand showed great stability over 100 ns as it remained mostly in equilibrium despite a single large momentary fluctuation and showed RMSD fluctuation range less than ~ 2.5 Å. This result suggested that ligand can preserve its bound position within a stable complex to be used as potential inhibitory molecule against Nsp10 binding. Compared frames also showed low displacement, indicating that ligand constantly remained within binding site (Figure B.12).

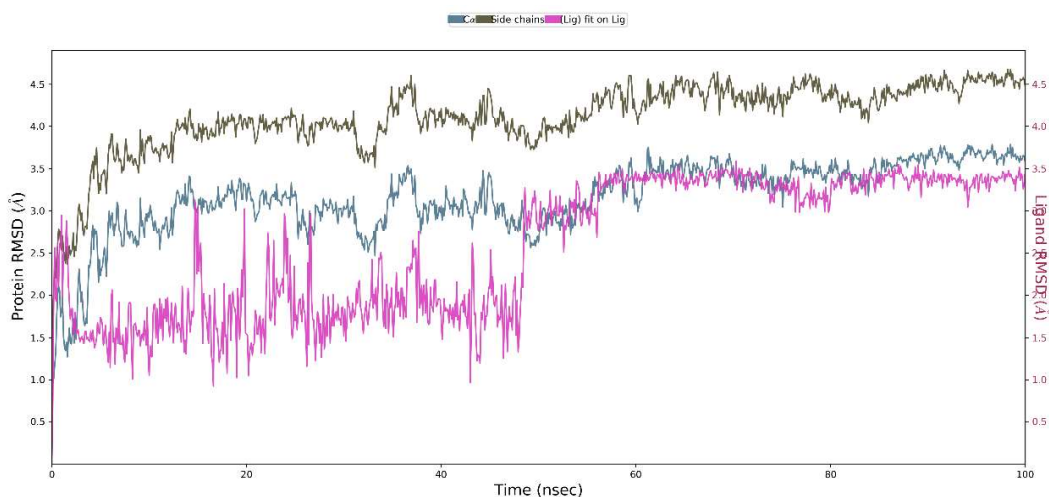


Fig 3.40: Changes in RMSD of Nsp16 C α and side chains (left y-axis) and ligand ZINC000002286111 (right y-axis) with respect to reference frame (t=0) over 100 ns atom displacement trajectory. The graph was generated by Maestro “Simulation Interaction Diagram” module.

According to trajectory of ligand ZINC000019497136 against Nsp 10 binding site of Nsp16 (Fig 3.41), Nsp16 was found to be mostly stable over the course of simulation as it reached a slight equilibrium and showed RMSD fluctuation range less than ~ 2.5 Å which can be considered acceptable. The ligand showed mediocre stability over 100 ns as it remained mostly in equilibrium state as it showed varying fluctuations to some degree and showed RMSD fluctuation range less than ~ 3.0 Å. This result suggested that ligand can preserve its bound position within a stable complex to be used as potential inhibitory molecule against Nsp10 binding. Compared frames also showed low displacement, indicating that ligand constantly remained within binding site (Figure B.13).

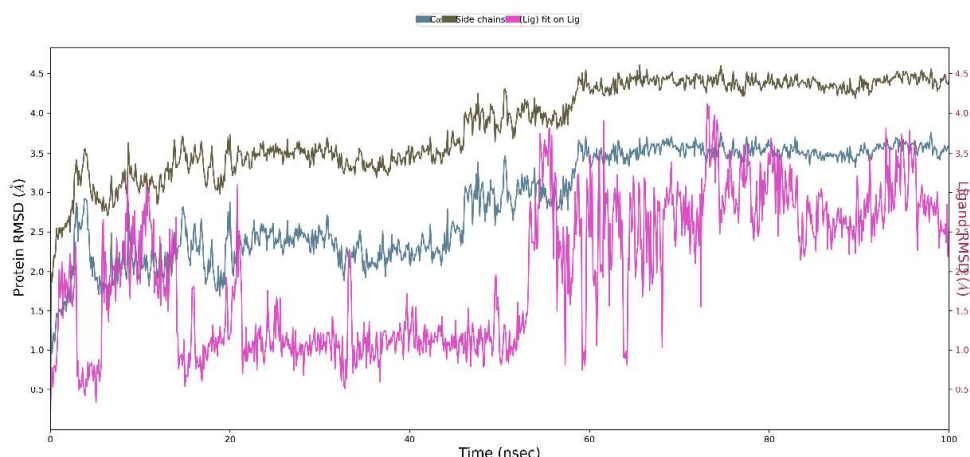


Fig 3.41: Changes in RMSD of Nsp16 Ca and side chains (left y-axis) and ligand ZINC000019497136 (right y-axis) with respect to reference frame ($t=0$) over 100 ns atom displacement trajectory. The graph was generated by Maestro “Simulation Interaction Diagram” module.

According to trajectory of ligand ZINC000009252898 against Nsp 10 binding site of Nsp16 (Fig 3.42), Nsp16 was found to be highly stable over the course of simulation as it reached a constant equilibrium and showed RMSD fluctuation range less than ~ 2.5 Å which can be considered acceptable. The ligand showed great stability over 100 ns as it constantly remained in equilibrium and showed RMSD fluctuation range less than ~ 2.0 Å. This result suggested that ligand can preserve its bound position within a stable complex to be used as potential inhibitory molecule against Nsp10 binding. Compared frames also showed low displacement, indicating that ligand constantly remained within binding site (Figure B.14).

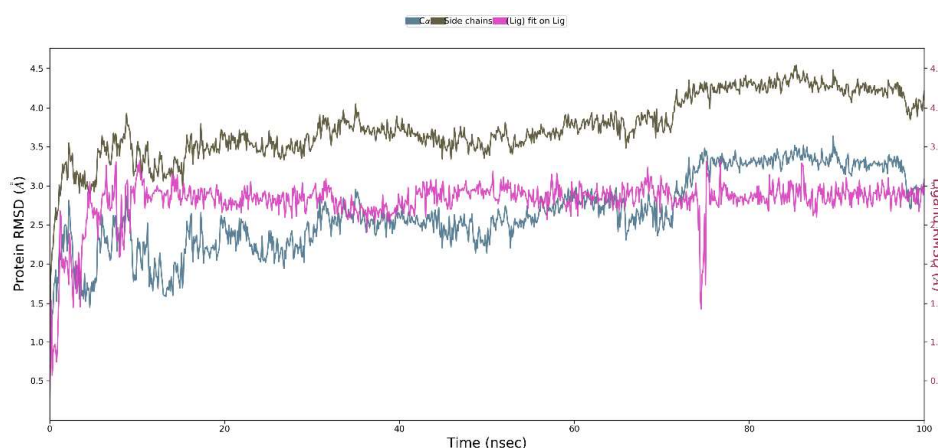


Fig 3.42: Changes in RMSD of Nsp16 Ca and side chains (left y-axis) and ligand ZINC000009252898 (right y-axis) with respect to reference frame ($t=0$) over 100 ns atom displacement trajectory. The graph was generated by Maestro “Simulation Interaction Diagram” module.

According to trajectory of ligand ZINC000009485446 against Nsp 10 binding site of Nsp16 (Fig 3.43), Nsp16 was found to be mostly stable over the course of simulation as it reached a slight equilibrium and showed RMSD fluctuation range less than ~ 2.5 Å which can be considered acceptable. The ligand showed a good stability over 100 ns as it reached a constant equilibrium after 30 ns showed RMSD fluctuation range less than ~ 2.5 Å. This result suggested that ligand can preserve its bound position within a stable complex to be used as potential inhibitory molecule against Nsp10 binding. Compared frames also showed low displacement, indicating that ligand constantly remained within binding site (Figure B.15).

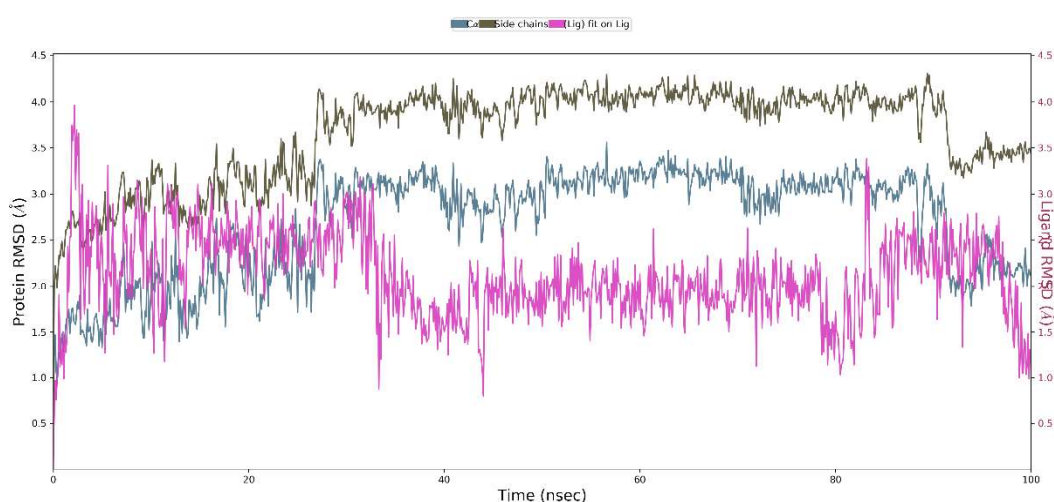


Fig 3.43: Changes in RMSD of Nsp16 C α and side chains (left y-axis) and ligand ZINC000009485446 (right y-axis) with respect to reference frame (t=0) over 100 ns atom displacement trajectory. The graph was generated by Maestro “Simulation Interaction Diagram” module.

According to trajectory of ligand ZINC000019373581 against SAM binding site of Nsp16 (Fig 3.44), Nsp16 was found to be highly stable over the course of simulation as it reached a constant equilibrium and showed RMSD fluctuation range less than ~ 2.5 Å which can be considered acceptable. The ligand showed great stability over 100 ns as it remained mostly in equilibrium state and showed RMSD fluctuation range less than ~ 1.0 Å. This result suggested that ligand can preserve its bound position within a stable complex to be used as potential inhibitory molecule against SAM binding. Compared frames also showed low displacement, indicating that ligand constantly remained within binding site (Figure B.16).

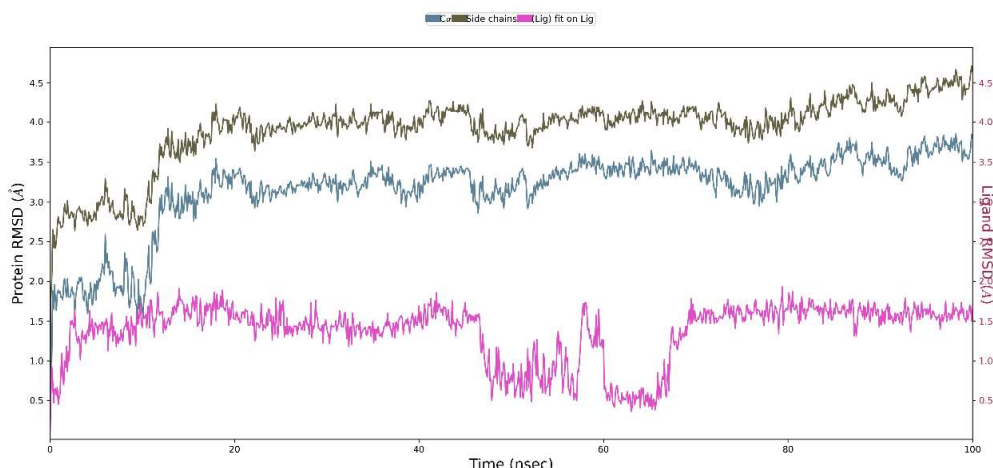


Fig 3.44: Changes in RMSD of Nsp16 C α and side chains (left y-axis) and ligand ZINC000019373581 (right y-axis) with respect to reference frame (t=0) over 100 ns atom displacement trajectory. The graph was generated by Maestro “Simulation Interaction Diagram” module.

According to trajectory of ligand ZINC000096137984 against SAM binding site of Nsp16 (Fig 3.45), Nsp16 was found to be mostly stable over the course of simulation as it reached a slight equilibrium and showed RMSD fluctuation range less than ~ 3.0 Å which can be considered adequate. The ligand showed great stability over 100 ns as it remained mostly in equilibrium state and showed RMSD fluctuation range less than ~ 2.0 Å. This result suggested that ligand can preserve its bound position within a stable complex to be used as potential inhibitory molecule against SAM binding. Compared frames also showed low displacement, indicating that ligand constantly remained within binding site (Figure B.17).

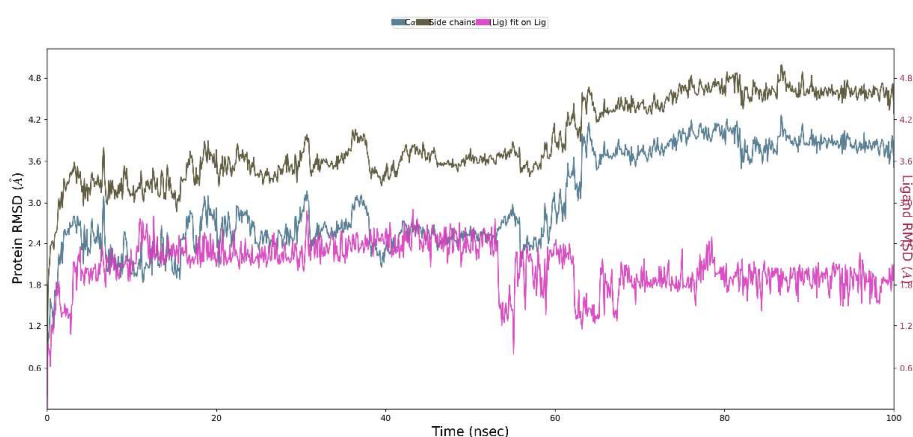


Fig 3.45: Changes in RMSD of Nsp16 C α and side chains (left y-axis) and ligand ZINC000096137984 (right y-axis) with respect to reference frame (t=0) over 100 ns atom displacement trajectory. The graph was generated by Maestro “Simulation Interaction Diagram” module.

According to trajectory of ligand ZINC000257313608 against SAM binding site of Nsp16 (Fig 3.46), Nsp16 was found to be mostly stable over the course of simulation as it reaches a slight equilibrium and showed RMSD fluctuation range less than ~ 3.0 Å which can be considered acceptable. The ligand showed a mediocre stability over 100 ns as it showed varying degree of fluctuations and showed RMSD fluctuation range less than ~ 3.0 Å. This result suggested that ligand can slightly preserve its bound position within a stable complex to be used as potential inhibitory molecule against SAM binding. Compared frames also showed moderate displacement, indicating that ligand remained mostly within binding site (Figure B.18).

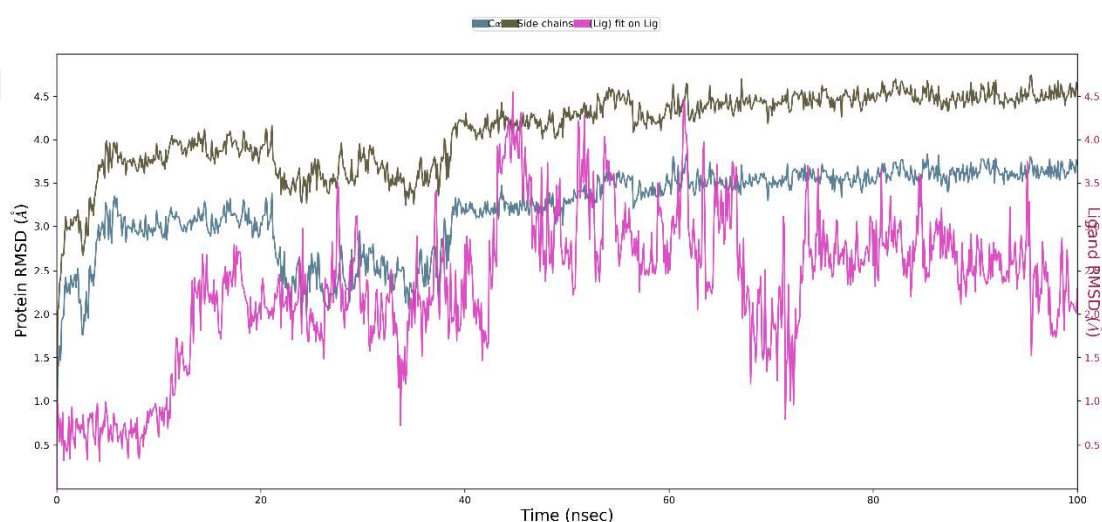


Fig 3.46: Changes in RMSD of Nsp16 C α and side chains (left y-axis) and ligand ZINC000257313608 (right y-axis) with respect to reference frame (t=0) over 100 ns atom displacement trajectory. The graph was generated by Maestro “Simulation Interaction Diagram” module.

According to trajectory of ligand ZINC000022289086 against SAM binding site of Nsp16 (Fig 3.47), Nsp16 was found to be mostly stable over the course of simulation as it reached a slight equilibrium and showed RMSD fluctuation range less than ~ 2.0 Å which can be considered acceptable. The ligand showed great stability over 100 ns as it remained mostly in equilibrium state and showed RMSD fluctuation range less than ~ 2.0 Å. This result suggested that ligand can preserve its bound position within a stable complex to be used as potential inhibitory molecule against SAM binding. Compared frames also showed low displacement, indicating that ligand constantly remained within binding site (Figure B.19).

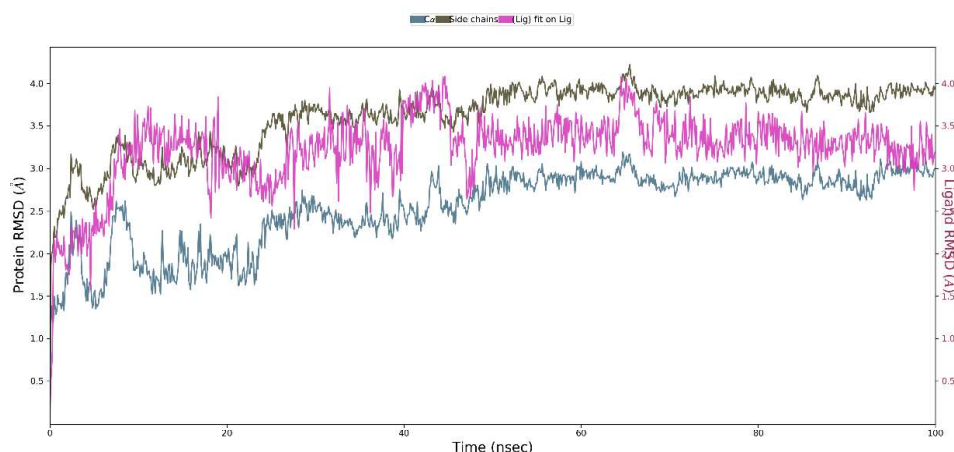


Fig 3.47: Changes in RMSD of Nsp16 C α and side chains (left y-axis) and ligand ZINC000022289086 (right y-axis) with respect to reference frame (t=0) over 100 ns atom displacement trajectory. The graph was generated by Maestro “Simulation Interaction Diagram” module.

According to trajectory of ligand ZINC000072395414 against SAM binding site of Nsp16 (Fig 3.48), Nsp16 was found to be mostly stable over the course of simulation as it reached a slight equilibrium and showed RMSD fluctuation range less than ~ 2.5 Å which can be considered acceptable. The ligand showed great stability over 100 ns as it remained mostly in equilibrium state and showed RMSD fluctuation range less than ~ 1.5 Å. This result suggested that ligand can preserve its bound position within a stable complex to be used as potential inhibitory molecule against SAM binding. Compared frames also showed low displacement, indicating that ligand constantly remained within binding site (Figure B.20).

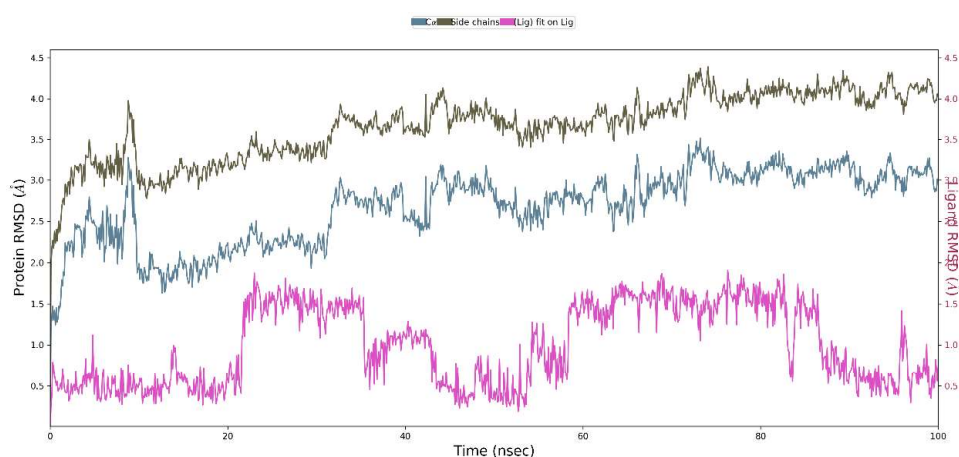


Fig 3.48: Changes in RMSD of Nsp16 C α and side chains (left y-axis) and ligand ZINC000072395414 (right y-axis) with respect to reference frame (t=0) over 100 ns atom displacement trajectory. The graph was generated by Maestro “Simulation Interaction Diagram” module.

4. DISCUSSION

AutoDock Vina docking calculations of ligands against Spike RBD, human ACE2 receptor, Nsp10 binding site of Nsp16 and SAM binding site of Nsp16 resulted in very high binding scores, indicating that selected ligands can bind to target protein with very high affinity. Additionally, second runs of all AutoDock Vina calculations gave higher binding scores and score averages; and this indicated that this second iteration of docking improved the docking results for tested ligands against their respective protein targets. The highest scored-ligands between two docking runs were also similar. This shows that there is high accuracy and high consistency between docking analyses.

Schrödinger Suite Glide XP docking calculations resulted in a different range of slightly lower docking scores for all target protein dockings. This is due to the fact that Glide XP utilized its own and more extensive scoring function so lower scores were expected compared to AutoDock Vina scoring calculations. However, these results could accurately reflect true docking score of ligands as these calculations excluded energetically unfavourable states by analyzing Epik penalty-added final docking scores. In the case of Spike RBD and ACE2, this can be also attributed to the fact that binding surface is highly accessible/open and flat instead of a more favorable deep pocket conformation (which is more proper for ligand binding) for the Spike RBD and ACE2 binding interface. On the other hand, ligands docked against Nsp10 and SAM binding sites of Nsp16 gave greater docking scores as docked site is a deep binding pocket. Despite these, Glide XP scores showed an average range from -6.00 kcal/mol to -9.00 kcal/mol docking score which can be considered as high binding affinity values for top-scored ligands since many researches considered docking scores less than -5,50 kcal/mol good and the ones less than -8.00 kcal/mol great in Glide XP calculations. Thus, it could be stated that majority of top ligands docked here had good docking scores and could bind their respective target protein with a high affinity. Resulting top ligands with highest binding affinity can be considered as potential small inhibitory molecules. Additionally, 6 of top-scored ligands against SAM binding site

of Nsp16 gave greater docking score than SAM so it can be stated that these ligands can highly compete with SAM in an antagonistic manner. For Spike RBD, it should be noted that results are attributed to 'up' conformation (active state) of Spike RBD where it can actively bind to ACE2 and initiate host cell entry; and thus, resulting ligands should be considered specific to infection-mediating state of protein. On the other hand, high-scored ligands for Nsp16 should be considered specific to the prevention of viral replication cycle.

Moreover, clustering fingerprinting results demonstrated that top ligand candidates are different from each other in terms of chemical and structural properties as members of clusters are similar within clusters and different between clusters. This diversity enables the increase the number of molecule options for possible future wet-lab experiments to test. ADME results showed that top ligands are statistically closer to optimum in terms of drug-likeness and pharmaceutical properties as all ADME/T values were in overall great range. Furthermore, a high number of hydrophobic interactions and hydrogen bonding was identified between top-scored docked ligands and their respective targets, Spike RBD, ACE2, Nsp10 and SAM binding sites of Nsp16. Most of the residues observed in this study for selected ligands were also the same with or next to readily known critical residues of target proteins (Spike RBD, ACE2 and Nsp16) required for their interaction with ACE2, RBD, Nsp10 and SAM respectively. This highly suggests that the use of these ligands can block target interactions as they can antagonistically overlap and bind to the same or close residues on target binding interfaces with high affinity.

According to analyses of displacement trajectories of top-scored ligands against Spike RBD, ACE2 and Nsp16, the state of RMSD fluctuation equilibrium represents the rigidity and stability of protein within ligand-bound complex. For ligands, such equilibrium indicates that the ligand is stable within the complex and can properly remain within the binding pocket of target protein. For the analysis, RMSD deviations between 1.0-3.0 Å generally means that atom displacements are low and thus, overall stabilities of protein and ligands were preserved in bound complex. In the majority of MD simulations for ligand-bound complexes, target proteins were generally found to be mostly stable over the course of simulation as they reached a slight or constant equilibrium. Few of these showed some exceptions by showing momentary large fluctuations that can be attributed to slight conformational loop changes. RMSD

fluctuations for proteins were also mostly less than $\sim 2.5\text{-}3.0$ Å which can be considered acceptable. Specifically, fluctuations seen in Spike RBD might be due to the fact that only RBD part of Spike were taken from the entire Spike structure and this might hinder some potential of simulations. Similar to this case, fluctuations seen in Nsp16 protein simulations could be due to its relaxing C-terminal tail as it generally folded near the end of simulations. The RMSD fluctuations due to these structural elements can be ignored as they were far from the target binding sites and did not affect the conformation of binding site which was docked by target ligands. Generally, conformations of binding sites remained constant in all MD simulations, indication that significant parts required for ligand interaction and stability were proper. Furthermore, majority of ligands also showed good stability over 100 ns as they constantly remained in equilibrium despite slight atom deviations and overall, showed RMSD fluctuation range less than $\sim 2.0\text{-}2.5$ Å. Ligands with constant RMSD values or slight RMSD fluctuations can be considered as ligands that can remain bound directly to the critical residues of target binding sites. On the other hand, ligands that showed moderate fluctuations were generally slightly displaced from their binding site but remained within binding pocket area. These ligands can be still considered for potential inhibitors but they are required more extensive simulation runs.

As a promising future perspective research, the same procedure can be performed on SARS-CoV-2 variants. In this approach, the same set of initial ligand library of identified ligands at the end of this thesis can be docked against mutant Spike proteins on their ACE2 binding site. The results coming from such experiment can give an insight about the efficiency of predicted ligands on Covid-19 variants to see if they can be utilized as small-molecule inhibitor candidates during the evolution of virus.

In conclusion, following ligands were identified to show the features of proper small-molecule inhibitors such as high-binding affinity, high stability, high drug-likeness, low toxicity. Ligands ZINC000067920242, ZINC000009828805, ZINC000009134748 and ZINC000008772453 were found to be potential small-molecule ligands against Spike RBD. Ligands ZINC000014879822, ZINC000000813934, ZINC000019742660 and ZINC000020149756 were found to be potential small-molecule ligands against ACE2. Ligands ZINC000009751488, ZINC000002286111, ZINC000009252898 and ZINC000009485446 were found to be potential small-molecule ligands against Nsp10 binding site of Nsp16. Ligands

ZINC000019373581, ZINC000096137984, ZINC000022289086 and ZINC000072395414 were found to be potential small-molecule ligands against SAM binding site of Nsp16. Identified ligands are also novel compounds which were not used previously and have unknown activities. These proposed ligands can be tested in future laboratory research such as in vitro protein binding assays and cell neutralization assays to verify their inhibitory use against SARS-CoV-2.



REFERENCES

- Al-Qahtani, A. A.** (2020). Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2): emergence, history, basic and clinical aspects. *Saudi journal of biological sciences*, 27(10), 2531-2538.
- Andrusier, N., Mashlach, E., Nussinov, R., & Wolfson, H. J.** (2008). Principles of flexible protein–protein docking. *Proteins: Structure, Function, and Bioinformatics*, 73(2), 271-289.
- Banerjee, A., Kulcsar, K., Misra, V., Frieman, M., & Mossman, K.** (2019). Bats and coronaviruses. *Viruses*, 11(1), 41.
- Berg, K. A., & Clarke, W. P.** (2018). Making sense of pharmacology: inverse agonism and functional selectivity. *International Journal of Neuropsychopharmacology*, 21(10), 962-977.
- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., & Bourne, P.E.** (2000) The Protein Data Bank *Nucleic Acids Research*, 28: 235-242.
- Bowers, K. J., Chow, D. E., Xu, H., Dror, R. O., Eastwood, M. P., Gregersen, B. A., ... & Shaw, D. E.** (2006, November). Scalable algorithms for molecular dynamics simulations on commodity clusters. In *SC'06: Proceedings of the 2006 ACM/IEEE Conference on Supercomputing* (pp. 43-43). IEEE.
- Carugo, O., & Pongor, S.** (2001). A normalized root-mean-square distance for comparing protein three-dimensional structures. *Protein science*, 10(7), 1470-1473.
- Cevik, M., Kuppalli, K., Kindrachuk, J., & Peiris, M.** (2020). Virology, transmission, and pathogenesis of SARS-CoV-2. *bmj*, 371.
- Coleman, C. M., & Frieman, M. B.** (2014). Coronaviruses: important emerging human pathogens. *Journal of virology*, 88(10), 5209-5212.
- Dhir, N., Jain, A., Mahendru, D., Prakash, A., & Medhi, B.** (2020). Drug repurposing and orphan disease therapeutics. In *Drug Repurposing-Hypothesis, Molecular Aspects and Therapeutic Applications*. IntechOpen.
- Duan, J., Dixon, S. L., Lowrie, J. F., & Sherman, W.** (2010). Analysis and comparison of 2D fingerprints: insights into database screening performance using eight fingerprint methods. *Journal of Molecular Graphics and Modelling*, 29(2), 157-170.
- Fehr, A. R., & Perlman, S.** (2015). Coronaviruses: an overview of their replication and pathogenesis. *Coronaviruses*, 1-23.

- Ferreira, L. G., Dos Santos, R. N., Oliva, G., & Andricopulo, A. D.** (2015). Molecular docking and structure-based drug design strategies. *Molecules*, 20(7), 13384-13421.
- Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., ... & Shenkin, P. S.** (2004). Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *Journal of medicinal chemistry*, 47(7), 1739-1749.
- Guterres, H., & Im, W.** (2020). Improving protein-ligand docking results with high-throughput molecular dynamics simulations. *Journal of Chemical Information and Modeling*, 60(4), 2189-2198.
- Hall Jr, D. C., & Ji, H. F.** (2020). A search for medications to treat COVID-19 via in silico molecular docking models of the SARS-CoV-2 spike glycoprotein and 3CL protease. *Travel medicine and infectious disease*, 35, 101646.
- Henderson, R., Edwards, R. J., Mansouri, K., Janowska, K., Stalls, V., Gobeil, S. M., ... & Acharya, P.** (2020). Controlling the SARS-CoV-2 spike glycoprotein conformation. *Nature structural & molecular biology*, 27(10), 925-933.
- Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., ... & Pöhlmann, S.** (2020). SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *cell*, 181(2), 271-280.
- Hollingsworth, S. A., & Dror, R. O.** (2018). Molecular dynamics simulation for all. *Neuron*, 99(6), 1129-1143.
- Humphrey, W., Dalke, A. and Schulten, K.,** (1996) "VMD - Visual Molecular Dynamics", *J. Molec. Graphics*, vol. 14, pp. 33-38.
- Irwin, J. J., & Shoichet, B. K.** (2005). ZINC— a free database of commercially available compounds for virtual screening. *Journal of chemical information and modeling*, 45(1), 177-182.
- Jo, S., Kim, T., Iyer, V. G., & Im, W.** (2008). CHARMM-GUI: a web-based graphical user interface for CHARMM. *Journal of computational chemistry*, 29(11), 1859-1865.
- Kastritis, P. L., & Bonvin, A. M.** (2013). On the binding affinity of macromolecular interactions: daring to ask why proteins interact. *Journal of The Royal Society Interface*, 10(79), 20120835.
- 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), 1-7.
- Kuo, L., Hurst-Hess, K. R., Koetzner, C. A., & Masters, P. S.** (2016). Analyses of coronavirus assembly interactions with interspecies membrane and nucleocapsid protein chimeras. *Journal of virology*, 90(9), 4357-4368.
- Lalchhandama, K.** (2020). The chronicles of coronaviruses: The electron microscope, the doughnut, and the spike. *Sci. Vis*, 20, 78-92.

- Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., ... & Wang, X.** (2020). Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*, 581(7807), 215-220.
- Lengauer, T., & Rarey, M.** (1996). Computational methods for biomolecular docking. *Current opinion in structural biology*, 6(3), 402-406.
- Li, J., Fu, A., & Zhang, L.** (2019). An overview of scoring functions used for protein–ligand interactions in molecular docking. *Interdisciplinary Sciences: Computational Life Sciences*, 11(2), 320-328.
- Lim, Y. X., Ng, Y. L., Tam, J. P., & Liu, D. X.** (2016). Human coronaviruses: a review of virus–host interactions. *Diseases*, 4(3), 26.
- Lin, S., Chen, H., Ye, F., Chen, Z., Yang, F., Zheng, Y., ... & Lu, G.** (2020). Crystal structure of SARS-CoV-2 nsp10/nsp16 2'-O-methylase and its implication on antiviral drug design. *Signal transduction and targeted therapy*, 5(1), 1-4.
- Mahdian, S., Ebrahim-Habibi, A., & Zarrabi, M.** (2020). Drug repurposing using computational methods to identify therapeutic options for COVID-19. *Journal of Diabetes & Metabolic Disorders*, 19, 691-699.
- Martin, W. R., & Cheng, F.** (2020). Repurposing of FDA-approved toremifene to treat COVID-19 by blocking the Spike glycoprotein and NSP14 of SARS-CoV-2. *Journal of proteome research*, 19(11), 4670-4677.
- Maurya, S. K., Maurya, A. K., Mishra, N., & Siddique, H. R.** (2020). Virtual screening, ADME/T, and binding free energy analysis of anti-viral, anti-protease, and anti-infectious compounds against NSP10/NSP16 methyltransferase and main protease of SARS CoV-2. *Journal of Receptors and Signal Transduction*, 40(6), 605-612.
- Morales, P., Curtis, N. L., Zárate, S. G., Bastida, A., & Bolanos-Garcia, V. M.** (2020). Interfering with mRNA methylation by the 2' O-methyltransferase (NSP16) from SARS-CoV-2 to tackle the COVID-19 disease. *Catalysts*, 10(9), 1023.
- Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J.** (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry*, 30(16), 2785–2791.
- Narkhede, R. R., Cheke, R. S., Ambhore, J. P., & Shinde, S. D.** (2020). The molecular docking study of potential drug candidates showing anti-COVID-19 activity by exploring of therapeutic targets of SARS-CoV-2. *Eurasian Journal of Medicine and Oncology*, 4(3), 185-195.
- Nguyen, H. L., Lan, P. D., Thai, N. Q., Nissley, D. A., O'Brien, E. P., & Li, M. S.** (2020). Does SARS-CoV-2 bind to human ACE2 more strongly than does SARS-CoV?. *The Journal of Physical Chemistry B*, 124(34), 7336-7347.
- O'Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., & Hutchison, G. R.** (2011). Open Babel: An open chemical toolbox. *Journal of cheminformatics*, 3(1), 1-14.

- Pantsar, T., & Poso, A.** (2018). Binding affinity via docking: fact and fiction. *Molecules*, 23(8), 1899.
- Petersen, O. H., Gerasimenko, O. V., & Gerasimenko, J. V.** (2020). Endocytic uptake of SARS-CoV-2: the critical roles of pH, Ca²⁺, and NAADP. *Function*, 1(1), zqaa003.
- Raj, R.** (2021). Analysis of non-structural proteins, NSPs of SARS-CoV-2 as targets for computational drug designing. *Biochemistry and biophysics reports*, 25, 100847.
- Ramírez Hernández, E., Hernández-Zimbrón, L. F., Martínez Zúñiga, N., Leal-García, J. J., Ignacio Hernández, V., Ucharima-Corona, L. E., ... & Zenteno, E.** (2021). The role of the SARS-CoV-2 S-protein glycosylation in the Interaction of SARS-CoV-2/ACE2 and immunological responses. *Viral Immunology*, 34(3), 165-173.
- Rosas-Lemus, M., Minasov, G., Shuvalova, L., Inniss, N. L., Kiryukhina, O., Brunzelle, J., & Satchell, K. J.** (2020). High-resolution structures of the SARS-CoV-2 2'-O-methyltransferase reveal strategies for structure-based inhibitor design. *Science signaling*, 13(651).
- Rosas-Lemus, M., Minasov, G., Shuvalova, L., Inniss, N. L., Kiryukhina, O., Wiersum, G., ... & Satchell, K. J.** (2020). The crystal structure of nsp10-nsp16 heterodimer from SARS-CoV-2 in complex with S-adenosylmethionine. *BioRxiv*.
- Salmaso, V., & Moro, S.** (2018). Bridging molecular docking to molecular dynamics in exploring ligand-protein recognition process: an overview. *Frontiers in pharmacology*, 9, 923.
- Santos, I. D. A., Grosche, V. R., Bergamini, F. R. G., Sabino-Silva, R., & Jardim, A. C. G.** (2020). Antivirals against coronaviruses: candidate drugs for SARS-CoV-2 treatment?. *Frontiers in microbiology*, 11, 1818.
- Sastry, G. M., Adzhigirey, M., Day, T., Annabhimoju, R., & Sherman, W.** (2013). Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *Journal of computer-aided molecular design*, 27(3), 221-234.
- Scannell, J. W., Blanckley, A., Boldon, H., & Warrington, B.** (2012). Diagnosing the decline in pharmaceutical R&D efficiency. *Nature reviews Drug discovery*, 11(3), 191-200.
- Schrödinger Suite**, Schrödinger, LLC, New York, NY, 2021.
- Sen, D., Debnath, P., Debnath, B., Bhaumik, S., & Debnath, S.** (2020). Identification of potential inhibitors of SARS-CoV-2 main protease and spike receptor from 10 important spices through structure-based virtual screening and molecular dynamic study. *Journal of Biomolecular Structure and Dynamics*, 1-22.
- Sexton, N. R., Smith, E. C., Blanc, H., Vignuzzi, M., Peersen, O. B., & Denison, M. R.** (2016). Homology-based identification of a mutation in the coronavirus RNA-dependent RNA polymerase that confers resistance to multiple mutagens. *Journal of virology*, 90(16), 7415-7428.

- Shang, J., Ye, G., Shi, K., Wan, Y., Luo, C., Aihara, H., ... & Li, F.** (2020). Structural basis of receptor recognition by SARS-CoV-2. *Nature*, 581(7807), 221-224.
- Shelley, J. C., Cholleti, A., Frye, L. L., Greenwood, J. R., Timlin, M. R., & Uchimaya, M.** (2007). Epik: a software program for pK_a prediction and protonation state generation for drug-like molecules. *Journal of computer-aided molecular design*, 21(12), 681-691.
- Shim, J. S., & Liu, J. O.** (2014). Recent advances in drug repositioning for the discovery of new anticancer drugs. *International journal of biological sciences*, 10(7), 654.
- Sliwoski, G., Kothiwale, S., Meiler, J., & Lowe, E. W.** (2014). Computational methods in drug discovery. *Pharmacological reviews*, 66(1), 334-395.
- QikProp User Manual** Copyright © 2015 QikProp, version 4.4, Schrödinger, LLC, New York, NY, 2015.
- Terali, K., Baddal, B., & Gülcan, H. O.** (2020). Prioritizing potential ACE2 inhibitors in the COVID-19 pandemic: Insights from a molecular mechanics-assisted structure-based virtual screening experiment. *Journal of Molecular Graphics and Modelling*, 100, 107697.
- Torres, P. H., Sodero, A. C., Jofily, P., & Silva-Jr, F. P.** (2019). Key topics in molecular docking for drug design. *International journal of molecular sciences*, 20(18), 4574.
- Trott, O., & Olson, A. J.** (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry*, 31(2), 455-461.
- Xue, H., Li, J., Xie, H., & Wang, Y.** (2018). Review of drug repositioning approaches and resources. *International journal of biological sciences*, 14(10), 1232.
- Van Rossum, G., & Drake Jr, F. L.** (1995). *Python reference manual*. Amsterdam: Centrum voor Wiskunde en Informatica.
- Veeramachaneni, G. K., Thunuguntla, V. B. S. C., Bobbillaipati, J., & Bondili, J. S.** (2020). Structural and simulation analysis of hotspot residues interactions of SARS-CoV 2 with human ACE2 receptor. *Journal of Biomolecular Structure and Dynamics*, 1-11.
- Vrbanac, J., & Slauter, R.** (2017). ADME in drug discovery. In *A Comprehensive Guide to Toxicology in Nonclinical Drug Development* (pp. 39-67). Academic Press.
- Wang, Y., Grunewald, M., & Perlman, S.** (2020). Coronaviruses: an updated overview of their replication and pathogenesis. *Coronaviruses*, 1-29.
- Wang, Z., Sun, H., Yao, X., Li, D., Xu, L., Li, Y., ... & Hou, T.** (2016). Comprehensive evaluation of ten docking programs on a diverse set of protein-ligand complexes: the prediction accuracy of sampling power and scoring power. *Physical Chemistry Chemical Physics*, 18(18), 12964-12975.

- Woo, H., Park, S. J., Choi, Y. K., Park, T., Tanveer, M., Cao, Y., ... & Im, W.** (2020). Developing a fully glycosylated full-length SARS-CoV-2 spike protein model in a viral membrane. *The Journal of Physical Chemistry B*, 124(33), 7128-7137.
- Woo, P. C., Huang, Y., Lau, S. K., & Yuen, K. Y.** (2010). Coronavirus genomics and bioinformatics analysis. *viruses*, 2(8), 1804-1820.
- World Health Organization (WHO)** (January 2020). Novel Coronavirus(2019-nCoV) Situation Report – 10
- World Health Organization (WHO)** (Retrieved January 2022). WHO Coronavirus (COVID-19) Dashboard.
- Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C. L., Abiona, O., ... & McLellan, J. S.** (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*, 367(6483), 1260-1263.
- Yadav, R., Chaudhary, J. K., Jain, N., Chaudhary, P. K., Khanra, S., Dhamija, P., ... & Handu, S.** (2021). Role of Structural and Non-Structural Proteins and Therapeutic Targets of SARS-CoV-2 for COVID-19. *Cells*, 10(4), 821.
- Yan, R., Zhang, Y., Li, Y., Ye, F., Guo, Y., Xia, L., ... & Zhou, Q.** (2021). Structural basis for the different states of the spike protein of SARS-CoV-2 in complex with ACE2. *Cell research*, 31(6), 717-719.

APPEDICES

APPENDIX A: Dendritic fingerprinting hierarchical clustering tables.

APPENDIX B: Desmond MD simulations figures.



APPENDIX A

Table A.1: 163 ligand clusters generated by CANVAS dendritic fingerprinting hierarchical clustering for Spike RBD.

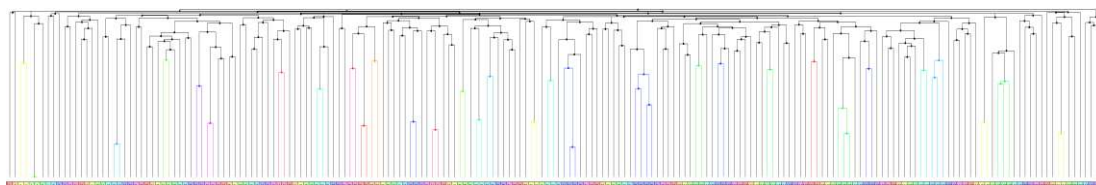


Table A.2: 166 ligand clusters generated by CANVAS dendritic fingerprinting hierarchical clustering for ACE2.

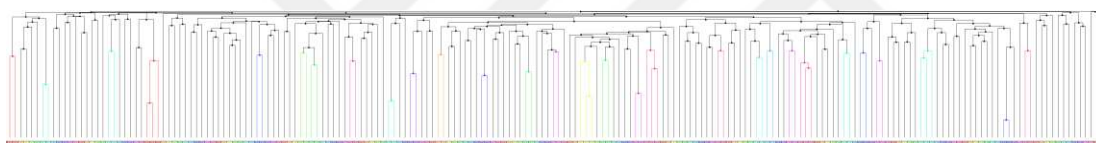


Table A.3: 161 ligand clusters generated by CANVAS dendritic fingerprinting hierarchical clustering for Nsp10 binding site of Nsp16.

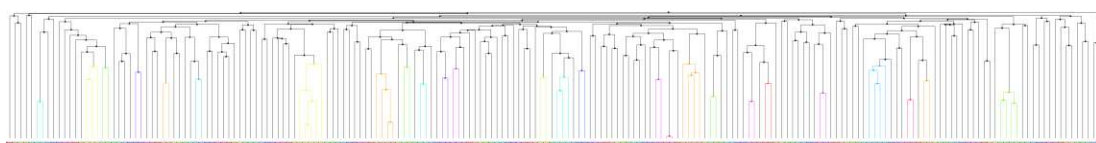
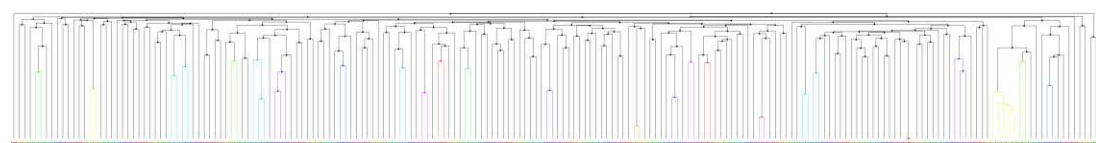


Table A.4: 164 ligand clusters generated by CANVAS dendritic fingerprinting hierarchical clustering for SAM binding site of Nsp16.



APPENDIX B

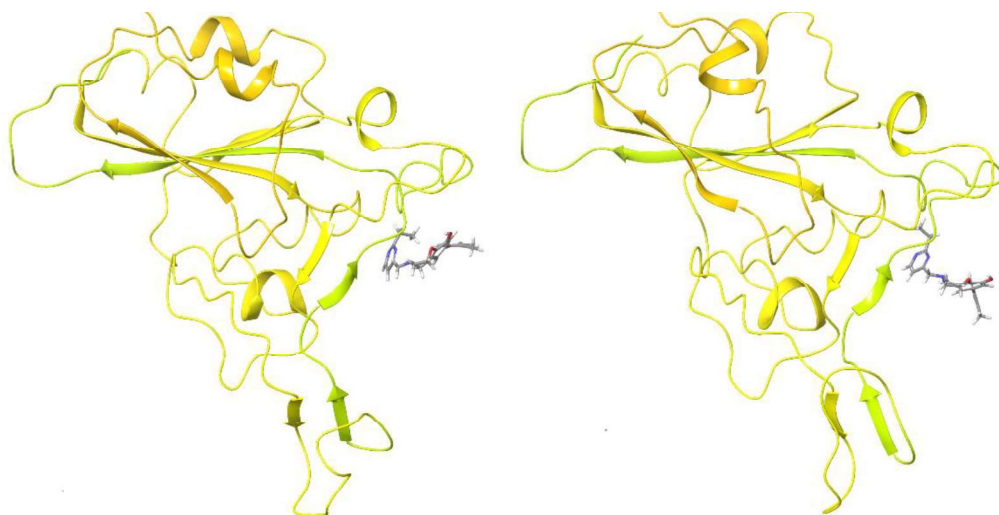


Figure B.1: Comparison of the first and the last frame of ZINC000067920242 within the ACE2 binding site of Spike RBD.

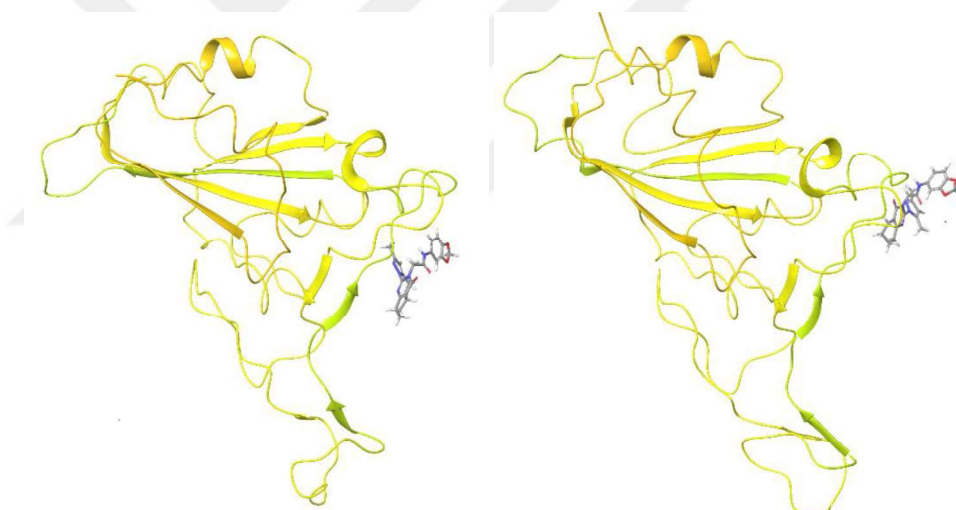


Figure B.2: Comparison of the first and the last frame of ZINC000009828805 within the ACE2 binding site of Spike RBD.



Figure B.3: Comparison of the first and the last frame of ZINC000014228954 within the ACE2 binding site of Spike RBD.

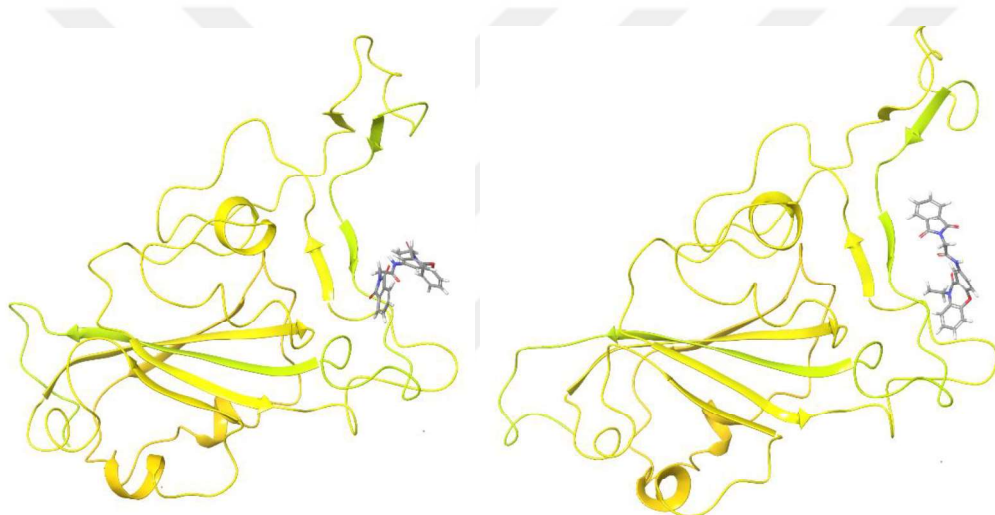


Figure B.4: Comparison of the first and the last frame of ZINC000009134748 within the ACE2 binding site of Spike RBD.

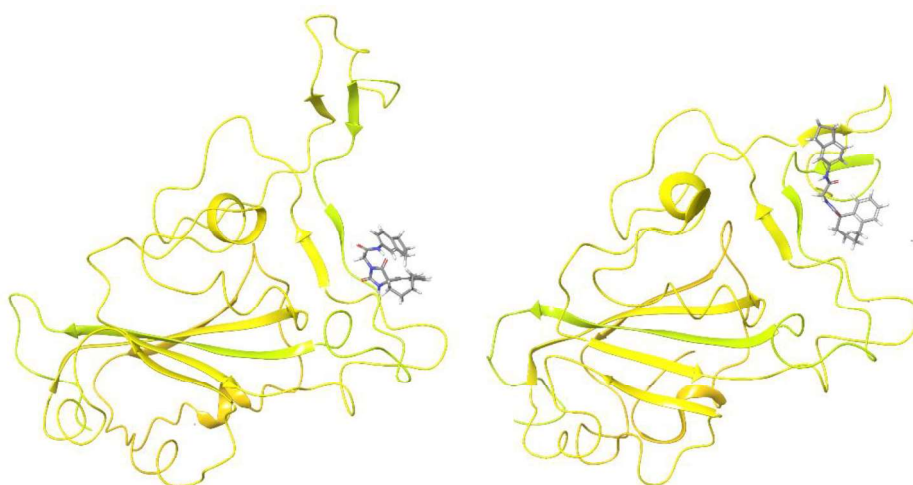


Figure B.5: Comparison of the first and the last frame of ZINC000008772453 within the ACE2 binding site of Spike RBD.

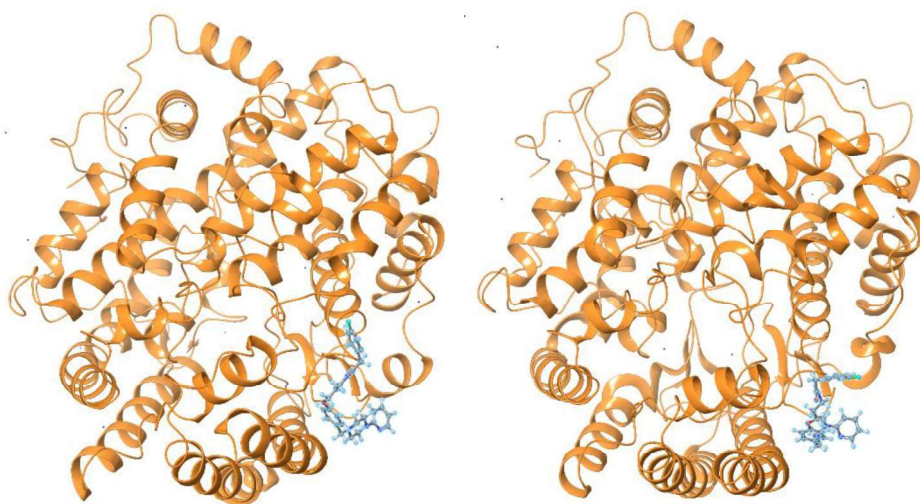


Figure B.6: Comparison of the first and the last frame of ZINC000014879822 within the Spike RBD binding site of ACE2.

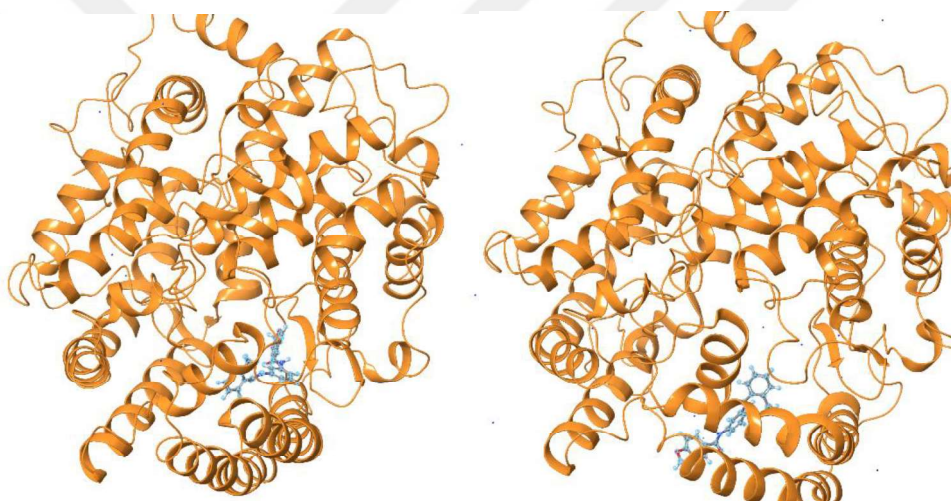


Figure B.7: Comparison of the first and the last frame of ZINC000000813934 within the Spike RBD binding site of ACE2.



Figure B.8: Comparison of the first and the last frame of ZINC000019742660 within the Spike RBD binding site of ACE2.

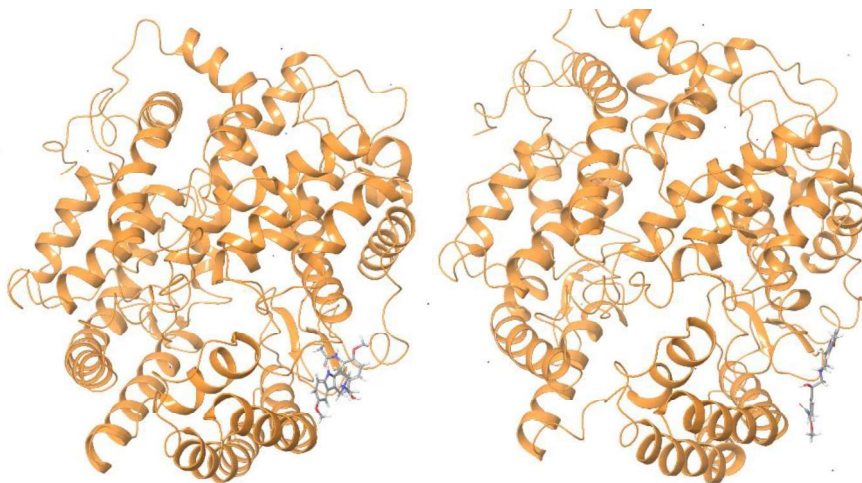


Figure B.9: Comparison of the first and the last frame of ZINC000019022189 within the Spike RBD binding site of ACE2.



Figure B.10: Comparison of the first and the last frame of ZINC000020149756 within the Spike RBD binding site of ACE2.

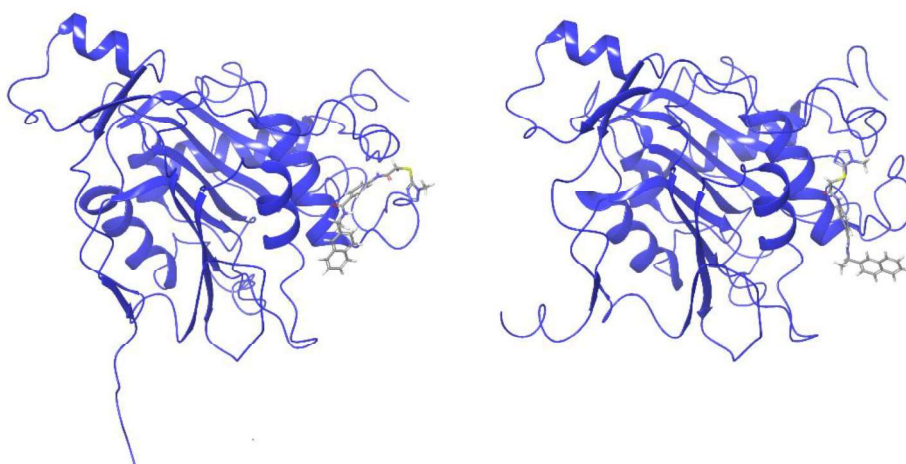


Figure B.11: Comparison of the first and the last frame of ZINC000009751488 within the Nsp10 binding site of Nsp16.

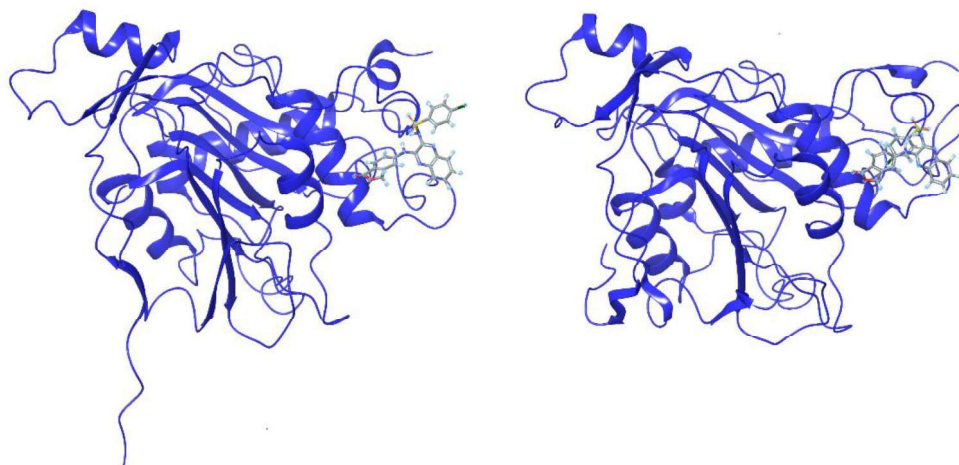


Figure B.12: Comparison of the first and the last frame of ZINC000002286111 within the Nsp10 binding site of Nsp16.

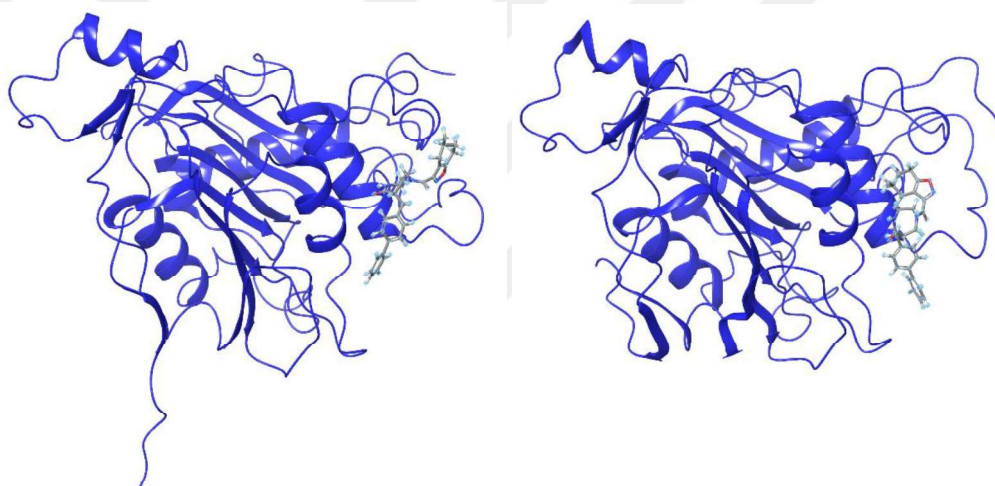


Figure B.13: Comparison of the first and the last frame of ZINC000019497136 within the Nsp10 binding site of Nsp16.

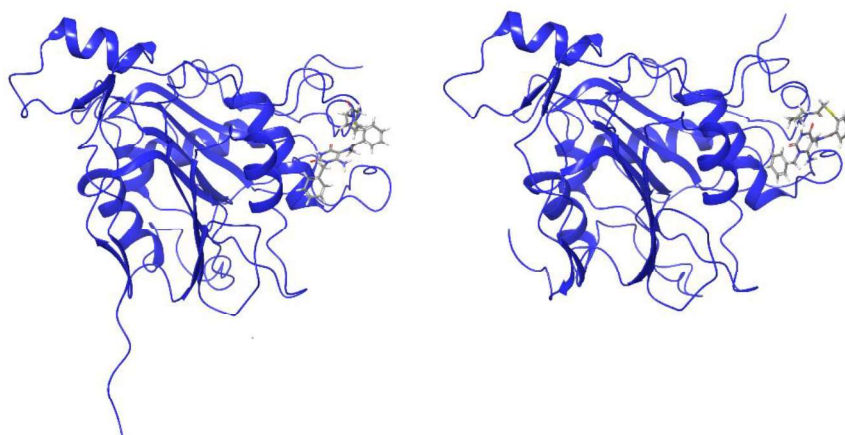


Figure B.14: Comparison of the first and the last frame of ZINC000009252898 within the Nsp10 binding site of Nsp16.

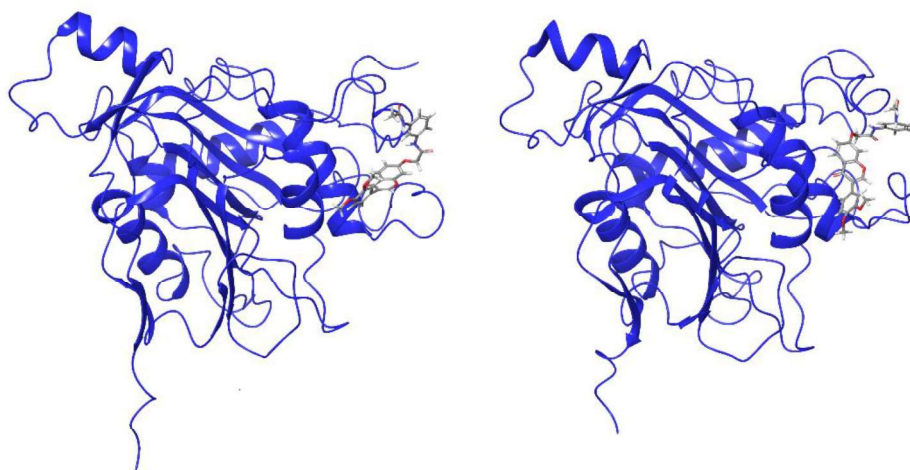


Figure B.15: Comparison of the first and the last frame of ZINC000009485446 within the Nsp10 binding site of Nsp16.



Figure B.16: Comparison of the first and the last frame of ZINC000019373581 within the SAM binding site of Nsp16.

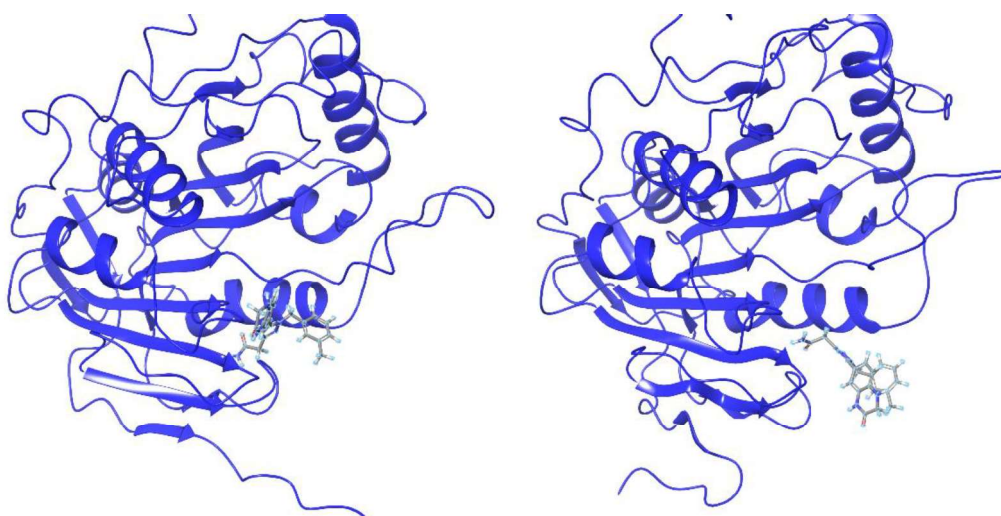


Figure B.17: Comparison of the first and the last frame of ZINC000096137984 within the SAM binding site of Nsp16.

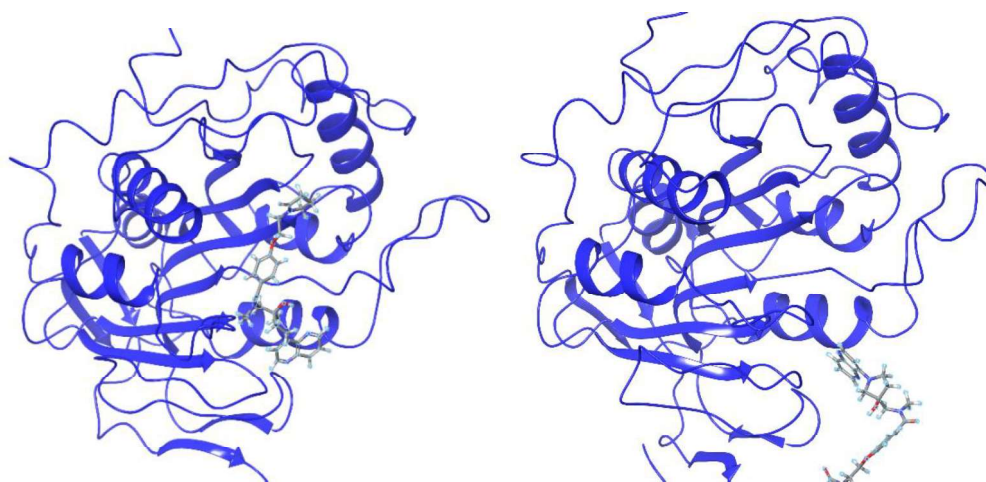


Figure B.18: Comparison of the first and the last frame of ZINC000257313608 within the SAM binding site of Nsp16.

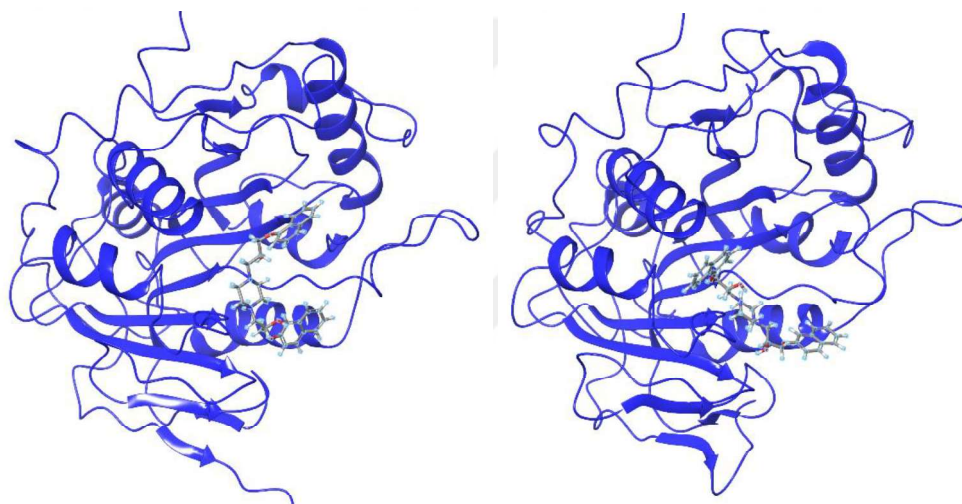


Figure B.19: Comparison of the first and the last frame of ZINC000022289086 within the SAM binding site of Nsp16.

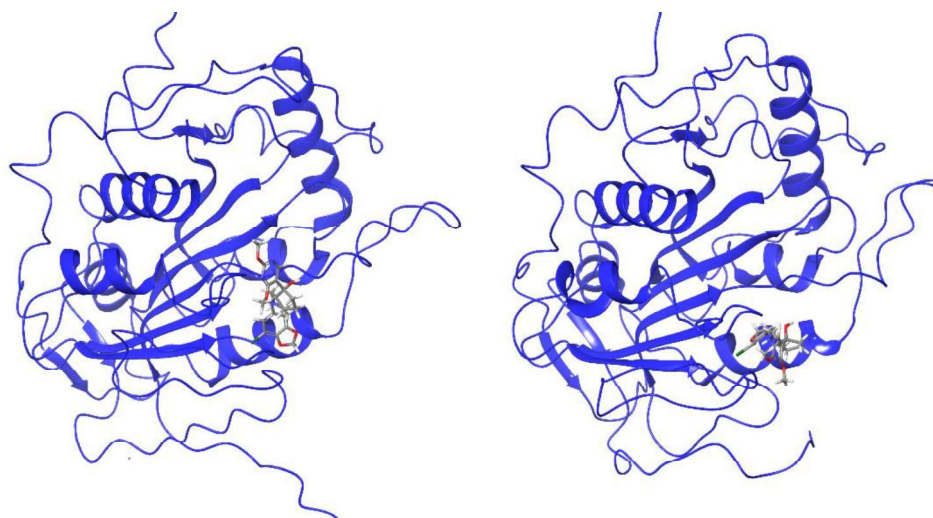


Figure B.20: Comparison of the first and the last frame of ZINC000072395414 within the SAM binding site of Nsp16.



CURRICULUM VITAE

Name Surname : Onur Özer

EDUCATION :

- **B.Sc.** : 2016, Bilkent University, Science Faculty, the Department of Molecular Biology and Genetics, Cum-GPA: 3.70/4.00

PROFESSIONAL EXPERIENCE AND REWARDS:

2018-2022 M.Sc. at Istanbul Technical University, Supervised by Assoc. Prof. Mert GÜR and Asst. Prof. Sefer BADAY – PREDICTING NOVEL SMALL INHIBITORS OF SARS-COV-2: TARGETING SARS-COV-2 SPIKE PROTEIN, HUMAN ACE2 PROTEIN AND SARS-COV-2 NSP16 VIA MOLECULAR DOCKING.

2019-2021 TÜBİTAK National Scholarship Programme for MSc Students

2015-2016 100% Merit Tuition and fee scholarship, Bilkent University

2015 ERASMUS+ Summer Internship Grant

2011-2015 50% Tuition and fee scholarship, Bilkent University

2013-2016 High honour certificates, Bilkent University

2012 Honour certificate, Bilkent University

PUBLICATIONS, PRESENTATIONS AND PATENTS IN THE THESIS

- M. Anwaar Nazeer, İsmail C Karaoglu, **Onur Ozer**, Cem Albayrak, Seda Kizilel “Neovascularization of engineered tissues for clinical translation: where we are, where we should be?” APL Bioengineering (This paper has been selected as Editor’s Pick), Feb 2021.