9-27-2021

Identifying Inhibitors Targeting the Nonstructural Protein 15 and Main Protease of Coronaviruses Using Molecular Docking and Molecular Dynamics Simulation

Nakoa Kristen Webber
Rowan University

IDENTIFYING INHIBITORS TARGETING THE NONSTRUCTURAL PROTEIN 15 AND MAIN PROTEASE OF CORONAVIRUSES USING MOLECULAR DOCKING AND MOLECULAR DYNAMICS SIMULATION

by

Nakoa Kristen Webber

A Thesis

Submitted to the
Department of Molecular and Cellular Biosciences
College of Science and Mathematics
In partial fulfillment of the requirement
For the degree of
Master of Science in Bioinformatics
at
Rowan University
July 23, 2021

Thesis Chair: Nathaniel V. Nucci Ph.D

Committee Members:
Benjamin Carone Ph.D
Yong Chen Ph.D

Dedication

I would like to dedicate this thesis to my mentor Dr. Nathaniel Nucci for his support that made all the difference and teaching me how to science.

Acknowledgments

Funding for this thesis was supported by NSF grant MCB 1942957 and by the NASA New Jersey Space Grant Consortium.

I would like to thank my thesis chair and mentor Dr. Nathaniel Nucci as well as my committee members Dr. Benjamin Carone and Dr. Yong Chen for their support throughout my master's degree and throughout composing my thesis. I would also like to acknowledge Dr. Chun Wu for teaching me computational methods as well as providing laboratory resources for this work.

I would like to acknowledge members of the Nucci lab as well as the Wu lab for their contributions to the projects presented in this thesis including Alyssa Sanders, Annie Tran, Phillip Lakernick, Katherine Hausman, and Julia Gabriel.

Nucci lab members (past and present) have created a lab environment where new scientists are encouraged and supported. Their continuing support has played a key role in my progress as a student and researcher.

I am extremely grateful for the support of my fiancé Jonas throughout my journey as a master's student and beyond. I would also like to acknowledge the support from friends who went above and beyond in the support of my education including Kristin Fesko, Chris Roselle, Sera Bayruns, and Adam Stuhltrager.

Abstract

Nakoa Kristen Webber
IDENTIFYING INHIBITORS TARGETING THE NONSTRUCTURAL PROTEIN 15 AND MAIN PROTEASE OF CORONAVIRUSES USING MOLECULAR DOCKING AND MOLECULAR DYNAMICS SIMULATION
2020-2021
Nathaniel V. Nucci Ph.D
Master of Science in Bioinformatics

The pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in 2020 has impacted daily life globally for over a year. While multiple vaccines have been authorized for emergency use and one oral medication has entered clinical trials, we are still seeking antiviral drugs for a long-term treatment for SARS-CoV-2 as well as other coronaviruses. Computational drug screenings of two SARS-CoV-2 protein target candidates are presented in this thesis: the nidoviral RNA uridylate-specific endoribonuclease (Nsp15) and the main protease (Mpro) of SARS-CoV-2. Nonstructural proteins of coronaviruses were selected as targets as they are more conserved across coronavirus strains than structural proteins. High throughput virtual screening of small molecule libraries including DrugBank and ZINC 15 resulted in several promising compounds for each of these targets. Molecular dynamics simulation allowed us to predict the binding energies for these compounds using molecular mechanics with generalized born surface area solvation calculations (MM-GBSA). Four top compounds were discovered for Nsp15 and eight compounds for Mpro.

Table of Contents

Table of Contents (Continued)

Table of Contents (Continued)

List of Figures

List of Tables

Chapter 1

Introduction

Coronaviruses

Coronaviruses are pathogens which target the respiratory system of both humans and animals (Shereen et al., 2020). The coronaviruses known to infect humans are all zoonotic in nature (Latinne et al., 2020). Currently, there are seven known human coronaviruses (H-CoVs) (Shereen et al., 2020). Most H-CoVs have origins in bats, with the exception of beta-CoVs, which are thought to originate from rodents (Forni et al., 2017). The most common and pathogenic coronaviruses that threaten human life are severe acute respiratory syndrome (SARS-CoV), middle east respiratory syndrome MERS-CoV, and SARS-CoV-2 (Liu et al., 2020).

Zoonotic viruses such as SARS-CoV2 remain a threat to public health as they all have potential to evolve into pandemics. There are two key factors that lead a virus to become a pandemic. First, the virus must be introduced into a population (Santacroce et al., 2020). Humans are exposed to a new virus through contact with blood, feces, saliva, food, or water contamination, or via an arthropod such as a mosquito or a tick (Woolhouse et al., 2012). Next, the virus must spread and maintain itself throughout that population (Santacroce et al., 2020). Virus spread is measured by a R_o value. A R_o greater than 1 indicates that a single case leads to more than one additional case (Woolhouse et al., 2012). Transmission of the virus through a population depends on how infectious the host is the duration of infectability, and the behavior of the infected population (Woolhouse et al., 2012). The most recent impactful coronaviruses (SARS-CoV, MERS-

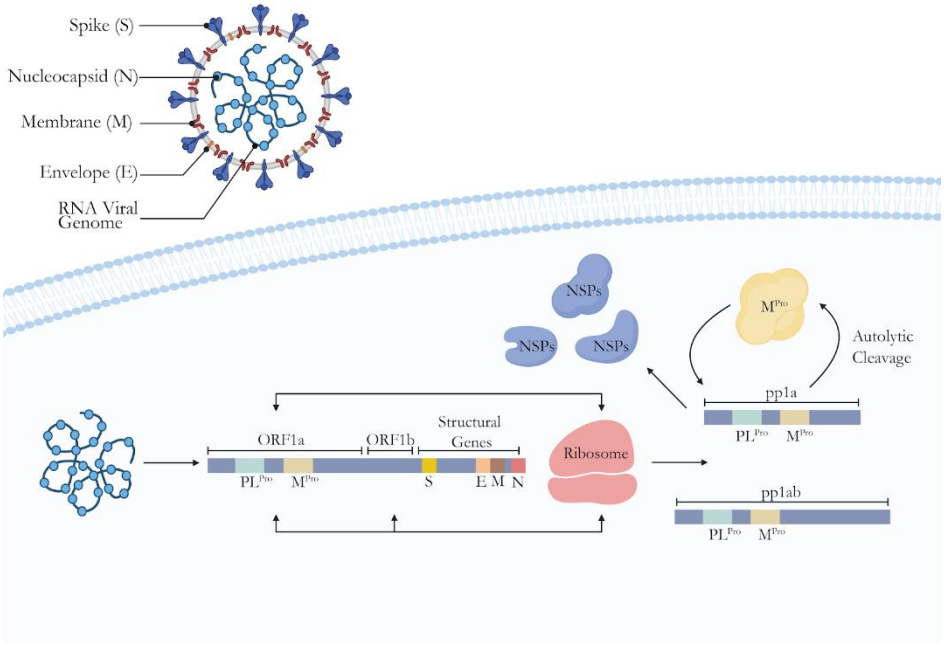
CoV, and SARS-CoV-2) had R_o values of 2.9, 1.3 and 5.7 prior to public intervention (Sanche et al., 2020) (Liu et al., 2020). Social distancing efforts to prevent the spread of coronaviruses by lowering R_o values is disruptive and devastating to local economies. Further understanding of these viruses and efficacious methods of treatment will directly save lives. We hope to be prepared should the world be faced with a similar (or worse) pandemic in the future.

SARS-CoV, MERS-CoV, and SARS-CoV-2

Coronaviruses SARS-CoV, MERS-CoV and SARS-CoV-2 are structurally and genetically conserved. Each virus is a positive-sense, single-stranded RNA virus (Alexandersen et al., 2020). While most RNA viruses tend to lack the 3' exonuclease proofreading capabilities that make DNA viruses less error-prone, Coronaviruses contain a 3' exonuclease domain found in Nsp14 (Sanjuán & Domingo-Calap, 2016). The function of this exonuclease is likely responsible for the maintenance of a genome of this size (Robson et al., 2020). The genomes of coronaviruses contain approximately 30,000 nucleotides with the majority of the genome coding for the replicase gene of the coronavirus (Sanders et al., 2021). The replicase gene consists of two open reading frames (ORF), ORF1a which encodes for polyprotein (pp) 1a and ORF1b. ORF1a and ORF1b encode for polyprotein pp1ab together. These polyproteins are cleaved into 16 nonstructural proteins (Nsps) by the chymotrypsin-like (Mpro) and papain-like (PLpro) proteases. The PLpro cleaves at five sites, working in the N-terminal direction, Nsp1-4. The Mpro, which is Nsp5, cleaves the polyprotein at eleven sites working in the C-terminal direction. Mpro auto-cleaves itself from the polyprotein, then cleaves Nsp6-16 (Sanders et al., 2021) (**Figure 1**).

Figure 1

Replication Cycle of SARS-CoV2



Note. In the replication cycle of SARS-CoV-2, viral RNA is released into the cell. From the RNA viral genome, ORF1a and ORF1b are translated into polyproteins pp1a and pp1ab. To begin the cleavage of nonstructural proteins, the M_{pro} first cleaves itself. Nonstructural proteins next combine to begin the replication process (Sanders et al., 2021).

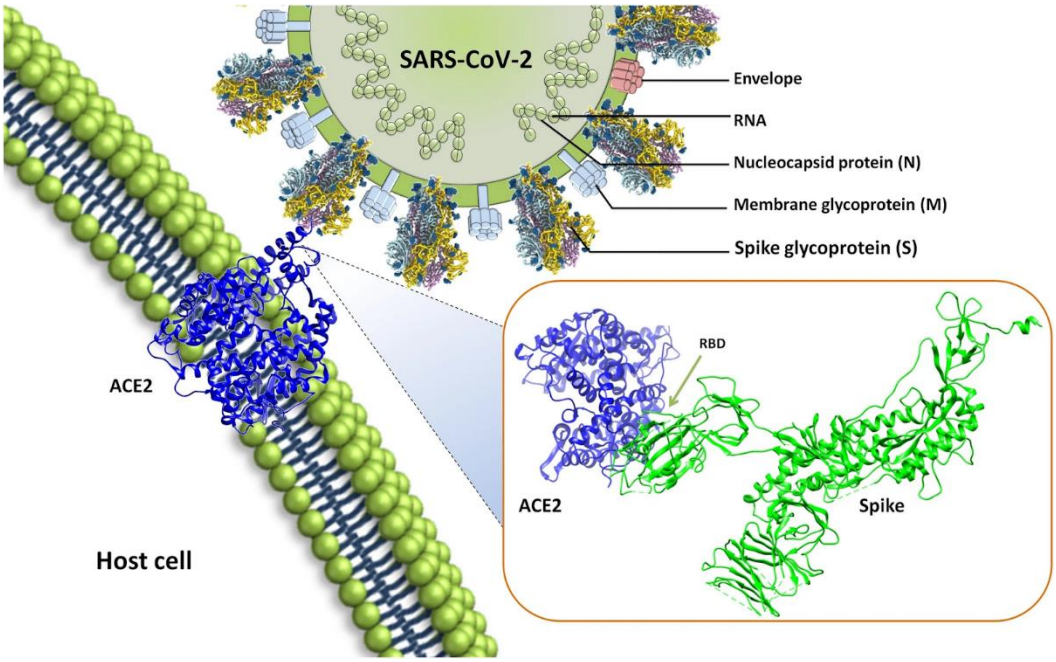
Structural Proteins of SARS-CoV2

There are four main structural proteins within SARS-CoV2: a spike glycoprotein, small envelope glycoprotein, membrane glycoprotein, and a nucleocapsid protein (Astuti & Ysrafil, 2020). The spike glycoprotein is a transmembrane protein approximately 150 kDa in size and is located on the outside of the virus structure (Astuti & Ysrafil, 2020). The spike protein is responsible for host-virus attachment. Small envelope glycoproteins form homotrimers within the viral surface which facilitate binding of envelope viruses to

host cells by attracting angiotensin-converting enzyme 2 (ACE2) (Astuti & Ysrafil, 2020). The small envelope glycoprotein is cleaved by a host cell furin-like protease into S1 and S2 (Astuti & Ysrafil, 2020). The S1 subunit contains a receptor-binding domain which recognizes and binds to the host receptor angiotensin-converting enzyme 2 (Huang et al., 2020). The S2 subunit forms a six-helical bundle that plays a role in mediating viral cell membrane fusion (Huang et al., 2020).

Figure 2

Structural Proteins of SARS-CoV2



Note. The structural proteins of SARS-CoV2 include nucleocapsid proteins, membrane glycoproteins, spike glycoproteins, and envelope proteins. The spike protein is shown in this figure interacting with ACE2 for host attachment (Saxena et al., 2020).

Nonstructural Proteins of SARS-CoV2

The nonstructural proteins of SARS-CoV2 play an important role in genome replication and transcription of the virus (Gasmalbari & Abbadi, 2020). Each Nsp plays a role in virus infection or replication as shown in **Table 1**. Nsp's are promising targets for vaccines and medications as targeting one or more can reduce or halt viral replication. Most nonstructural proteins have conserved amino acid sequences among different coronaviruses which would allow one drug to target them all.

Table 1

Nonstructural Proteins of Coronaviruses and their Function

NSP1	Suppresses host gene expression by degrading the host cell's RNA (Sanders et al., 2021), prevents hosts cells from performing antiviral functions (Gasmalbari & Abbadi, 2020)
NSP2	Thought to play a crucial role in viral RNA synthesis (Sanders et al., 2021), the specific mechanism of action requires further research
NSP3	Papain-like protease (Plpro) protein responsible for processing of the viral polypeptide (Cornillez-Ty et al., 2009)
NSP4	Required for viral replication by assembly of, and localizing to, double-membrane cytoplasmic vesicles (with nsp3) (Sakai et al., 2017)
NSP5	Mpro (3CL), main protease, cleaves at 11 sites (Zhang et al., 2020)
NSP6	Induces double membrane vesicles, interferes with delivery of viral factors to lysosomes for destruction (Angelini et al., 2013)
NSP7	Forms a super complex with NSP8, works as a cofactor for the RNA-dependent RNA polymerase nsp12 (Snijder et al., 2016)
NSP8	Forms a super complex with NSP 7 that supports viral replication works as a cofactor for the RNA-dependent RNA polymerase nsp12 (Snijder et al., 2016)

NSP 9	Plays a role in the binding of ssRNA and dsDNA and affects viral growth (Sanders et al., 2021)
NSP 10	Forms a part of the viral mRNA cap methylation complex (Sanders et al., 2021)
NSP 11	Function of NSP 11 is unknown
NSP 12	Responsible for priming the dependent RNA polymerase (Sanders et al., 2021)
NSP 13	A helicase with RNA and DNA unwinding capabilities, possesses dNTPase activity, helping form the 5' cap of viral mRNA (Sanders et al., 2021)
NSP 14	Responsible for proofreading during RNA replication and viral mRNA capping (Snijder et al., 2016)
NSP 15	Processes viral RNA, aids in evasion of hosts defense system (Hsu 2021)
NSP 16	Adds Nsp10 and Nsp14 to form the mRNA cap methylation complex. (Sanders et al., 2021)

Current Variants of SARS-CoV-2

As of June 2021, there are currently seven variants of interest of novel SARS-CoV-2: B.1.525 originally detected in the United Kingdom/Nigeria (2020), B.1.526 and B.1526.1 originally detected in the United States (2020), B.1617, B.1.617.1, and B.1.617.3 originally detected in India (2021 and 2020), and P.2 originally detected in Brazil (2020) (CDC). A variant is classified as of interest if it contains genetic markers that have been associated with changes to receptor binding, reduced neutralization by antibodies generated against previous infection or vaccination, reduced efficacy of treatments, potential diagnostic impact, or predicted increase in transmissibility or disease severity (CDC 2020).

According to the CDC there are currently 6 variants of concern: B.1.1.7 originally detected in UK(2020), B.1.351 originally detected in South Africa (2020), B.1.427 and B.1.429 originally detected in the US (2021), B.1.617.2 originally detected in India (2020), and P.1 originally detected in Japan/Brazil (2021). A variant is considered to be of concern when there is evidence of increased transmissibility, increased severity of disease, significant reduction in neutralization by antibodies generated during previous infection or vaccination, reduced effectiveness of treatments or vaccines, or diagnostic detection failures (CDC 2020).

In silico Screening for a SARS-CoV-2 Inhibitor

Lack of treatment or vaccine for SARS CoV-2 led to extreme loss of life and economic disruption in the years 2020 and 2021 (V'kovski et al., 2020). Scientists from all over the world are searching diligently for an effective inhibitor-target duo to lessen the severity of this virus (V'kovski et al., 2020). One beneficial starting point in the search for a potential drug suitable as a specific inhibitor for SARS-CoV-2 is in silico screening (Chandra et al., 2020). When searching for potential inhibitors for SARS-CoV-2, researchers are initially screening previously developed compounds in libraries including the Food and Drug Administration (FDA) approved drug database for structure based virtual screening (SBVS) as well using molecular docking to compare drugs that are used in the treatment of other viruses (Chandra et al., 2020). The first strategy is designing the compound based on the existing broad spectrum of antivirals. The advantage of this approach is that the established pharmacological properties for these compounds allow them to be readily used (Baby et al., 2021). The downside to this strategy is the limitation of available compounds within the library search.

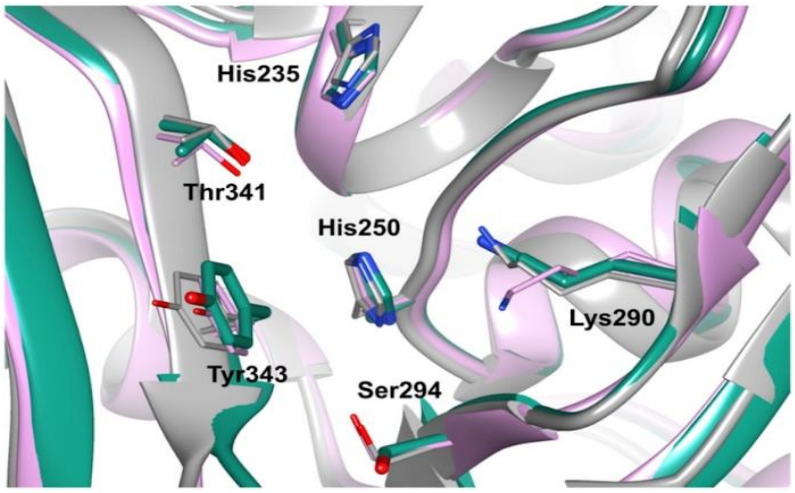
In our study, we investigated small molecules and their potential to inhibit Nsp15 and the Mpro (Nsp 5) of SARS-CoV2 using in-silico screening. Targeting Nsp proteins for inhibition as opposed to the structural proteins of coronaviruses is advantageous due to strong conservation across viral strains. The spike protein, for example, is highly mutagenic (Zhou et al., 2021). All variants of SARS-CoV-2 (as classified by the CDC) contain spike protein amino acid substitutions. In a study by Khan et al, a multiple sequence alignment was performed comparing sequences of the SARS-CoV-2 Mpro from thirteen different countries. From these alignments, only a single point mutation was found from the Vietnam strain (Khan et al., 2020). There are no known mutations of the Nsp15 protein in any of the new SARS-CoV-2 strains (CDC 2020). Our objective is to find an inhibitor not only for SARS-CoV2, but one with potential to treat its variants and future coronaviruses as well.

Nonstructural Protein 15 of SARS-CoV2

NSP 15 is a nidoviral RNA uridylate-specific endoribonuclease (NendoU). Nsp 15 contains a carboxy-terminal catalytic domain that functions by cleaving RNA at the 3′-position of uridylates to form a 2′-3′ cyclic phosphodiester product (Chandra et al., 2020). The structure of Nsp15 is 84.38 kDa in size (Kim et al., 2020). The C- terminal NendoU domain contains two antiparallel β-sheets (β16–β17–β18 and β19–β20–β21) with the active site between them. The active site carries six key residues which are conserved among SARS-CoV-2, SARS-CoV, and MERS-CoV coronaviruses (**figures 3 & 4**) : His235, His250, Lys290, Ser 294, Thr341, Tyr343) and is thought to play a role in processing viral RNA as well as aiding in the evasion of host defense mechanisms (Kim et al., 2020).

Figure 3

Active Site Overlay of Nsp 15 of SARS-CoV, MERS, SARS-CoV-2

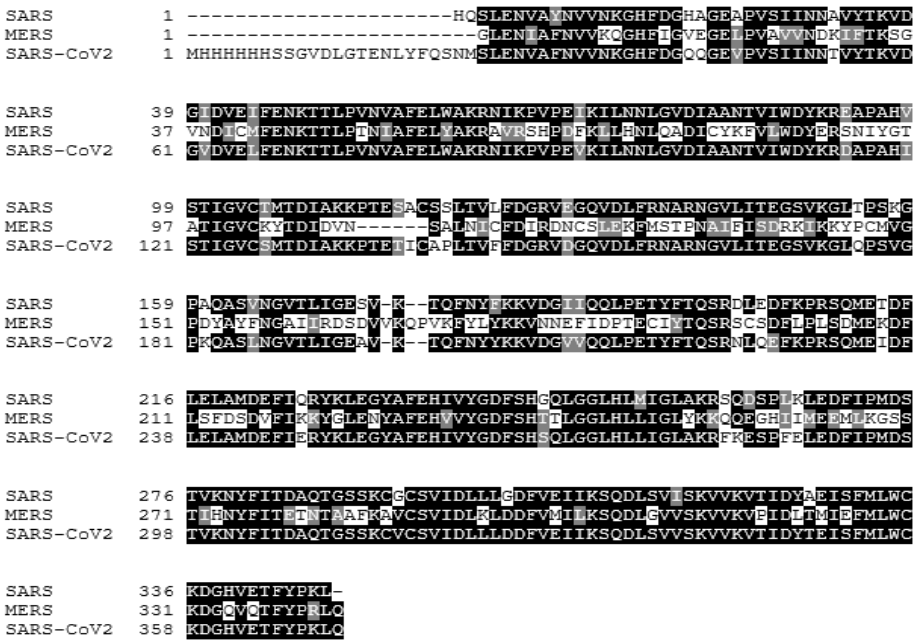


Note. An active site overlay of Nsp 15 of SARS-CoV2 (teal) SARS (pink), and MERS (grey) shows structural similarity among these Coronaviruses. Key residues involved in ligand binding (His235, His250, Lys290, Ser294, Thr341, and Tyr343) are highlighted (Kim et al., 2020).

While its exact function is unclear, conservation among coronaviruses, confirmed by multiple sequence alignment presented in **Figure 4**, suggests that Nsp15 is essential for viral replication. Previous studies suggest that there are multiple Nsp15 cleavage targets that are important to regulate the accumulation of viral RNA and prevent activation of RNA-activated antiviral responses (Pillon et al., 2021) (Ancar et al., 2020). By targeting this protein with a small molecule inhibitor, we can disrupt the viral replication process or potentially prevent the evasion of the body's immune response to the virus.

Figure 4

Multiple Sequence Alignment of Nsp15 of SARS, MERS, and SARS-CoV2



Note. Amino acid multiple sequence alignment of Nsp15 of SARS, MERS, and SARS-CoV2 performed using T-Coffee (Notredame et al., 2000). Amino acids highlighted in black are conserved among SARS, MERS, and SARS-CoV2. Individual amino acids highlighted in grey show amino acids with similar characteristics or properties to those in the same position on different coronaviruses. Figure created using BoxShade (https://embnet.vital-it.ch/software/BOX_form.html).

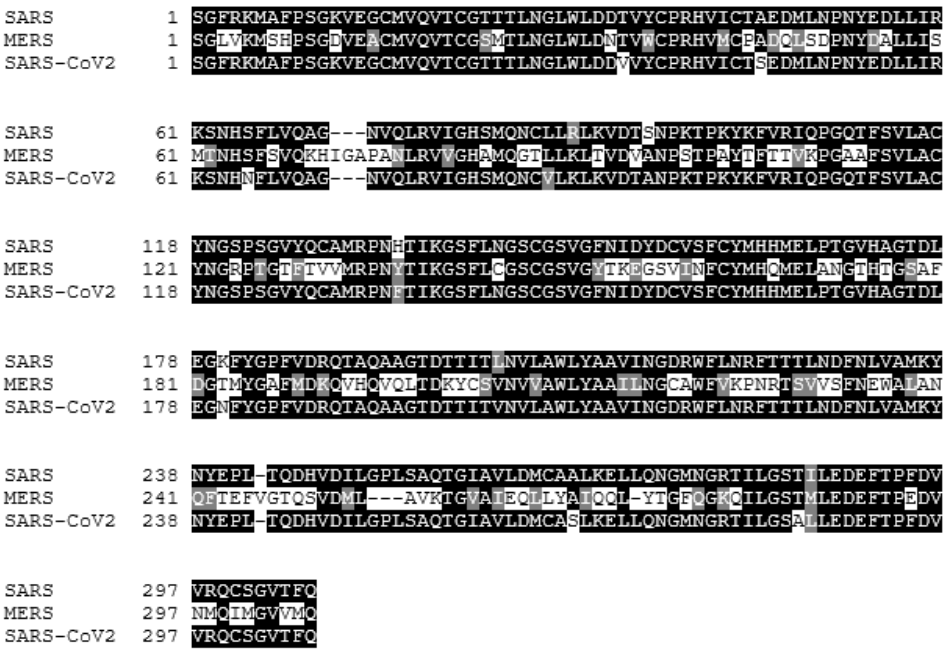
Main Protease of SARS-CoV2

The Mpro (also referred to as 3CLpro) was selected due to its significant role in viral replication by processing polypeptides after translation from viral RNA (Zhang et al., 2020). This enzyme is ~34 kDa in size (Kneller et al., 2020). The Mpro's function is cleaving 11 sites on polyprotein 1ab to continue the replication cycle which suggests inhibition would block replication by preventing cleavage of non-structural proteins from pp1a and pp1ab (Zhang et al., 2020) (Sanders et al., 2021). As one of the most

characterized targets of the SARS-CoV2 virus, the Mpro displays high genetic and structural conservation with earlier coronaviruses with no human analogues (Mengist et al., 2021). A multiple sequence alignment shown in **Figure 5** confirms the conserved amino acid sequence of the Mpro across coronaviruses SARS, MERS, and SARS CoV2.

Figure 5

Multiple Sequence Alignment of Mpro of SARS, MERS, and SARS-CoV2

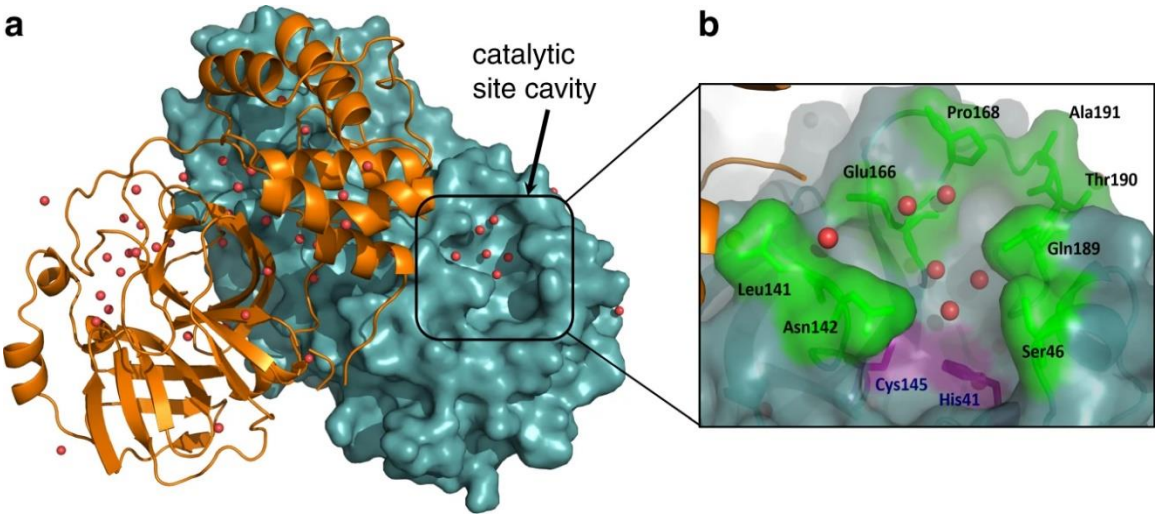


Note. Amino Acid multiple sequence alignment (MSA) of the Mpro of SARS, MERS, and SARS CoV2. Amino acids highlighted in black are conserved among SARS, MERS, and SARS-CoV2. Individual amino acids highlighted in grey show amino acids with similar characteristics or properties to those in the same position on different coronaviruses. MSA performed by T-Coffee (Notredame et al., 2000). Figure created using BoxShade.

Structurally, the Mpro is functional as a homodimer with each monomer consisting of three domains (Sanders et al., 2021). The active site is formed by antiparallel beta sheets from domains I (residues 1-101) and II (residues 102-184) containing a Cys-His dyad (**Figure 6 A, B**). Eight residues (Ser 46, leu141, Asn 142, Glu 166, Pro168, Ala191, Thr190, and Gln189) surround the active site and play roles in ligand binding (**Figure 6 B**). Domain III (residues 185-200) contains five alpha helices and facilitates dimerization (Kneller et al., 2020) (Verschueren et al., 2008). Dimerization is essential for function as the monomer is not catalytically active (Kneller et al., 2020).

Figure 6

Active Site of Mpro



Note. The three-dimensional structure of the SARS-CoV-2 Mpro. **A.** The dimerized Mpro is shown with one monomer in orange and the other (containing the catalytic domain) is shown in teal. **B.** The catalytic site of Mpro is shown with the catalytic dyad of Cys145 and His41 colored purple. Nearby residues interacting within the binding pocket are shown in green. (Figure created by Kneller et al., 2020 and is used under Creative

First Small Molecule Inhibitor Targeting Mpro

Pfizer recently announced its first antiviral small molecule inhibitor targeting the Mpro of SARS-CoV2, PF-07321332. This inhibitor is currently in Phase 1 clinical trials (NCT04535167) and is delivered orally (Halford, 2020). Scientists originally discovered this compound in an attempt to target SARS in 2002, and with most targets of coronaviruses remaining structurally conserved, they began testing this compound for SARS-CoV2. Early data suggests that compound PF-07321332, which is a prodrug, releases active antiviral compound PF-00835231in tissue (Halford, 2020). PF-00835231 exhibits antiviral activity in multiple strains of SARS-CoV2 as well as other coronaviruses suggesting the possibility of a broad spectrum therapeutic for coronaviruses (Halford, 2020). Currently, the projected effective dose of this proposed inhibitor is 500mg/day delivered intravenously which is considered high.

Conclusion

With its first small molecule inhibitor in clinical trials, the Mpro of coronaviruses remains a promising drug target. We continue to search for to a small molecule inhibitor for the Nsp15 of Coronaviruses. Structures of these key proteins involved in coronavirus replication are well characterized making them a prime candidate for high throughput virtual screening to discover an inhibitor. Using high throughput virtual screening, molecular docking, and molecular dynamics simulation, we have presented several promising compounds from the Zinc15 library targeting the NSP15 and Mpro proteins.

Chapter 2

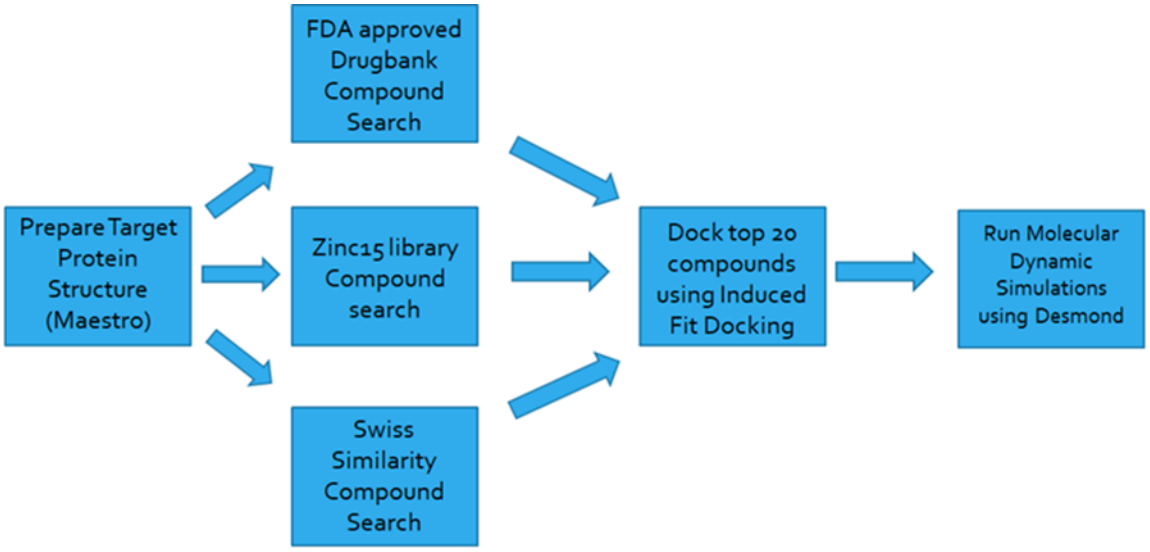
Methods

Introduction

In Silico high-throughput virtual screening provided us with a method to search for compounds with potential to bind favorably to our protein targets (Nsp15 and Mpro) for inhibition. Because the Nsp15 and the Mpro structures are well characterized, we were able to prepare their structures from PDB files 6WXC and 6LU7 respectively in maestro for molecular docking. Using pharmacophore screening of each protein's natural ligand, we searched the FDA approved drug bank, Zinc15 library, and the SwissSimilarity library for compounds with structural similarities to each natural ligand (Zoete et al., 2016). Following the workflow presented in **Figure 7**, compounds were narrowed down to top candidates which were first docked using Glide docking in Maestro. Top candidates were then induced fit docked to each target protein where those with the best docking scores were evaluated by molecular dynamics (MD) simulations. The molecular mechanics generalized Born surface area solvation calculation (MM-GBSA) was used to predict the binding energy of each ligand to the Nsp15 and Mpro.

Figure 7

High Throughput Virtual Screening Workflow



Note. High throughput virtual screening workflow starting with protein structure preparation using Maestro, compound search of three different libraries, molecular docking to identify hit compounds, and finally, molecular dynamics simulation.

Multiple Sequence Alignment

Multiple sequence alignment (MSA) of the amino acid sequences of the Nsp15 and Mpro of SARS, MERS and SARS-CoV2 was performed using T-Coffee (Notredame et al., 2000). MSA allowed us to confirm structural conservation of key residues involved in binding among these different coronaviruses.

Protein Structure and Receptor Grid Preparation

The structures of Nsp 15 and of Mpro were retrieved from the Protein Data Bank as PDB ID: 6WXC 6LU7, respectively. The protein structure was prepared for molecular docking using the Maestro protein preparation wizard (Madhavi Sastry et al., 2013).

Water molecules beyond 5 Å from the surface of each protein were deleted, and both proteins were optimized at pH 7.0. The optimized proteins were subject to a restrained minimization to relax the protein structure using an OPLS3 force field. The active site was located by the center of the natural ligand for receptor grid preparation.

Ligand Preparation

The three-dimensional ligand structures were prepared using Maestro Elements 2.2, a feature within the Maestro 10.2 software. Ionization/tautomeric states were generated at pH =7 using EPIK which employs the refined Hammett and Taft method (Madhavi Sastry et al., 2013). The lowest ionization/tautomeric state was selected. The ligand structure was relaxed via restrained minimization in Maestro.

Glide Docking of the Natural Ligand

The natural ligand of both Nsp15 and the Mpro of SARS-CoV2 was first docked using Glide dock. Docking the natural ligand back to the protein allows us to validate our docking methods. Glide XP docking searches for the most favorable ligand-receptor conformations for a protein-drug complex. Standard Glide dock was used to dock each crystal ligand into its respective receptor grid under default parameters.

High Throughput Virtual Screening: Zinc Drug-Like Library

To search for promising small molecule inhibitors, virtual libraries were searched using the natural ligand as a model for structure similarity. Libraries included within the search for Nsp15 inhibitors were SwissSimilarity, DrugBank, and Zinc15 (Sterling & Irwin, 2015). The library searched for Mpro inhibitors was Zinc15. Initial screening of

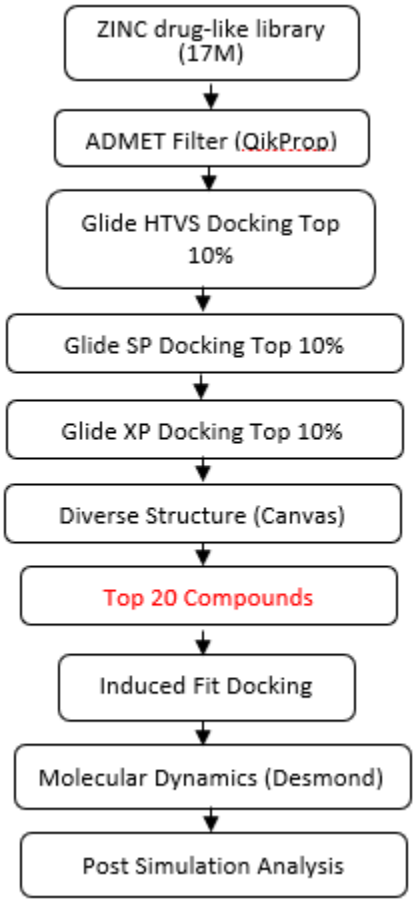
over 7 million compounds presented several promising compounds with structural similarity to the natural ligand for both protein targets.

QikProp ADME Filtering of Compounds

QikProp from Schrodinger was used as an ADME (absorptions, distribution, metabolism, excretion) filter to quickly reduce molecular candidates which fall outside of the normal range of known drug limits of properties such as molecular weight, logP value, .and oral absorption.

Figure 8

Workflow of Virtual Screening Compounds for Inhibition of Coronaviruses



Note. Workflow of virtual screening compounds for inhibition of coronaviruses

Glide Docking

Promising molecules for inhibition were docked to the prepared Nsp15 (PDB: 6WXC) (Kim et al., 2020) and Mpro (PDB:6LU7) (X. Liu et al., 2020) protein structures using the Glide feature of Maestro software. Glide is a quick and rigid method of molecular docking which calculates a docking score based on protein-ligand coulomb-vdW energy with a small contribution from GlideScore (Friesner et al., 2004). The

GlideScore is a value meant to separate molecules with strong versus weak binding potential (Friesner et al., 2004). While the docking scoring function used by Glide is proprietary, we do know it rewards Van der Waals and Coulomb energy contributions, hydrophobic terms, hydrogen bonds, and polar interactions within the binding site. Glide penalizes the inhibition of rotatable bonds and other binding interactions (Friesner et al., 2006). Molecules with docking scores greater than the score of the natural ligand were considered for further evaluation. Compounds with high structure similarity were eliminated to diversify the screening.

Induced Fit Docking

Induced fit docking was performed on the top 19 compounds of Nsp15 and top 30 compounds of Mpro to validate top compounds by allowing conformational changes within the binding pocket of the protein. Induced fit docking files were ordered by highest docking score and processed for molecular dynamics simulation.

Molecular Dynamics Simulation

Molecular dynamics simulations were performed using Desmond Molecular Dynamics simulation by Schrodinger. Molecular dynamics allows us to observe how compounds may bind to a protein target by estimating kinetics and binding energy. Each system was solvated in an orthorhombic water box using the SPC water model with a 10 Å water buffer (Mark & Nilsson, 2001). To neutralize the systems, Na+ ions were added with a salt concentration of 0.15 M NaCl. After successful solvation of each system, the OPLS3 force field (Harder et al., 2015) was used to represent the receptor-ligand complex. For each system, the default relaxation protocols were followed in the

Desmond simulation package (Jorgensen et al., 1996). Simulations were run for 200 nanoseconds at a temperature of 300 K.

Protein and Ligand Root-Mean-Square Deviation of Atomic Position (RMSD) Calculation

Root mean square deviation calculations were performed for both the protein and ligand for conformation of binding stability throughout the simulation using equation 1:

$$\text{RMSD} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(r_i'(t_x)\right) - r_i(t_{ref}))^2)} \quad (1)$$

Where N is the number of atoms in the atom selection; t_{ref} is the reference time, (the first frame is used as the reference as time t=0); and r_i' is the position of the selected atoms in frame x after superimposing on the reference frame, where frame x is recorded at time t_x. This calculation is repeated for every frame in the simulation trajectory.

Protein RMSD values were expected to stabilize toward the end of each simulation which indicated a steady state within the binding pocket. Ligand RMSD values significantly larger than the protein RMSD suggest the ligand has likely moved from the binding site.

Protein Root Means Square Fluctuation (RMSF) Calculation

Protein RMSF was calculated to determine individual amino acid changes throughout the simulation. RMSF was calculated using equation 2:

$$\text{RMSF} = \sqrt{\frac{1}{T}\sum_{t=1}^{T} < \left(r_i'(t)\right) - r_i(t_{ref}))^2) >} \quad (2)$$

Where T is the trajectory time over which the RMSF is calculated, t_{ref} is the reference time, r_i is the position of residue i; r_i' is the position of atoms in residue i after superposition on the reference, and the angle brackets indicate that the average of the square distance is taken over the selection of atoms in the residue.

Binding Energy Calculation

Molecular Mechanics-General Born Surface Area (MM-GBSA) binding energies were calculated for the last 10 nanoseconds of the combined trajectory for each system where the RMSD calculation suggested binding stability within the system. This method of binding energy prediction was selected as it balances accuracy and computational power. The OPLS3 force field, VSGB 2.0 solvation model (Li et al., 2011) and the default prime protocol (Cournia et al., 2017) was used to separately minimize the receptor, ligand, and receptor-ligand complex using equation 3 for the total binding free energy:

$$\Delta G_{(bind)} = E_{complex\ (minimized)} - (E_{ligand\ (minimized)} + E_{receptor\ (minimized)}) \quad (3)$$

Binding components (Coulombic + H-bond + GB solvation+ van der Waals + π-π packing + self-contact + hydrophobic) were evaluated into separate groups: $E_{electrostatic}$, E_{vdW}, and $E_{hydrophobic,}$ where ($E_{electrostatic} = E_{coulombic} + E_{H\text{-}bond} + E_{GB\text{-}solvation}$) and ($E_{vdW} = E_{vdW} + E_{pi\text{-}pi\ stacking} + E_{self\text{-}contact}$).

Simulation Interaction Diagram Analysis

The Simulation Interaction Diagram in Desmond and Glide was used to view interactions between the protein and ligand in each simulation. Residues interacting with the protein for at least 30 % of the simulation are highlighted. 2D interaction diagrams are included for proteins and compounds simulated in F**igures 13 and 18**.

ADME Property Prediction

Predicted ADME (absorption, distribution, metabolism, and excretion) properties were evaluated using the Swiss ADME server (http://www.swissadme.ch) to determine the potential for any compounds to be functional as a drug from pharmacokinetic and physicochemical properties (Daina et al., 2017).

Chapter 3

Molecular Docking and Molecular Dynamics Simulation of Inhibitors Targeting Nsp15

Introduction

The Nsp15 protein of Coronaviruses plays a key role in viral replication making it a potential target for a small molecular inhibitor. In this chapter, I present the results of our high throughput virtual screening of small molecule libraries including DrugBank and ZINC 15. Four promising compounds were identified for Nsp15. Molecular dynamics simulation allowed us to predict the binding energies (calculated by MM-GBSA) for these compounds and further understand the interactions supporting binding. Using the SwissADME server we were able to identify structurally based ADME properties that determine the likelihood of our compounds to perform successfully as drugs.

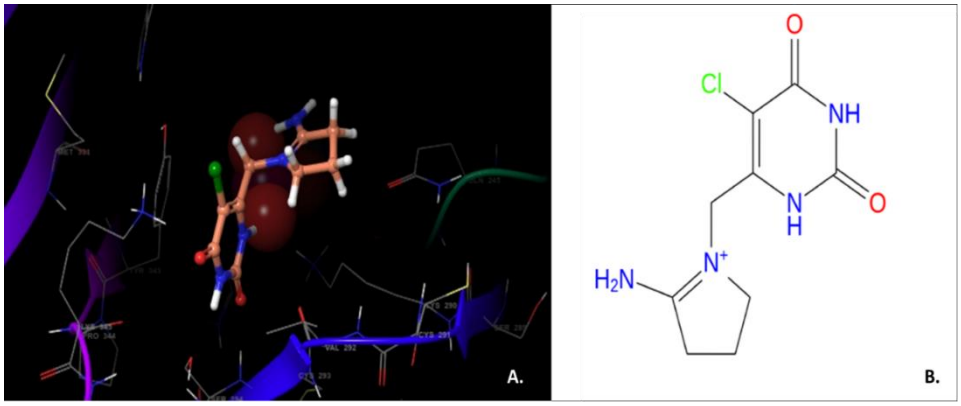
Molecular Docking

Natural Ligand Docking

Using Glide docking software in Maestro, we measured the binding score of the natural ligand of Nsp15 to validate our docking protocol (**Figure 9**). The binding score of the natural ligand docked to the binding site of PDB:6LU7 was calculated at -7.295 kcal/mol. When evaluating compounds for molecular docking, we searched for compounds with preliminary docking scores better (more negative) or equivalent to this value.

Figure 9

NSP15 Complex with its Natural Ligand and 2D Structure



Note. A. Nsp15 is shown in complex with its natural ligand (PDB:6WXC). B. The two-dimensional structure of the natural ligand of Nsp15 is presented.

Glide and Induced Fit Docking

Molecules from the high throughput virtual screening of the DrugBank and Zinc15 databases were docked to NSP15 using GlideXP. Compounds found to have an initial docking score greater than or equal to the docking score of the natural ligand were selected for further evaluation by induced fit docking. Compounds with high structure similarity to each other were eliminated to diversify the screening.

Nineteen compounds from the Zinc15 database were selected for further consideration because their docking scores ranged from -7.33 kcal/mol to -10.32 kcal/mol. The docking score allows us to easily compare binding probability among compounds. No targets were selected for further evaluation from the DrugBank library due to low docking scores. An induced fit docking protocol as described in Chapter 2,

was employed to optimize the docking pose of those compounds for molecular dynamics simulation.

Molecular Dynamics Simulation

Molecular dynamics simulation was performed as described in Chapter 2 to further evaluate binding of hit compounds to NSP15. Output of the MD simulations included docking scores from the previous induced fit docking, MM-GBSA free energy, and RMSD values for the protein and ligand. The RMSD value for each compound averaged over the last 20 ns of the simulation was presented. The last 20 ns were averaged as we expected the complex to be stable for this portion of the simulation. Results of these simulations are presented in **Table 2**.

Table 2

Properties of the Top 19 Compounds were Determined from Molecular Docking and MD Simulations

Zinc ID	Docking Score (kcal/mol)	MM-GBSA (kcal/mol)	Receptor RMSD[1] (Å)	Ligand RMSD[1] (Å)
Natural Ligand	-7.295	-28±2.7	1.67±0.1	4.48±1.0
ZINC000013545806	-10.327	-67.3±5.3	3.14±0.3	22.73±3.7
ZINC000004096690	-10.007	-21.0±4.5	2.87±0.3	44.93±6.8
ZINC000247434422	**-9.799**	**-76.5±6.7**	**4.07±0.5**	**7.97±0.4**
ZINC000004095545	-9.758	-2.7±3.3	2.78±0.7	44.56±8.6
ZINC000030690671	-9.267	-28.7±13.0	2.63±0.3	45.52±7.7
ZINC000004096060	-9.067	-6.6±3.2	6.44±0.6	46.49±6.8
ZINC000043898683	-8.666	-38.3±16.7	2.52±0.2	12.49±3.8
ZINC000043772626	**-8.168**	**-47.5±9.7**	**3.19±0.3**	**5.19±2.5**
ZINC000257311522	-7.708	-41.1±10.0	2.35±0.1	24.65±7.5
ZINC000002573902	-7.66	-33.8±4.6	4.1±0.4	13.2±1.1
ZINC000097814854	-7.633	-40.0±9.0	2.2±0.3	6.92±0.8
ZINC000005811925	-7.585	-19.8±6.9	2.7±0.3	21.0±0.9
ZINC000095447896	-7.478	-30.5±9.9	2.4±0.2	15.9±6.5
ZINC000033902452	**-7.478**	**-49.0±3.3**	**2.7±0.2**	**3.0±0.5**
ZINC000044020013	**-7.463**	**-50.0±3.0**	**2.7±0.3**	**7.4±1.0**
ZINC000584892418	-7.461	-28.3±3.8	3.1±0.3	21.3±1.1
ZINC000257288396	-7.423	-36.0±9.0	4.5±0.5	36.25±10.7
ZINC000009716294	-7.388	-39.1±8.1	2.8±0.3	7.05±2.9
ZINC000014728394	-7.338	-28.9±19.3	2.7±0.2	17.53±3.1

Note. Properties presented are the docking score, MM-GBSA calculation, and Receptor and Ligand RMSD. [1] Receptor and Ligand RMSD values are averaged from the last 20 ns of the simulation. Top promising compounds are shown in bold.

MM-GBSA Binding Energy Calculations Predict Top Compounds for Nsp15 Inhibition

MM-GBSA was used to calculate the enthalpic terms of the binding energy as previously described in Chapter 2. The MM-GBSA binding energy makes several assumptions, including the use of an implicit solvent model. Using an implicit model assumes the solvent is one continuous medium with a specified dielectric. This significantly decreases protein-solvent interactions as compared to an explicit solvent model. These interactions play a non-negligible role in protein-ligand binding and should be considered in the future (J. Zhang et al., 2017). Using this method also removes viscosity from consideration, which speeds up the conformational search of the solvent. Despite these assumptions, MM-GBSA provides an accurate prediction of binding energy, while lowering time and computational costs (Onufriev & Case, 2019).

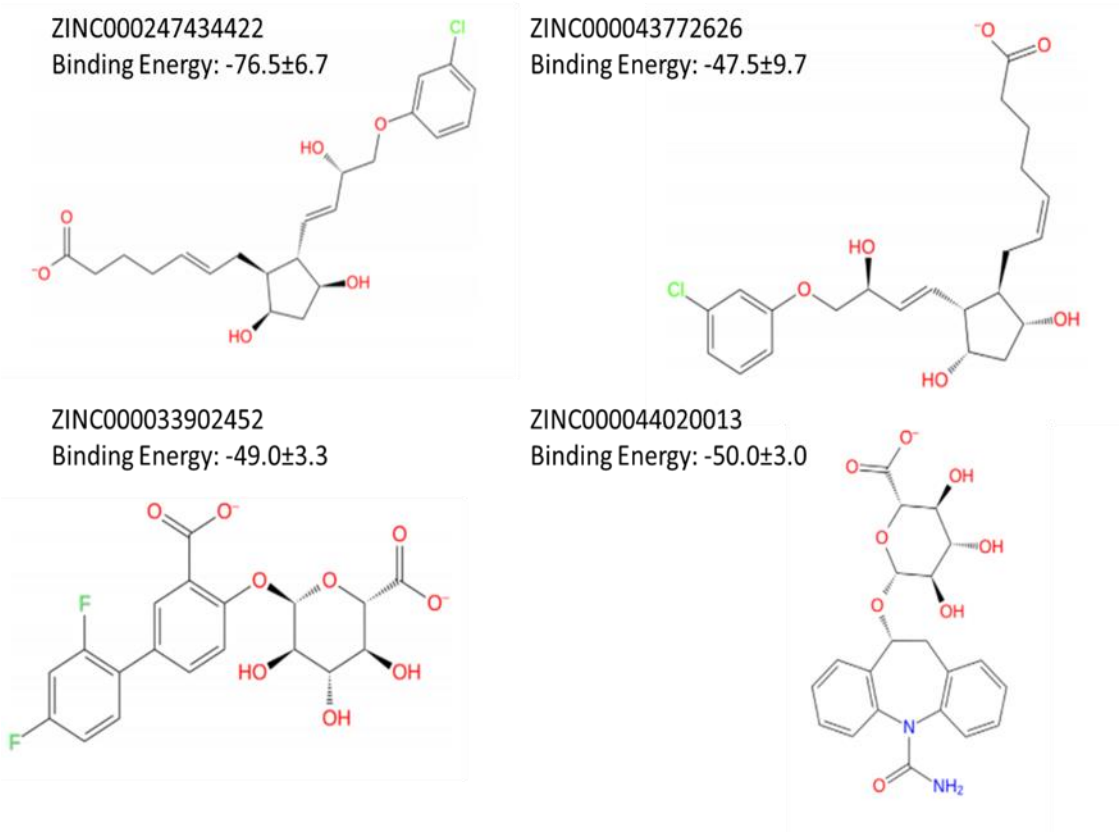
It is important to note that entropic terms are not considered in these calculations due to high computational cost. Conformational entropy is known to contribute significantly (and usually unfavorably) to binding energy (Genheden & Ryde, 2015) (Singh & Warshel, 2010). There are additional computational methods available that can be used to accurately predict this contribution including normal mode analysis. Including entropic terms would offer a more accurate predicted binding energy.

Analysis of the output from the MD simulations showed four promising compounds: ZINC000247434422, ZINC000043772626, ZINC000033902452,

and ZINC000044020013. These compounds were considered to be the most promising ones because they meet the following criteria: a better (more negative) docking score than the natural ligand, a low predicted binding energy as calculated by MM-GBSA, and each remained bound to the protein when the trajectory stabilized. A two-dimensional structure of each top compound with its predicted binding energy is presented in **Figure 10**.

Figure 10

Top Four Compounds and their Predicted Binding Energy



Note. The top four compounds for Nsp15 inhibition were determined by docking score, predicted binding energy, and the ability to remain bound to the protein

when the trajectory stabilized. Predicted binding energies according to MM-GBSA calculations are listed for each compound in kcal/mol.

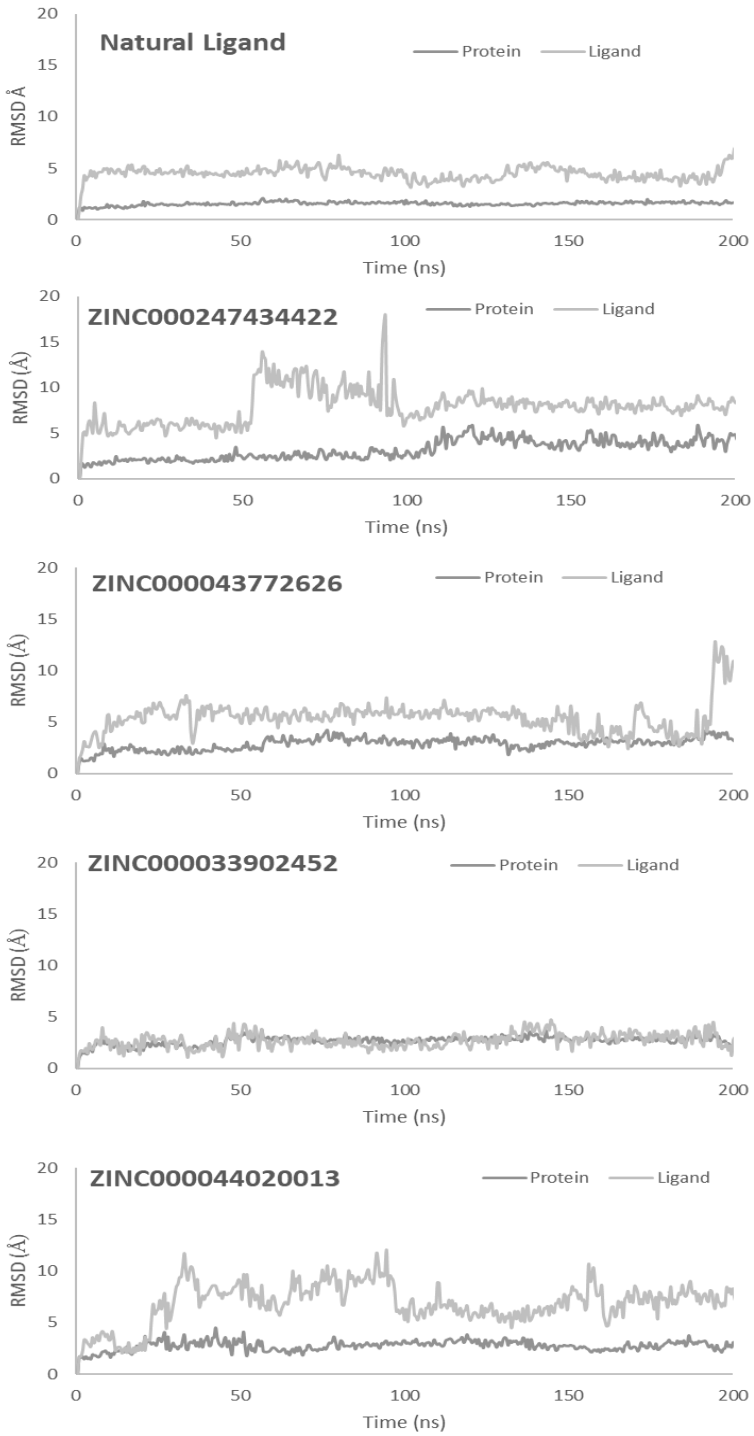
RMSD Plots Show the MD Simulation has Stabilized

Binding stability was evaluated by measuring the RMSD values of each complex throughout the entire simulation. At the end of each simulation, we expect the RMSD values to converge around a fixed value to confirm equilibration of the simulation. The values listed for the RMSD of the protein and ligand in **Table 2** are averaged over the last 20 ns where we expect the simulation to be stabilized. Plots of RMSD values throughout the entire simulation for the most promising compounds are presented in **Figure 11**. Small fluctuations of the RMSD values (between 1-3 Å) are expected throughout the simulation. Larger fluctuations suggest a larger conformational change and may indicate the ligand has moved away from the binding site.

All compounds appear to have stable RMSD values for the last 40 seconds of each simulation except for ZINC000043772626. This compound remained stable for most of the simulation, then saw a sharp increase within the last 10 ns with the ligand RMSD measuring 10 Å. The protein RMSD remained stable at around 4 Å. These findings suggest the ligand may have moved away from the binding pocket. A longer simulation may be needed to reach equilibration for this simulation. All other compounds appear stable for the last 40 ns of the simulation as expected suggesting equilibration.

Figure 11

RMSD of NSP15 Protein and Top 4 Compounds



Note. Root Mean Square Deviation (RMSD) plots are presented for the MD simulation runs of each top protein-ligand complex over the length of the trajectory. The Cα-RMSD for the protein is shown in dark grey and the ligand RMSD is shown in light grey. Cα-RMSD is based on initial protein alignment.

MM-GBSA Binding Energy Components

The predicted binding energy calculation combines energy contributions from non-covalent interactions including Van der Waals, hydrophobic, and electrostatic terms. The predicted binding energy contribution from each of these terms is presented in **Table 3**. These terms are also rewarded by the Glide docking scoring function, so it is expected that compounds with a favorable (more negative) docking score will also have a favorable predicted binding energy. An example of this relationship is shown by compound ZINC000247434422 which presented the most favorable binding energy at -76.5±6.7 kcal/mol and one of the highest docking scores at -9.7. The hydrophobic term provided the greatest contribution to the total binding energy at -40.7±2.3 kcal/mol with an additional heavy contribution from the Van der Waals term at -35.3±0.7 kcal/mol. These contributions to the predicted binding energy can be better understood when we examine them separately.

Table 3

The MM-GBSA Binding Energy Calculation Components Including the Van der Waals, Hydrophobic, and Electrostatic Terms are Presented

Zinc ID	Van der Waals (kcal/mol)	Hydrophobic (kcal/mol)	Electrostatic (kcal/mol)
Natural Ligand	-22.8±2.2	-7.6±1.4	1.9±1.4
ZINC000013545806	-35.1±2.8	-43.4± 2.0	11.2±1.5
ZINC000004096690	-16.6±5.5	-7.0±1.2	2.5±1.1
ZINC000247434422	**-35.3±0.7**	**-40.7±2.3**	**-0.5±5.2**
ZINC000004095545	-2.7±3.4	-1.1±1.3	1.2±1.6
ZINC000030690671	-11.5±4.6	-17.1±8.1	-0.1±7.1
ZINC000004096060	2.2±3.3	-1.2±1.8	-7.7±4.6
ZINC000043898683	-24.0±9.1	-23.6±10.2	9.4±2.4
ZINC000043772626	**-25.9±4.1**	**-24.5±2.1**	**3.0±5.9**
ZINC000257311522	-18.0±3.5	-13.3±5.8	-9.9±2.7
ZINC000002573902	-21.6±2.5	-13.9±1.7	1.8±1.4
ZINC000097814854	-33.3±4.0	-16.8±3.8	10.0±2.7
ZINC000005811925	-10.2±1.7	-2.6±0.8	-7.0±7.5
ZINC000095447896	-22.1±4.2	-12.0±4.4	3.6±3.2
ZINC000033902452	**-23.5±4.8**	**-16.9±0.3**	**-8.6±6.9**
ZINC000044020013	**-29.3±2.2**	**-20.7±1.0**	**0.0±4.6**
ZINC000584892418	-15.3±4.5	-12.1±2.9	-0.8±6.1
ZINC000257288396	-18.5±6.6	-12.6±1.9	-4.9±4.8
ZINC000009716294	-26.3±4.4	-16.6±2.0	3.7±3.9
ZINC000014728394	-19.9±12.4	-12.4±7.3	3.4±2.0

Note. Top promising compounds are shown in bold

As described in chapter 2, the Van der Waals term is the sum of Van der Waals, π-π stacking, and self-contact energies. The hydrophobic term is measured alone while the electrostatic term is the sum of energy contributions made by Coulomb interactions, Hydrogen bonds, and general born solvation energies. A breakdown of these terms allows us to further understand how the ligand interacts with the binding pocket.

The Van der Waals and hydrophobic terms of the predicted binding energy equation dominate ligand binding for these compounds. In Table 3, top compounds all had large contributions from both of these terms. It is expected that Van der Waals interactions to contribute significantly to the binding energy due to the number of these interactions involved in molecular recognition (L. Li et al., 2015). Hydrophobic interactions are extremely valuable and frequent in supporting ligand binding. These interactions drive nonpolar interactions to displace water molecules from interacting surfaces within the binding pocket (Wermuth et al., 2015).

There were minimal favorable and unfavorable contributions to the total binding energy from the electrostatic term in all cases as expected. The energy from hydrogen bonds can vary in intermolecular interactions, and in these simulations their contributions were very small. The contribution from Coulombic energy and binding solvation are similar in magnitude, but opposite in their signs leaving the sum of the total electrostatic contribution to be a small number of either positive or negative value. It is important to note that the electrostatic term is highly dependent on the parameters of the simulation including the force field

chosen, the dielectric constant of the protein and the refinement of the proteins structure (Talley et al., 2008).

The Protein Cα RMSF Confirms the Stability of Compound Binding by Residue

Residue fluctuation was measured throughout the simulation and is presented by RMSF values in F**igure 12**. Higher RMSF values indicate more fluctuation of amino acids throughout the MD simulation. It is common for the N and C terminal tails of the protein to fluctuate significantly, and we observe that for all simulations. We expect that secondary structures (alpha helices and beta strands) remain more rigid with lower RMSF values, while loop regions have a higher rate of fluctuation throughout each simulation.

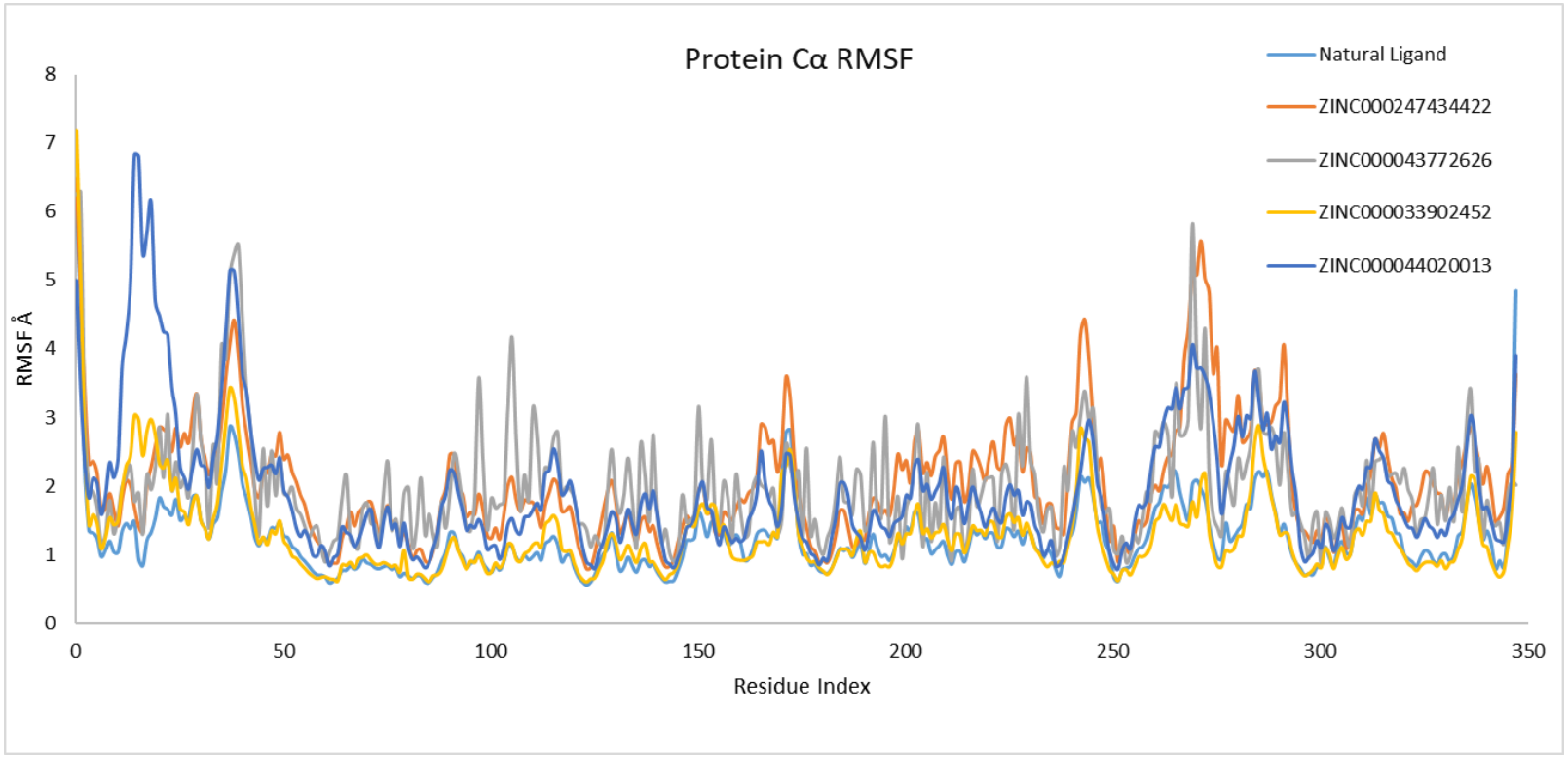
Residues fluctuated similarly for each compound and the natural ligand with a differing fluctuation observed for the complex with compound ZINC000044020013 in residues 13-25. These residues are a part of a loop region between an alpha helix structure and a beta strand. An amino acid sequence of Nsp15 including secondary structure assignments is presented as **Appendix Figure 1.** This particular compound relies heavily on its binding interaction with only its carboxylic acid group. Its binding may not stabilize this loop region as well as the other top compounds allowing for some additional protein flexibility. Compound ZINC000043772626 also shows additional fluctuation between residues 103 through 110 which is a loop region.

Amino acids interacting within the binding pocket, including His235, His250, Lys290, Ser 294, Thr341, Tyr343, appear stable in each simulation with

RMSF values under 3 Å. As expected, loop regions show the most flexibly, this is apparent in regions 45-52 and 256-265 where RMSF values reach approximately 5.5 Å.

Figure 12

Residue Fluctuation Plotted by RMSF



Note. The residue fluctuation plotted by RMSF during binding with each top compound (Cyan: Natural Ligand, Orange: ZINC000247434422, Grey: ZINC000043772626, Yellow: ZINC000033902452, Dark Blue: ZINC000044020013). Residues fluctuated similarly throughout the simulation for all compounds with an exception in residues 13 through 25 for ZINC000044020013.

Simulation Interaction Diagrams Reveal Key Binding Residues for Top Compounds

Key residues supporting compound binding to Nsp15 for at least 30 % of the simulation were identified by the Desmond simulation interaction diagram (**Figure 13**). These interactions are expected to agree with docking score and MM-GBSA binding energy prediction value. As shown previously in **Table 3**, most of the predicted binding energy is composed of the Van der Waals and hydrophobic terms so we expect interactions to support these energy contributions. Most of the interactions displayed in the simulation diagrams were driven by hydrophobic or polar interactions with the binding pocket which agrees with these findings. The residues responsible for these interactions within the binding pocket are His235, His250, Lys290, Ser 294, Thr341, and Tyr343.

For ZINC000247434422, key interactions include polar interactions through water molecules with His250 and Ser 294 for 41% and 39% of the simulation, respectively. Hydrophobic interactions with residues Leu346 and Pro344 were maintained for 42% and 46% of the simulation respectively. This compound had the highest docking score and predicted binding energy which was supported in large by its hydrophobic and Van der Waals energy contributions.

Interactions supporting ZINC000043772626 included a positive charge with Lys290 for 40% and 36% of the simulation with two separate oxygens. A polar

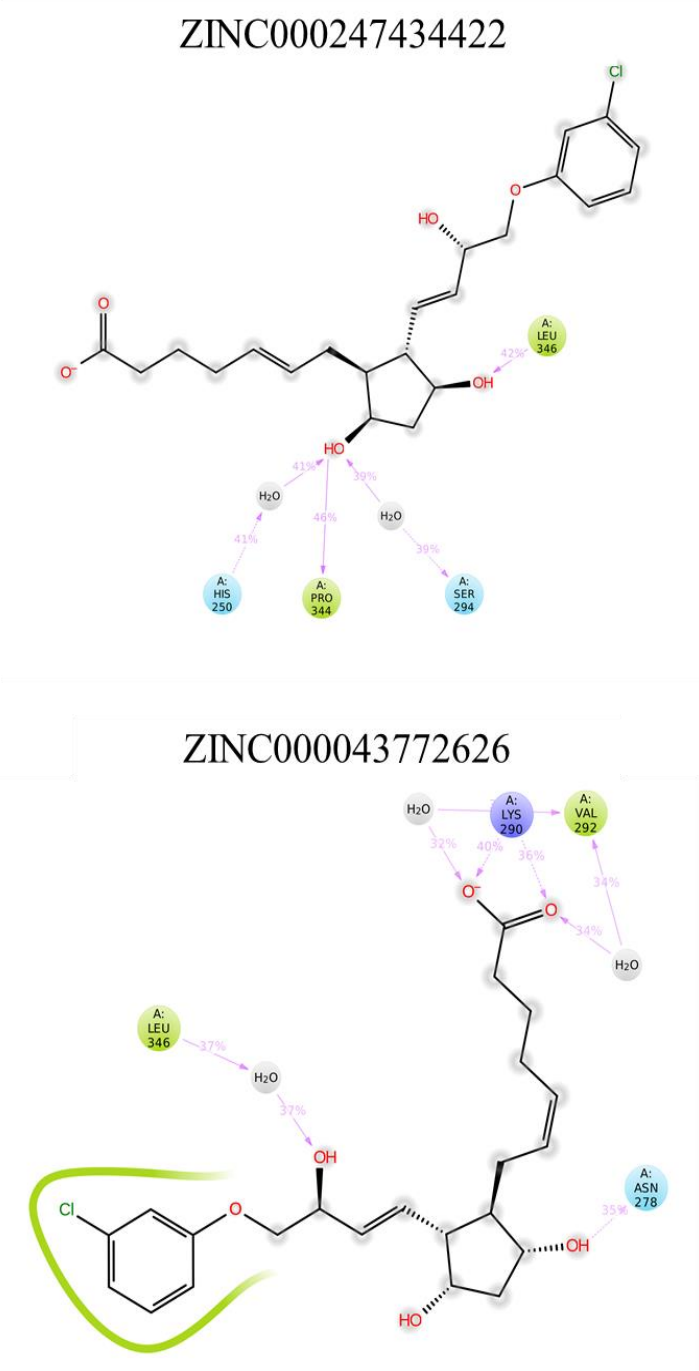
interaction with Asn278 for 35% of the simulation, and hydrophobic interactions with waters and Leu 346 (37%) and Val292 (34%) also contributed to the compounds binding to Nsp15. This compound had similar contributions from hydrophobic and Van der Waals terms.

ZINC000033902452 has a positively charged interaction with Lys290 for 69% of the simulation and Lys 345 for 37% of the simulation. Hydrophobic interactions for ZINC000033902452 include: Val292 (58%), Trp333 (62%), Leu346 through water for 51%, and Tyr343 through interactions with two separate waters for 40% and 36% of the simulation. Polar interactions include His250 through a water for 35% and Ser294 for 57% of the simulation. Several interactions between this compound and the binding pocket are hydrophobic which is reasonable as the predicted binding energy of this compound was supported heavily by its hydrophobic term.

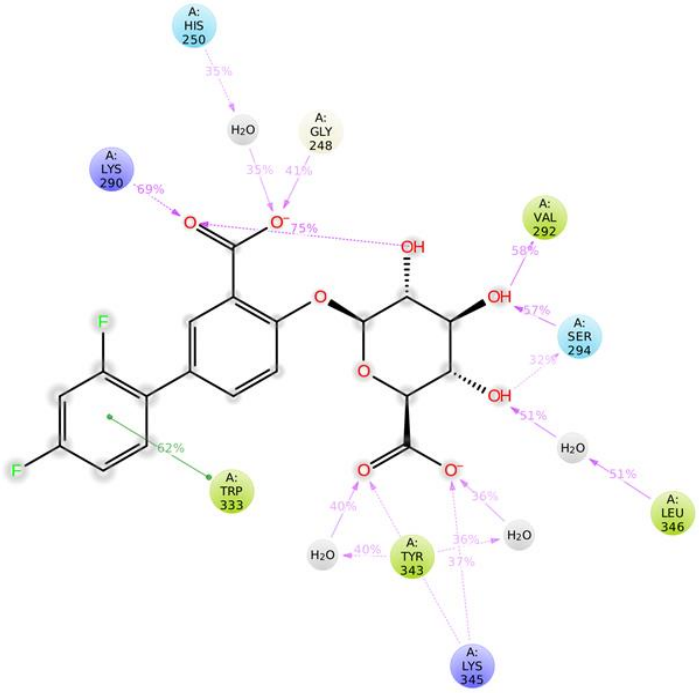
Compound ZINC000044020013 has a positively charged interaction with Lys290 with two oxygens for 37% and 32% of the simulation. Gly248 maintained interaction with Nsp15 for 36% of the simulation. This compound had the least amount of interaction with the binding pocket. This resulted in a less favorable predicted binding energy when compared to the other top compounds.

Figure 13

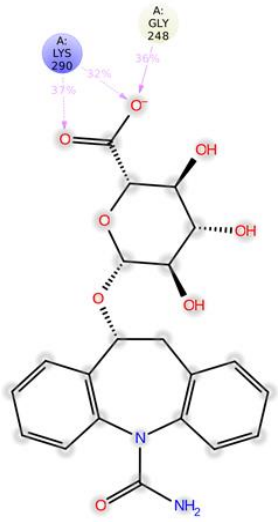
2D Ligand Interaction Diagrams of Top 4 Compounds

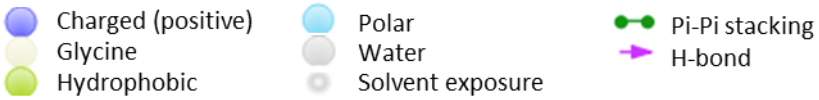


ZINC000033902452



ZINC000044020013





Note. 2D ligand interaction diagrams from the MD trajectory for the top four compounds show the interactions supporting binding to Nsp15. The residues shown interacted with each ligand for a minimum of 30% of the simulation time.

SwissADME Predicts Important Drug Discovery Parameters

Important ADME properties for each compound were predicted using the SwissADME server. When evaluating a compound for use in a drug, we must consider how it will reach its target and assess absorption, distribution, metabolism, and excretion within the body. Properties highlighted in our search were GI absorption, blood brain barrier permeability, Lipinski rule of 5 violations, Cytochrome P450 enzyme inhibition, PAINS alerts, and Brenk alerts.

Gastrointestinal absorption and blood brain barrier permeability is predicted by the Swiss ADME server using the *Brain Or IntestinaL EstimateD permeation* (BOILED-Egg) method. This method measured the lipophilicity (by Log P) and polarity (by polar surface area (PSA) of each compound and sorts them based on their likelihood to be passively absorbed by the gastrointestinal tract or passively diffused through the blood brain barrier. Parameters of the BOILED -Egg considered good gastrointestinal absorption are PSA lower than 142 Å and log *P* between −2.3 and +6.8 (Daina & Zoete, 2016). Parameter considered good for blood brain barrier permeability are a PSA <79 Å and a

lipophilic log *P* from 0.4 to 6.0. SARS-CoV-2 and other coronaviruses are most commonly detected in the upper airways, lungs, mouth, and the gastrointestinal tract so blood brain barrier permeability is not necessary for reaching Nsp15 (Trypsteen et al., 2020). We are most interested in gastrointestinal tract absorption for this study as that is promising for an oral drug which simplifies delivery.

Lipinski's rule of five helps determine if a drug meets a set of standards known to allow drug absorption and permeability. These five rules include a molecular weight less than 500 daltons, a calculated octanol-water partition coefficient (LogP) less than or equal to 5, includes hydrogen bonding acceptors (less than or equivalent to 10), as well as hydrogen bonding donors less than or equivalent to 5 (Benet et al., 2016). Lipinski's rules consider any compound with more than one violation to be a poor candidate for drug use.

The Swiss ADME server also screens compounds for their ability to inhibit top CYP enzymes. Cytochrome P450 enzymes are isoenzymes that play a key role in drug elimination (Daina et al., 2017). Inhibition of the enzymes can lead to drug-drug interactions which may result in toxic or other adverse effects caused by low clearance of the drug (Daina et al., 2017). Understanding the potential for a new drug to inhibit the cytochrome P450 enzymes plays an important role in predicting potential interactions between drugs and drug excretion.

Pan Assay Interference Compounds (PAINS), bind non-specifically to drug targets. These compounds are known to cause false positives in high

throughput virtual screenings. Lack of PAINS alerts suggest a good outlook for these compounds to bind specifically to their target in drug form (Baell & Holloway, 2010).

Brenk alert determines lead-likeness of a molecule and focuses on physiochemical boundaries. Leads may be subject to chemical modifications to optimize the compound which may increase the size of the molecule or its lipophilicity. In anticipation of these modifications, leads should begin smaller in size and less hydrophobic than other molecules.

Results of the Swiss ADME predictions are listed in **Table 4**. Two compounds had good gastrointestinal absorption: ZINC000247434422 and ZINC000043772626 while none displayed the ability to permeate the blood brain barrier. The top four compounds for Nsp15 do not show potential to be inhibitors for four of the five subtypes of cytochrome P450 enzymes (CYPs) including CYP1A2, CYP2C19, CYP2C9, CYP2D6 (McDonnell, PharmD, BCOP & Dang, PharmD, BCPS, 2013). Two compounds (ZINC000247434422 and ZINC000043772626) show potential to inhibit CYP3A4. Compounds had no Lipinski rule of five violations, PAINS or Brenk alerts.

Table 4

The Predicted ADME Properties for the Top 4 Best Compounds by the SwissSimilarity Server are Presented

Compound	GI absorption	BBB Permeant	Lipinski Rule Violations	CYP1A2	CYP2C19	CYP2C9	CYP2D6	CYP3A4	PAINS Alerts	Brenk Alerts
ZINC000247434422	+	-	-	-	-	-	-	+	-	-
ZINC000043772626	+	-	-	-	-	-	-	+	-	-
ZINC000033902452	-	-	-	-	-	-	-	-	-	-
ZINC000096232566	-	-	-	-	-	-	-	-	-	-

Note. Predicted properties include gastrointestinal absorption, blood brain barrier permeability, Cytochrome P450 enzyme inhibition, Lipinski Rule of five violations, PAINS alerts, and Brenk alerts + indicates high GI absorption, BBB permeability, a Lipinski rule violation, inhibition of a cytochrome P450, or a PAINS or Brenk alert. - indicates low GI absorption, no BBB permeability, no Lipinksi Rule Violations, no CYP inhibition, and no PAINS or Brenk alerts.

Based on the SwissADME server screening, all four top compounds remain promising. The two compounds without high gastrointestinal absorption (ZINC000033902452 and ZINC000096232566) would need further consideration for a delivery method as an oral administration may not be suitable. Interactions between CYP3A4 and compounds ZINC000247434422 and ZINC000043772626 should be further investigated. Several effective antiviral drugs are inhibitors of CYP3A4, it is just important to consider toxicity levels of the drug and any other potential drug-drug interactions.

Conclusion

We have presented four promising compounds targeting the Nsp15 protein of SARS-CoV-2. The results of our Molecular docking and MD simulations conclude that each top compound has a favorable predicted binding energy to the Nsp15 active site. The compound performing best through all analyses for Nsp15 was ZINC000247434422 with the most favorable predicted binding energy at -76.5±6.7 kcal/mol. The SwissADME screening allowed us to identify preliminary parameters suggesting that these compounds may succeed in drug form. Each compound remains a candidate for further experimental validation as a Nsp15 small molecule inhibitor.

Chapter 4

Molecular Docking and Molecular Dynamics Simulation of Inhibitors Targeting M_{pro}

Introduction

The M_{pro} of Coronaviruses plays a significant role in processing viral replication making it a potential target for a small molecular inhibitor. In this chapter, I present our high throughput virtual screening of ZINC 15. Eight promising compounds were identified for M_{pro} Molecular dynamics simulation allowed us to predict the binding energies (calculated by MM-GBSA) for these compounds and further understand the interactions supporting binding to their target. Using the SwissADME server we were able to identify structurally based ADME properties that determine the likelihood of our compounds to perform successfully as drugs.

Molecular Docking

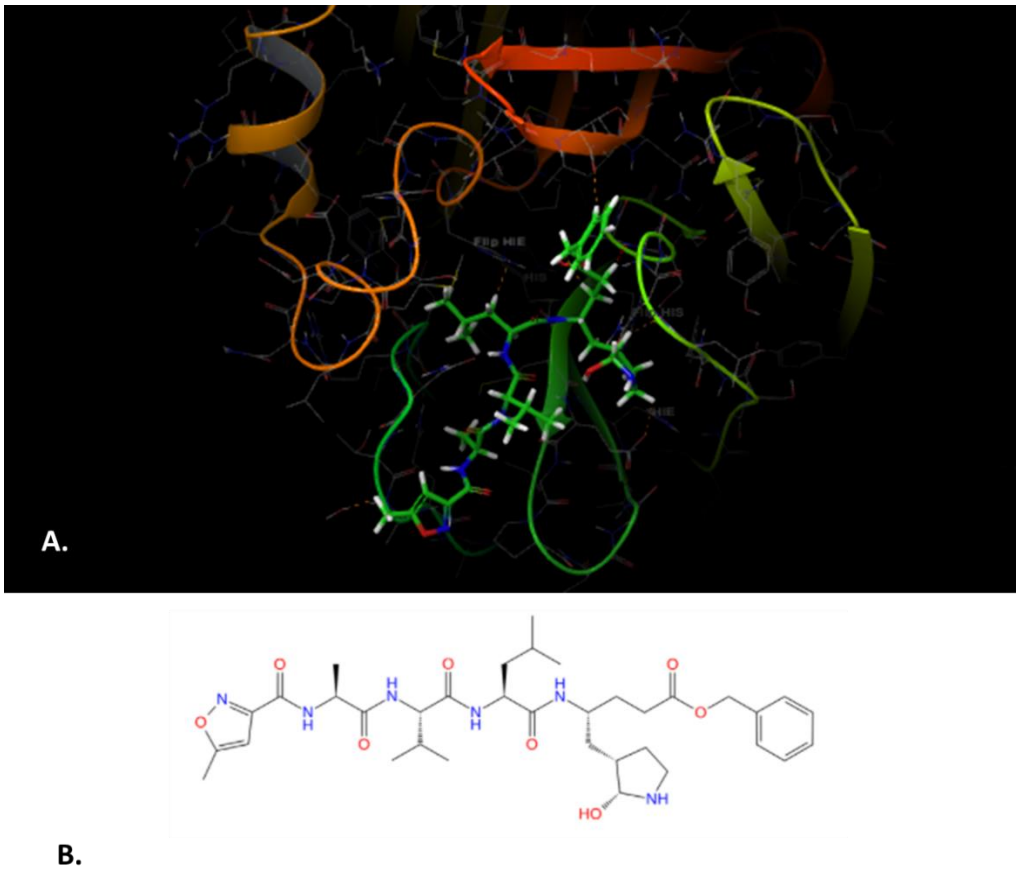
Natural Ligand Docking

Using Glide docking software in Maestro, we measured the binding score of the natural ligand of the M_{pro} (**Figure 14**). The binding score of the natural ligand to the binding site used in our molecular docking and molecular dynamics simulations was calculated at -9.607 kcal/mol. When evaluating molecules for docking, we originally searched for compounds with preliminary docking scores better (more negative) or equivalent to this value. Most compounds in the high throughput virtual screening did not meet these criteria so compounds with a Glide docking score less than -8.4 kcal/mol were

considered. Values with docking scores close to the natural ligand score are acceptable as there is variability in the scoring function.

Figure 14

Mpro in Complex with its Natural Ligand and 2D Structure



Note. A. M_{pro} is shown in complex with its natural ligand (PDB:6LU7). B. The two-dimensional structure of the natural ligand of M_{pro} is presented.

High Throughput Virtual Screening and Molecular Docking

From our high throughput virtual screening of the DrugBank and Zinc15 database, 31 molecules were selected from the Zinc15 database for preliminary molecular docking to M_{pro} using structure PDB:6LU7. The compounds listed in **Table 5** were found to have an initial docking score greater than or equal to -8.4 kcal/mol. After eliminating molecules with high structure similarity, remaining compounds selected by docking score, were further evaluated by induced fit docking. Docking scores selected for further consideration ranged from -8.4 kcal/mol to -10.1 kcal/mol. These compounds were next evaluated by molecular dynamics simulation.

Molecular Dynamics Simulation

Molecular dynamics simulation was performed to evaluate binding of hit compounds to M_{pro}. Output of the MD simulations included docking scores, MM-GBSA free energy, and protein and ligand RMSD. Results of this simulation are presented in **Table 5**.

Table 5

Properties of the Top 31 Compounds from Molecular Docking and MD Simulations are Presented

Zinc ID	Docking Score (kcal/mol)	MM-GBSA (kcal/mol)	Receptor RMSD[1] (Å)	Ligand RMSD[1] (Å)
Natural Ligand	-8.4	-70.08 ± 12.5	1.6±0.1	5.1±1.0
ZINC000004600917	-10.1	-43.9 ± 10.6	2.7±0.1	3.1±0.4
ZINC000057312352	**-9.4**	**-73.0 ± 13.0**	**2.1±0.1**	**4.3±0.9**
ZINC000000090720	-9.4	-51.3 ± 9.0	2.1±0.2	8.8±0.6
ZINC000020988539	-9.3	-54.8 ± 6.4	2.3±0.1	6.3±0.3
ZINC000004899522	-9.1	-55.8 ± 8.4	3.0±0.2	5.9±0.3
ZINC000253630002	-9.1	-62.9 ± 7.9	1.9±0.2	6.0±0.2
ZINC000014728050	-9.1	-54.5 ± 6.9	2.7±0.1	6.6±0.5
ZINC000064568387	-9.0	-68.8 ± 8.4	2.3±0.1	10.4±0.4
ZINC000004897405	-9.0	-62.3 ± 13.6	3.2±0.2	7.7±0.5
ZINC000223270144	-9.0	-61.7 ± 11.4	2.5±0.2	9.7±0.4
ZINC000096447388	-8.9	-68.3 ± 14.1	2.7±0.1	7.2±0.3
ZINC000012119172	-8.8	-39.2 ± 9.4	2.6±0.1	17.8±0.5
ZINC000000121038	-8.8	-51.6 ± 9.0	2.0±0.1	5.4±0.4
ZINC000070216736	-8.8	-40.3 ± 8.2	6.7±0.4	8.3±0.6
ZINC000663523562	**-8.8**	**-72.6 ± 12.1**	**2.7±0.2**	**9.3±1.5**
ZINC000005273576	-8.7	-44.5 ± 14.5	3.0±0.1	8.3±2.5
ZINC000064568512	**-8.7**	**-83.4 ± 8.5**	**3.1±0.1**	**2.4±0.2**
ZINC000000632530	-8.7	-59.7 ± 13.9	2.8±0.3	12.2±0.6
ZINC000261493176	-8.7	-33.6 ± 15.7	2.7±0.1	35.8±2.7
ZINC000426359607	-8.7	-69.0 ± 13.2	2.7±0.2	8.7±0.3
ZINC000263585674	-8.6	-69.1 ± 10.7	2.3±0.1	9.2±0.3
ZINC000057774900	-8.6	-56.5 ± 12.6	3.1±0.2	10.5±0.6

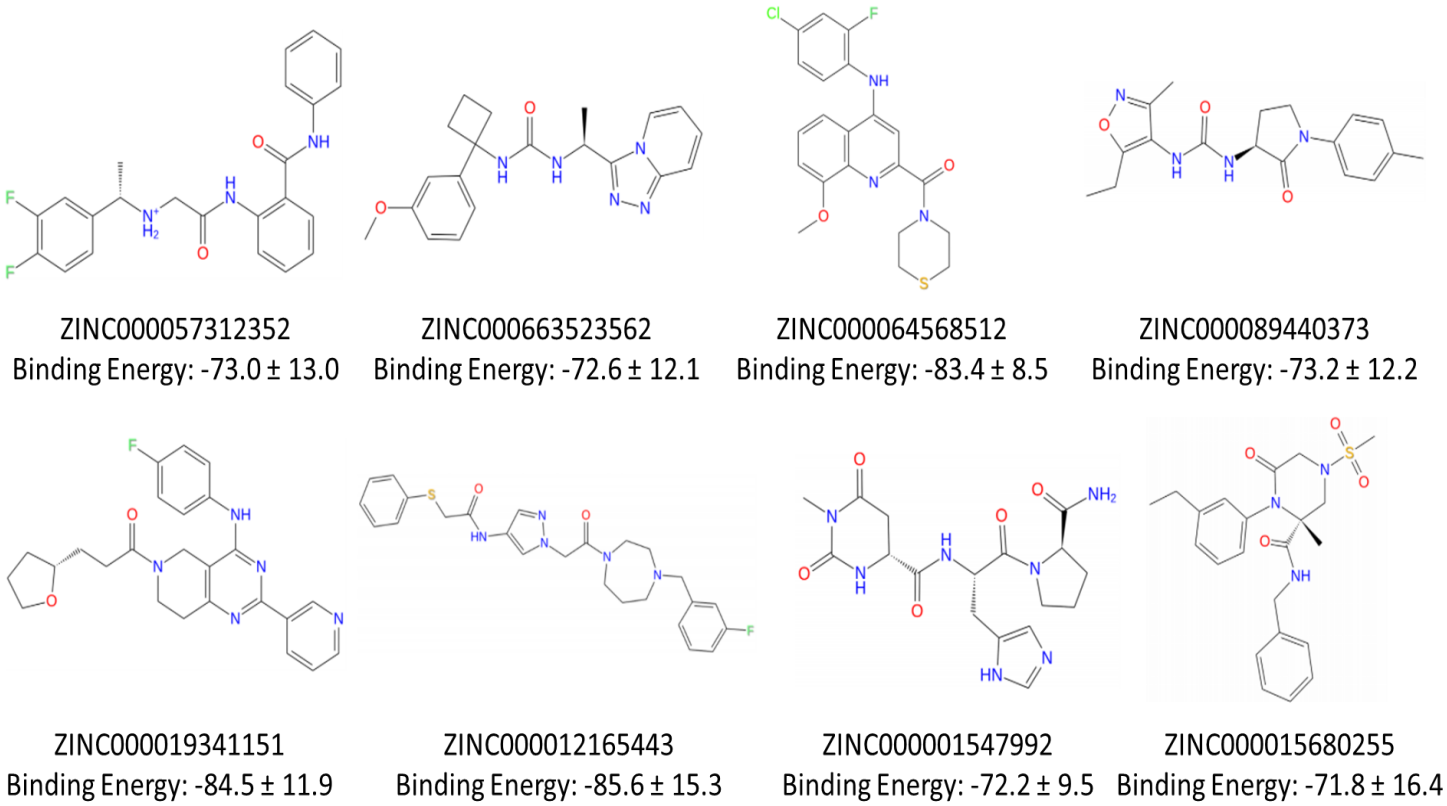
Zinc ID	Docking Score (kcal/mol)	MM-GBSA (kcal/mol)	Receptor RMSD[1] (Å)	Ligand RMSD[1] (Å)
ZINC000089440373	**-8.6**	**-73.2 ± 12.2**	**2.3±0.1**	**9.4±0.6**
ZINC000019341151	**-8.6**	**-84.5 ± 11.9**	**2.2±0.2**	**3.1±0.4**
ZINC000193716208	-8.6	-56.7 ± 21.2	3.2±0.2	24.4±4.3
ZINC000008876585	-8.5	-63.1 ± 7.7	2.9±0.1	4.8±0.4
ZINC000012165443	**-8.5**	**-85.6 ± 15.3**	**2.7±0.1**	**10.2±0.8**
ZINC000012990014	-8.5	-43.0 ± 9.4	2.6±0.2	3.0±0.3
ZINC000001547992	**-8.5**	**-72.2 ± 9.5**	**2.5±0.2**	**2.9±0.3**
ZINC000015680255	**-8.4**	**-71.8 ± 16.4**	**2.1±0.1**	**6.5±0.5**
ZINC000005553602	-8.4	-57.8 ± 12.3	2.1±0.1	7.4±0.6

Note. Properties including the docking score, total binding energy as calculated by MM-GBSA, and the RMSD for the ligand and protein. [1] Receptor and Ligand RMSD values are averaged from the last 20 ns of the simulation. Top promising compounds are shown in bold

Analysis of the output from the MD simulation showed eight promising compounds: ZINC000057312352, ZINC000663523562, ZINC000064568512, ZINC000089440373, ZINC00019341151, ZINC000012165443, ZINC000001547992, and ZINC000015680255. These compounds and their docking scores are presented in **Figure 15**. Each compound had a docking score higher than -8.4 kcal/mol, a low predicted binding energy, and remained bound to the protein after the trajectory stabilized.

Figure 15

Top 8 Compounds and their Predicted Binding Energy



Note. Top eight compounds for Nsp15 inhibition were determined by docking score, predicted binding energy, and the ability to remain bound to the protein when the trajectory stabilized. Predicted binding energies according to MM-GBSA calculations are listed for each compound.

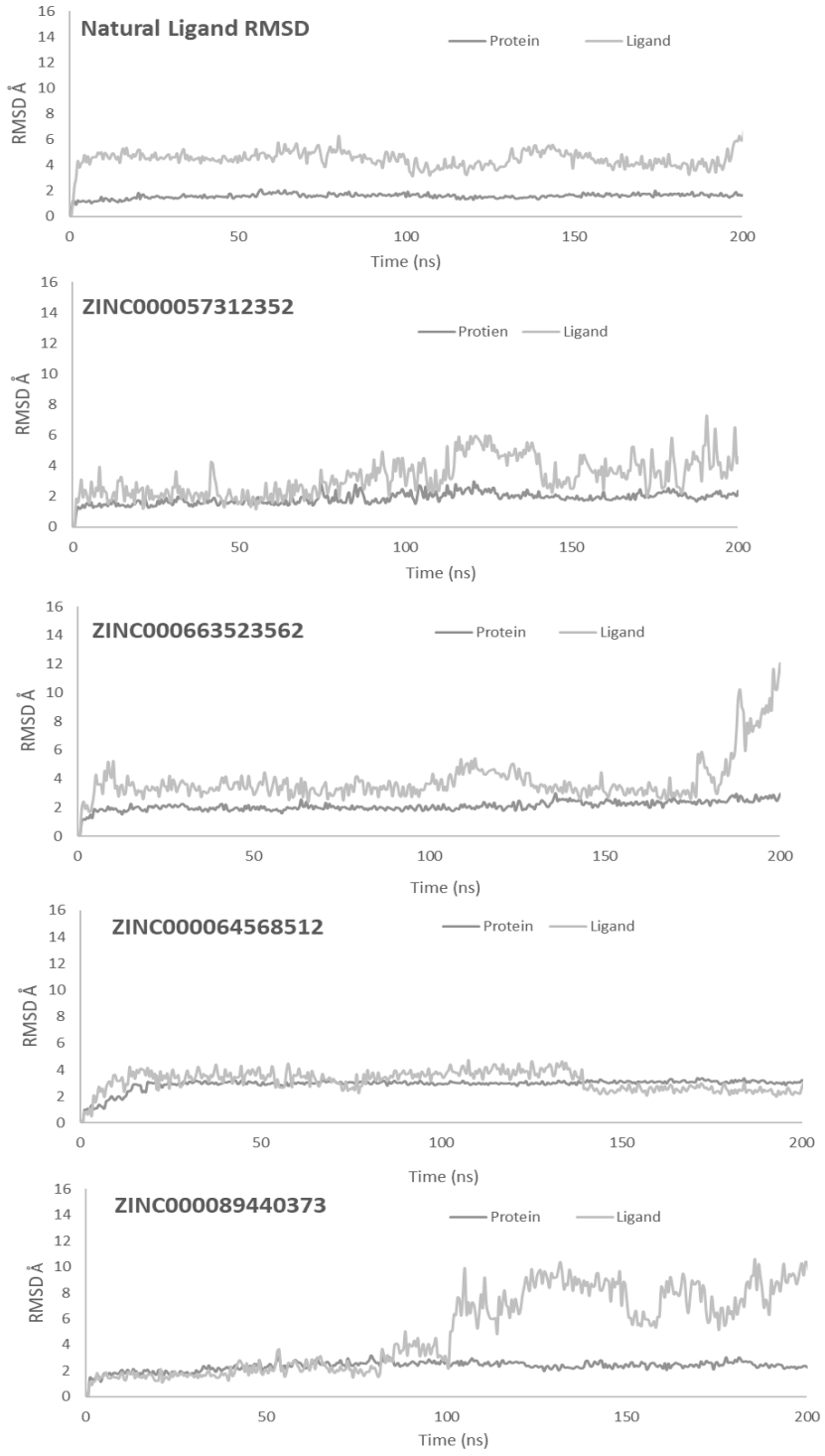
RMSD Plots Show the Simulation has Stabilized

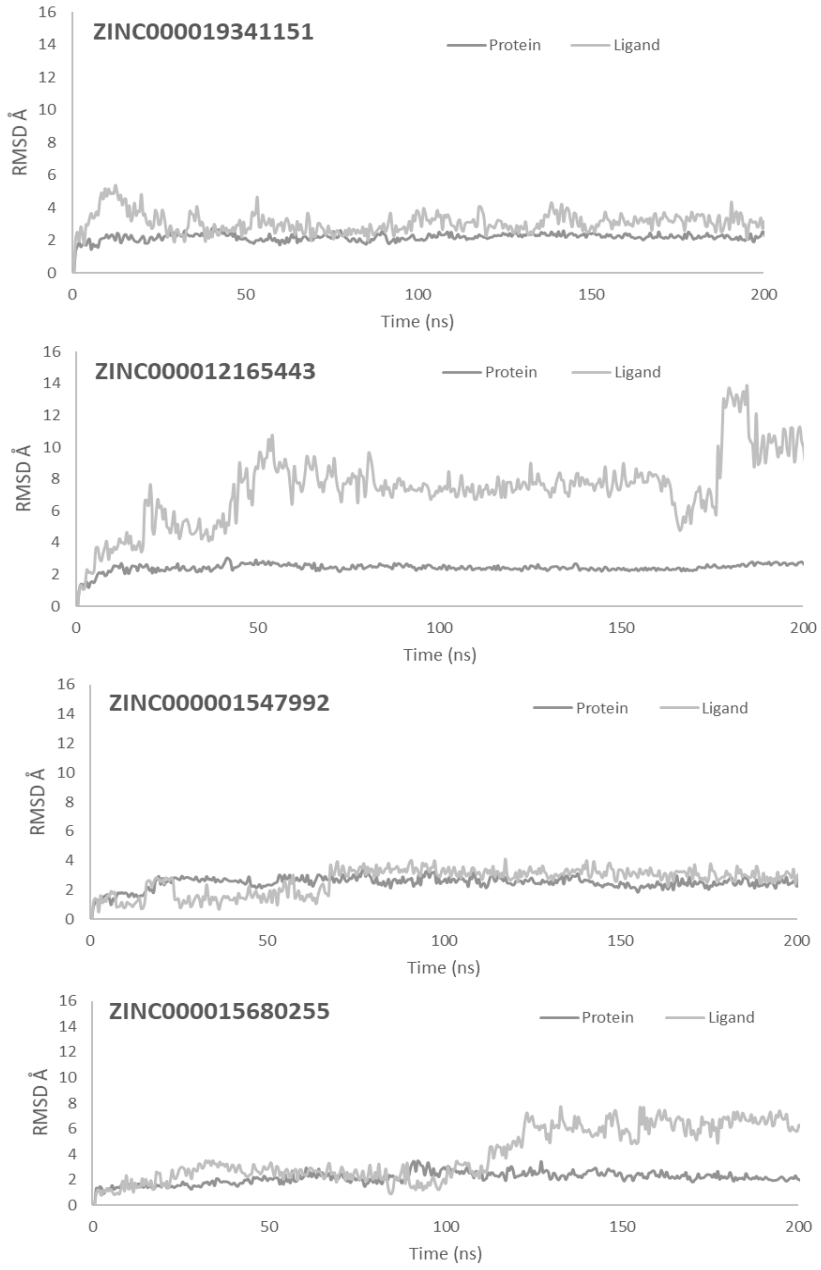
Binding stability was evaluated by measuring the RMSD values of each complex throughout the entire simulation. The values listed for the RMSD of the protein and ligand in **Table 5** are averaged over the last 20 ns. As previously described in chapter 3, we expect the RMSD values to converge around toward the end of the simulation to confirm equilibration. Plots of RMSD values throughout the entire simulation for the eight most promising compounds are presented in **Figure 16**. Small fluctuations of the RMSD values (between 1-3 Å) are expected throughout the simulation. Larger fluctuations suggest a larger conformational change for the protein or may indicate the ligand has moved away from the binding site.

The RMSD values of each complex throughout the simulation is presented in **figure 16**. For all simulations, the protein RMSD remained stable. The RMSD for compounds ZINC000057312352, ZINC000064568512, ZINC00019341151, ZINC000001547992, and ZINC000015680255 remained stable throughout the simulation which suggests the ligand is stable within the binding pocket of the protein. Compounds ZINC000663523562, ZINC000089440373, and ZINC000012165443 saw a large change in the ligand RMSD value towards the end of the trajectory. A longer simulation may allow us to see if stabilization occurs.

Figure 16

RMSD of Mpro and Top 8 Compounds





Note. Root Mean Square Deviation (RMSD) plots are presented for the MD simulation runs of each top protein-ligand complex over the length of the trajectory. The Cα-RMSD for the protein is shown in dark grey and the ligand RMSD is shown in light grey. Cα-RMSD is based on initial protein alignment.

MM-GBSA Binding Energy Components

The MM-GBSA predicted binding energy calculation combines energies contributed from the Van der waals, hydrophobic, and electrostatic terms. The predicted binding energy from each of these terms is presented in **Table 6**. These terms are rewarded by the Glide docking score scoring function, so it is reasonable to expect that compounds with a favorable (more negative) docking score will also have a favorable predicted binding energy. A breakdown of these terms allows us to further understand how the amino acids within the binding pocket interact with the ligand.

Ligand interactions with the M_{pro} binding pocket were supported heavily by the Van der Waals term. As previously discussed, the Van der Waals term is the sum of energy contribution made by Van der Waals interactions, pi-pi stacking, and self-contact energies. The hydrophobic term is measured alone while the electrostatic term is the sum of energy contributions made by Coulomb interactions, hydrogen bonds, and general Born solvation energies. The Van der Waals and hydrophobic terms of the predicted binding energy equation are the largest contributors for all top compounds. Electrostatic contributions were minimal and mostly unfavorable for this target.

Table 6

The MM-GBSA Binding Energy Calculation Components Including the Van der Waals, Hydrophobic, and Electrostatic Terms are Presented

Zinc ID	Van der Waals (kcal/mol)	Hydrophobic (kcal/mol)	Electrostatic (kcal/mol)
Natural Ligand	-49.0 ± 5.2	-31.1 ± 5.2	10.2 ± 4.8
ZINC000004600917	-27.7 ± 4.6	-16.7 ± 2.8	0.5 ± 6.3
ZINC000057312352	**-46.5 ± 7.1**	**-31.4 ± 5.5**	**4.9 ± 5.8**
ZINC000000090720	-32.0 ± 4.1	-17.9 ± 2.2	-1.5 ± 6.5
ZINC000020988539	-36.4 ± 5.9	-24.5 ± 3.6	6.1 ± 4.6
ZINC000004899522	-33.8 ± 4.9	-24.5 ± 3.4	2.5 ± 5.8
ZINC000253630002	-42.7 ± 4.6	-27.6 ± 3.0	7.3 ± 3.3
ZINC000014728050	-31.7 ± 2.6	-21.1 ± 2.4	-1.8 ± 5.8
ZINC000064568387	-41.9 ± 5.0	-35.2 ± 5.3	8.4 ± 3.2
ZINC000004897405	-43.1 ± 5.4	-27.2 ± 4.9	7.9 ± 8.0
ZINC000223270144	-38.1 ± 3.8	-23.9 ± 3.6	0.3 ± 6.5
ZINC000096447388	-39.2 ± 6.7	-30.1 ± 7.5	0.9 ± 5.2
ZINC000012119172	-31.7 ± 6.2	-17.5 ± 4.8	10.0 ± 2.8
ZINC000000121038	-33.3 ± 5.4	-25.1 ± 4.0	6.7 ± 3.1
ZINC000070216736	-26.6 ± 5.5	-16.7 ± 2.8	3.0 ± 3.4
ZINC000663523562	**-42.0 ± 6.3**	**-34.5 ± 4.5**	**4.0 ± 5.1**
ZINC000005273576	-29.1 ± 9.6	-17.2 ± 6.2	1.9 ± 7.8

Zinc ID	Van der Waals (kcal/mol)	Hydrophobic (kcal/mol)	Electrostatic (kcal/mol)
ZINC000064568512	**-48.9 ± 4.9**	**-44.0 ± 3.8**	**9.5 ± 4.4**
ZINC000000632530	-35.0 ± 8.0	-31.0 ± 6.5	6.3 ± 2.3
ZINC000261493176	-21.8 ± 9.4	-16.4 ± 8.1	4.6 ± 3.2
ZINC000426359607	-40.4 ± 6.8	-33.9 ± 7.1	5.4 ± 4.4
ZINC000263585674	-44.2 ± 5.9	-31.3 ± 4.8	6.4 ± 4.5
ZINC000057774900	-35.7 ± 6.7	-25.1 ± 5.0	4.2 ± 5.1
ZINC000089440373	**-41.7 ± 6.8**	**-30.0 ± 5.3**	**-1.6 ± 6.0**
ZINC000019341151	**-51.8 ± 6.1**	**-36.4 ± 5.3**	**3.6 ± 5.1**
ZINC000193716208	-28.8 ± 9.2	-24.5 ± 9.6	-3.4 ± 5.6
ZINC000008876585	-44.2 ± 5.0	-25.3 ± 4.1	6.4 ± 4.6
ZINC000012165443	**-51.6 ± 8.2**	**-39.2 ± 6.9**	**5.2 ± 3.0**
ZINC000012990014	-31.3 ± 4.0	-17.3 ± 4.0	5.7 ± 4.7
ZINC000001547992	**-51.3 ± 7.6**	**-24.0 ± 3.1**	**3.0 ± 4.7**
ZINC000015680255	**-45.0 ± 8.0**	**-27.4 ± 8.7**	**0.6 ± 4.5**
ZINC000005553602	-36.1 ± 5.8	-26.2 ± 7.3	4.5 ± 3.3

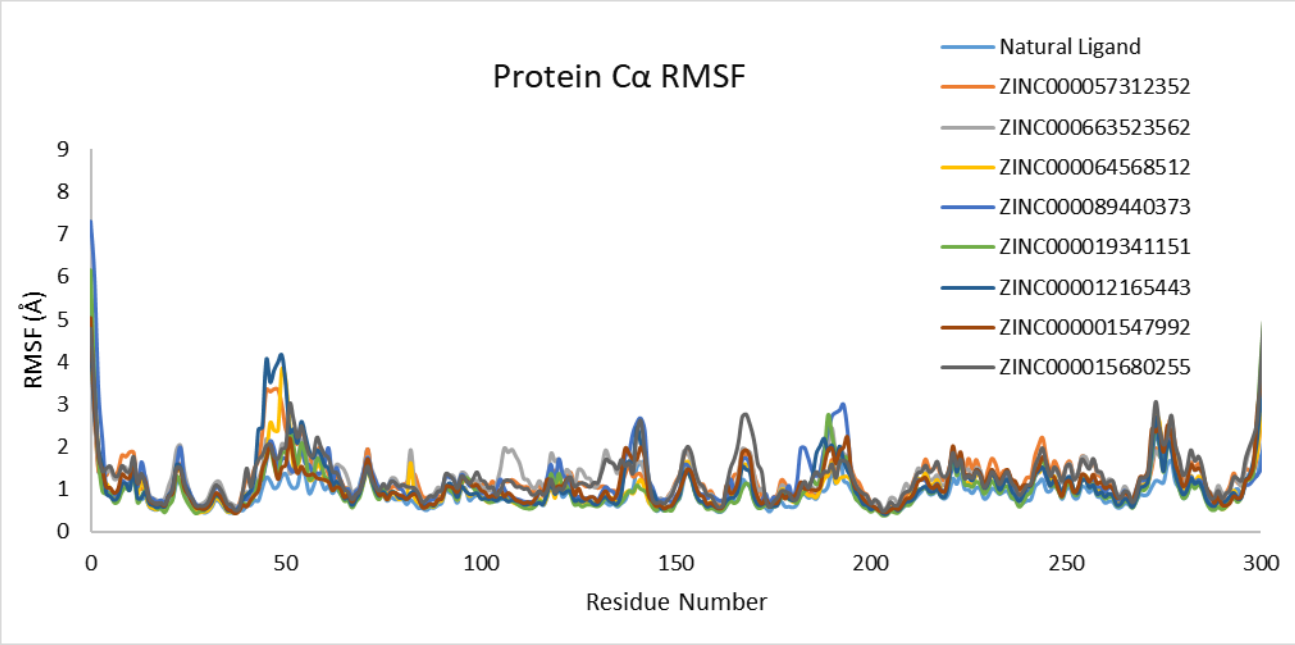
Note. Top promising compounds are shown in bold.

The Protein Cα Root Mean Square Fluctuation Confirms the Stability of Compound Binding.

Residue fluctuation was measured throughout the simulation and is presented by RMSF values in F**igure 17**. Higher RMSF values indicate more fluctuation of amino acids throughout the MD simulation. We see the largest fluctuations at the N and C terminals which is expected. Residues fluctuated similarly for each compound and the natural ligand with larger fluctuations found in loops between secondary structures which was observed in residues 40-53, 140-148, 152-155, 165-170, 180-200, and 275-290. An amino acid sequence of the M_{pro} including secondary structure assignments is presented as **Appendix Figure 2** which allows us to identify these loop regions**.** Key amino acids interacting within the binding pocket including Ser 46, leu141, Asn 142, Glu 166, Pro168, Ala191, Thr190, and Gln189 appear stable with RMSF values under 3 Å.

Figure 17

Residue Fluctuation Plotted by RMSF



Note. Residue fluctuation plotted by RMSF (Cyan: Natural Ligand, Orange: ZINC000057312352, Light Grey: ZINC000663523562, Yellow: ZINC000064568512, Light Blue: ZINC000089440373, Green: ZINC00019341151 Dark Blue: ZINC000012165443 Red: ZINC000001547992 Dark Grey: ZINC000015680255)

Simulation Interaction Diagrams Reveal Key Binding Residues for Top Compounds

Key residues supporting compound binding to Nsp15 for at least 30 % of the simulation were identified by the Desmond simulation interaction diagram (**Figure 18**). These interactions are expected to agree with docking score and MM-GBSA binding energy prediction value. As shown in **Table 6**, most of the predicted binding energy is composed of the Van der Waals and hydrophobic terms. Most of the interactions displayed in the simulation diagrams for M_{pro} were driven by hydrophobic, polar, and charged interactions. This agrees with known characteristics of residues found within the binding pocket including: Cys145, His4, Ser 46, leu141, Asn 142, Glu 166, Pro168, Ala191, Thr190, and Gln189

Interactions supporting ZINC000057312352 binding include hydrophobic interactions with Met49 and Met165. Positive charge interactions included His41 for 34% of the simulation as well as Gln189 for 35% of the simulation. This compound had the most favorable predicted binding energy, with most of its energy contributions coming from the Van der Waals term.

ZINC000663523562 has a single positive charge interaction with Gln189 for 34% of the simulation. ZINC000064568512 has a negatively charged interaction with Asp187 for 30% while most of the compound remains solvent exposed. These compounds also received most of their predicted binding energy contribution from the Van der Waals term.

Compound ZINC000089440373 interacts with the M_{pro} via positively charged interactions with His41 (37%) and Thr45 (42%) and negative charged interactions with

Glu166 through an interaction with a water molecule (39%) and Asp187 (35%). ZINC000019341151 interacts with Gly143 96 % of the simulation. Positively charged interactions supporting binding include with Ser144 (35%) and His164 (53%). A single negatively charged interaction with Glu166 through an interaction with a water molecule takes place for 38% of the simulation.

Binding of ZINC000012165443 is supported by two hydrophobic interactions with Phe185 and Val186 for 58% of the simulation. Other residues involved in binding include a polar interaction with His164 (34%) and a positive charged interaction with Arg40 (44%). Half of this compound remains solvent exposed.

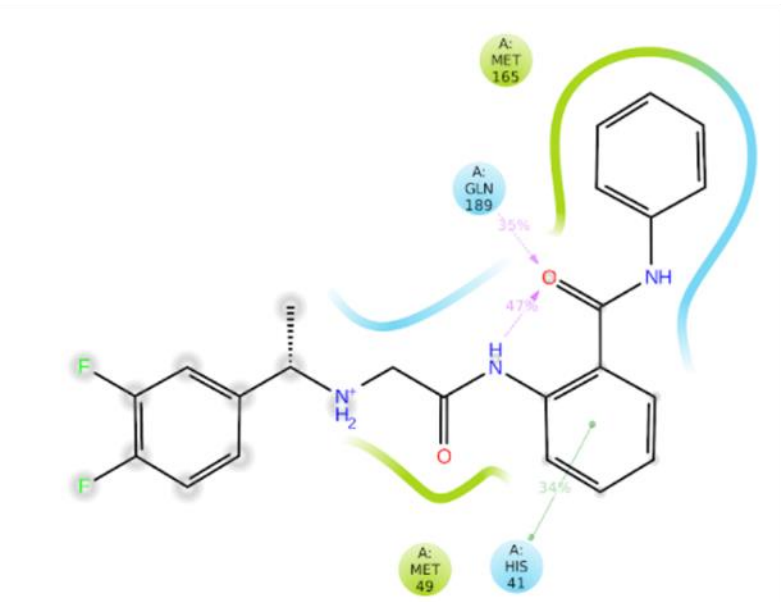
ZINC000001547992 binding energy is mostly supported by Van der Waals interactions. This compound has two negatively charged interactions with Asp187 and Glu166. Asp187 interacts with the ligand directly for 50% of the simulation and through a water bridge for 32% of the simulation. Glu166 interacts with the ligand directly for 82% of the simulation and through an interaction with water for 45% of the simulation. The compound also has two polar charged interactions with His41 (30%) and His164 (33%).

ZINC000015680255 has one hydrophobic interaction with Cys145 for 40% of the simulation. Three polar interactions with the protein include His41 in two separate areas of the compound for 52% and 47% of the simulation, Ser144 for 40% of the simulation, and Gln189 for 37% of the simulation. There is a single negatively charged interaction with Glu166 for 98% of the simulation. The binding energy for this compound is also supported heavily by Van der Waals interactions.

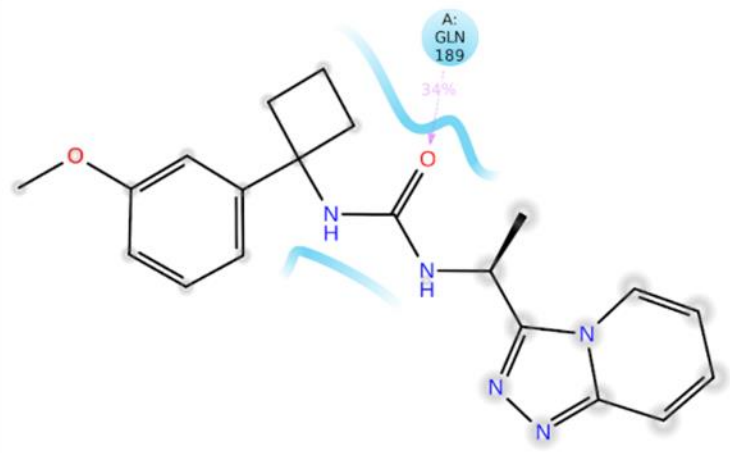
Figure 18

2D Ligand Interaction Diagrams of Top 8 Compounds

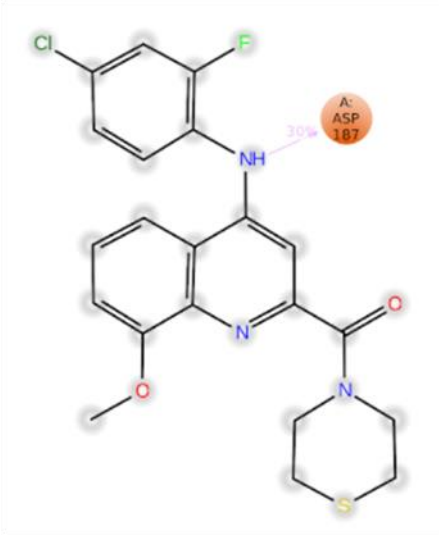
ZINC000057312352



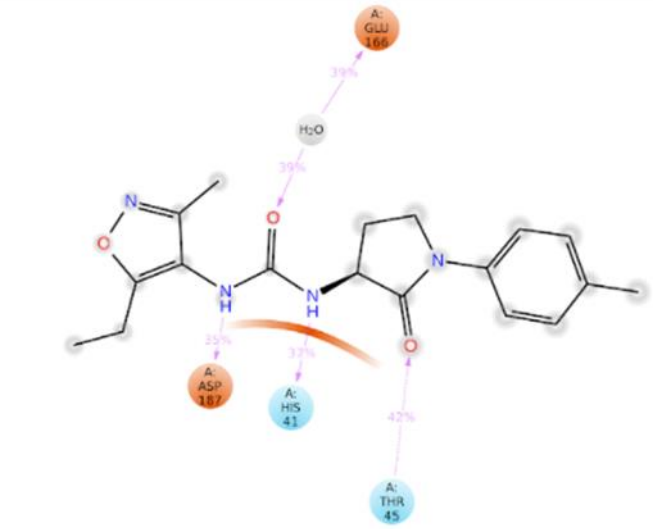
ZINC000663523562



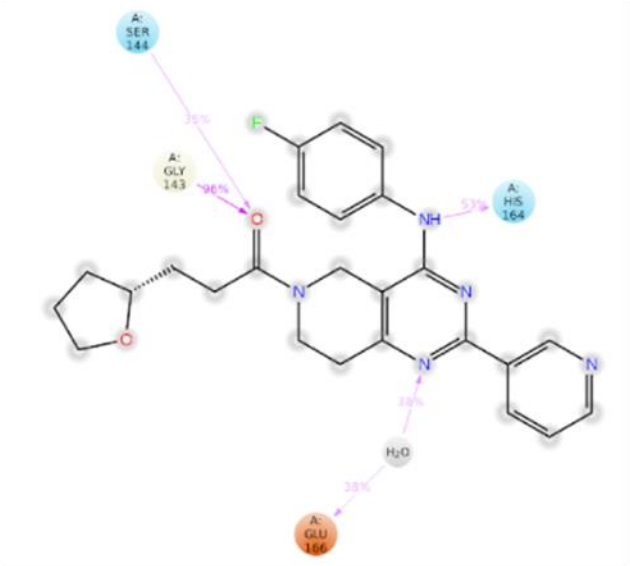
ZINC000064568512



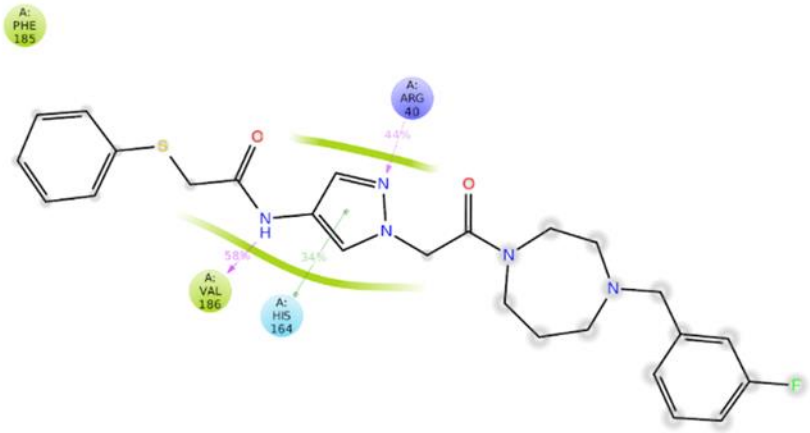
ZINC000089440373



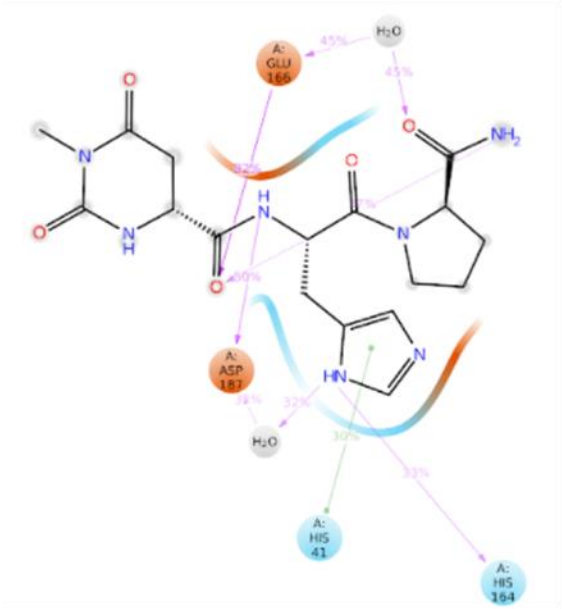
ZINC000019341151



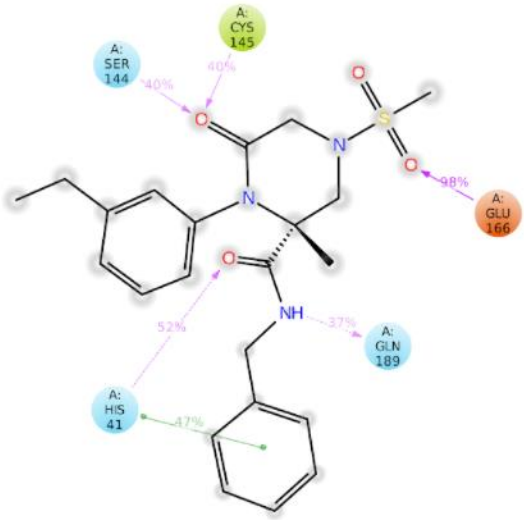
ZINC000012165443

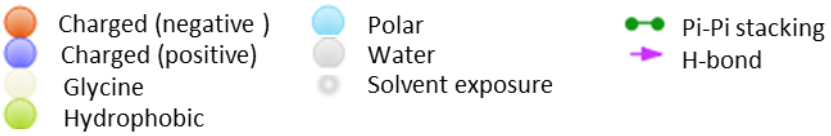


ZINC000001547992



ZINC000015680255





Note. 2D ligand interaction diagrams from the MD trajectory for top eight compounds show the interactions supporting binding of top compounds. Residues displayed interacted with each ligand for a minimum of 30% of the simulation time.

SwissADME Predicts Important Drug Discovery Parameters

Predicted ADME properties including gastrointestinal absorption, blood brain barrier permeability, Lipinski rule of 5 violations, inhibition of five cytochrome P450 enzymes, PAINS alerts, and Brenk alerts for the top eight compounds are presented in **Table 7**. A full description of each of these properties was presented previously in Chapter 3.

Table 7

The Predicted ADME Properties for the Top 8 Best Compounds by the SwissSimilarity Server are Presented

Note. Predicted properties include gastrointestinal absorption, blood brain barrier permeability, Cytochrome P450 enzyme inhibition, Lipinski Rule of five violations, PAINS alerts, and Brenk alerts. + indicates high GI absorption, BBB permeability, a Lipinski rule violation, inhibition of a cytochrome P450, a PAINS alert, or a Brenk alert. - indicates low GI absorption, no BBB permeability, no Lipinksi Rule Violations, no CYP inhibition, and no PAINS or Brenk alerts.

Compound	GI absorption	BBB Permeant	Lipinski Rule Violations	CYP1A2	CYP2C19	CYP2C9	CYP2D6	CYP3A4	PAINS Alerts	Brenk Alerts
ZINC000057312352	+	+	-	-	+	-	+	-	-	-
ZINC000663523562	+	-	-	-	+	+	+	+	-	-
ZINC000064568512	+	-	-	+	+	+	+	+	-	-
ZINC000089440373	+	-	-	-	+	+	-	+	-	-
ZINC000019341151	+	-	-	+	+	+	+	+	-	-
ZINC000012165443	+	-	-	-	+	+	+	+	-	-
ZINC000001547992	-	-	+	-	-	-	-	-	-	-
ZINC000015680255	+	-	-	-	+	+	+	+	-	-

All compounds show high intestinal absorption properties apart from ZINC000001547992. This suggests most compounds would be suitable for oral delivery. Compound ZINC000057312352 was the only compound that showed potential for blood brain barrier permeability. As previously discussed, SARS-CoV-2 and other coronaviruses are more commonly found in the upper airways, lungs, mouth, and the gastrointestinal tract so blood brain barrier permeability is not necessary.

Compounds targeting the M_{pro} show varying potential to be inhibitors for the five subtypes of cytochrome P450 enzymes (CYPs) including CYP1A2, CYP2C19, CYP2C9, CYP2D6, and CYP3A4. Two compounds, ZINC000064568512 and ZINC000019341151, show potential to inhibit all five of the CYPs in this analysis. It is likely these compounds would have trouble being metabolized and would likely interact with several other drugs which raises concern. Compounds ZINC000663523562, ZINC000012165443, and ZINC000015680255 inhibit four out of the five CYP's presented which is also concerning.

The top compounds presented did not have any Lipinski rule of five violations apart from ZINC000001547992 which has one violation. This compound has more than ten hydrogen bond acceptors. One Lipinski rule violation does not entirely rule out the potential for this compound to work as a drug. There were no PAINS or Brenk alerts for any of the compounds.

In summary, compounds that show the ability to inhibit several CYP enzymes including ZINC000064568512, ZINC000019341151, ZINC000663523562, ZINC000012165443, and ZINC000015680255 require further examination to understand if they can be

metabolized in the body. These compounds should also be evaluated for potential drug-drug interactions. Two compounds, ZINC000057312352 and ZINC000001547992, look the most promising in their ability to work in drug form.

Conclusion

We have presented eight promising compounds targeting the M_{pro} of SARS-CoV-2. The results of our Molecular docking and MD simulations conclude that each top compound has a favorable predicted binding energy to the M_{pro} active site. SwissADME screening allowed us to identify any preliminary signs that these compounds may not succeed in drug form. Two compounds: ZINC000057312352 and ZINC000001547992 look the most promising with favorable binding energies and good predicted ADME properties. RMSD and RMSF values support stability of these compounds interacting with the M_{pro} binding pocket.

Chapter 5

Conclusion and Future Directions

Conclusion

Since March of 2020 when SARS-CoV-2 was declared a pandemic, we have experienced the effects of this deadly virus across the globe. The development of vaccines has allowed us to return to some sense of normalcy, but the threat of a vaccine-resistant variant, or another coronavirus remains. A small molecule inhibitor targeting conserved proteins of coronaviruses may improve the outlook of public health now and in the future. Our high throughput virtual screening resulted in several promising compounds for the conserved Nsp15 and M_{pro} protein targets of coronaviruses.

The compound performing best through all the analyses for Nsp15 was ZINC000247434422 with the most favorable predicted binding energy at -76.5±6.7 kcal/mol, stability throughout the simulation, and a good outlook on important ADME properties. The other 3 top compounds also performed well overall and should continue to be considered for experimental analysis.

The M_{pro} screening resulted in 8 top compounds for further consideration as inhibitors. Two compounds, ZINC000057312352 and ZINC000001547992, look the most promising with favorable binding energies and predicted ability to work in drug form. RMSD and RMSF values support stability of these compounds interacting with the M_{pro} binding pocket. Further assessment is needed to determine whether the other six promising compounds have potential to work as a small molecule inhibitor for this target. While their predicted binding energy remains favorable, it is important that the

compounds remain stable in complex with the target and structurally agree with ADME property parameters to be used as a drug.

While our results have narrowed down a large search of compounds targeting Nsp15 and M_{pro}, further experimental analysis is needed to fully characterize binding of these compounds to their targets. The entropic terms that were unaccounted for in our analysis should also be considered in future work.

Future Directions

As previously mentioned, the computational methods used for these projects neglected the entropic terms of binding energy. To address this, I would consider using additional computational tools available to estimate the entropic properties of compound binding. The configurational entropy component of binding can be predicted by normal mode analysis (NMA) (Forouzesh & Mishra, 2021). NMA calculates vibrational modes as well as protein flexibility reliably, but at a large computational cost (Alexandrov, 2005). To overcome these steep computational costs, many groups use a truncated version of their protein target. Using this method with MM-GBSA calculations would offer a more complete prediction of which compounds bind most favorably.

Statistical approaches for analyzing MD simulations vary but would be useful to verify these methods in future work. Clustering analyses and principal component analyses (PCA) are commonly applied. Clustering analyses would allow us to group similar molecular configurations found in the simulation, into groups in an unbiased manner (Shao et al., 2007). This method of separation would minimize variance in our predicted free energy calculations. A principal component analysis of each MD

simulation could offer insight into which conformational changes in the simulation are relevant versus those that are simulation-based fluctuations which are unlikely to be repeated. Repeating MD simulations is unpractical due to time and cost therefore these analyses are important for identifying outliers in the large amount of data produced by each simulation.

In addition to considering other computational tools, compounds should be verified experimentally. Both Nsp15 and M_{pro} plasmids are commercially available for recombinant protein production (Altincekic et al., 2021). Each compound presented in this thesis is also available commercially. Nuclear magnetic resonance (NMR) spectroscopy would be a powerful tool in quantitatively measuring both binding affinity and the conformational changes that occur during binding. Chemical shift assignments are available for both the Nsp15 and M_{pro} proteins at https://covid19-nmr.de/ making these experiments even more accessible.

References

Shereen, M. A., Khan, S., Kazmi, A., Bashir, N., & Siddique, R. (2020). COVID-19 infection: Emergence, transmission, and characteristics of human coronaviruses. *Journal of Advanced Research*, *24*, 91–98. https://doi.org/10.1016/j.jare.2020.03.005

Latinne, A., Hu, B., Olival, K. J., Zhu, G., Zhang, L., Li, H., Chmura, A. A., Field, H. E., Zambrana-Torrelio, C., Epstein, J. H., Li, B., Zhang, W., Wang, L. F., Shi, Z. L., & Daszak, P. (2020). Origin and cross-species transmission of bat coronaviruses in China. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-17687-3

Forni, D., Cagliani, R., Clerici, M., & Sironi, M. (2017b). Molecular Evolution of Human Coronavirus Genomes. *Trends in Microbiology*, *25*(1), 35–48. https://doi.org/10.1016/j.tim.2016.09.001

Woolhouse, M., Scott, F., Hudson, Z., Howey, R., & Chase-Topping, M. (2012). Human viruses: discovery and emergence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1604), 2864–2871. https://doi.org/10.1098/rstb.2011.0354

Liu, J., Xie, W., Wang, Y., Xiong, Y., Chen, S., Han, J., & Wu, Q. (2020). A comparative overview of COVID-19, MERS and SARS: Review article. *International Journal of Surgery*, *81*, 1–8. https://doi.org/10.1016/j.ijsu.2020.07.032

Santacroce, L., Charitos, I. A., Carretta, D. M., De Nitto, E., & Lovero, R. (2020). The human coronaviruses (HCoVs) and the molecular mechanisms of SARS-CoV-2 infection. *Journal of Molecular Medicine*, *99*(1), 93–106. https://doi.org/10.1007/s00109-020-02012-8

Sanche, S., Lin, Y. T., Xu, C., Romero-Severson, E., Hengartner, N., & Ke, R. (2020). High Contagiousness and Rapid Spread of Severe Acute Respiratory Syndrome Coronavirus 2. *Emerging Infectious Diseases*, *26*(7), 1470–1477. https://doi.org/10.3201/eid2607.200282

Alexandersen, S., Chamings, A., & Bhatta, T. R. (2020). SARS-CoV-2 genomic and subgenomic RNAs in diagnostic samples are not an indicator of active replication. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-19883-7

Sanjuán, R., & Domingo-Calap, P. (2016). Mechanisms of viral mutation. *Cellular and Molecular Life Sciences*, *73*(23), 4433–4448. https://doi.org/10.1007/s00018-016-2299-6

Robson, F., Khan, K. S., Le, T. K., Paris, C., Demirbag, S., Barfuss, P., Rocchi, P., & Ng, W. L. (2020). Coronavirus RNA Proofreading: Molecular Basis and Therapeutic Targeting. *Molecular Cell*, *79*(5), 710–727. https://doi.org/10.1016/j.molcel.2020.07.027

Sanders, A., Ricci, S., Uribe, S., Boyle, B., Nepper, B., & Nucci, N. (2021). A Survey of Inhibitors for the Main Protease of Coronaviruses with the Potential for Development of Broad-Spectrum Therapeutics. *American Journal of Undergraduate Research*, *17*(4), 71–84. https://doi.org/10.33697/ajur.2020.037

Astuti, I., & Ysrafil. (2020). Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2): An overview of viral structure and host response. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, *14*(4), 407–412. https://doi.org/10.1016/j.dsx.2020.04.020

Saxena, S. K., Kumar, S., Baxi, P., Srivastava, N., Puri, B., & Ratho, R. K. (2020). Chasing COVID-19 through SARS-CoV-2 spike glycoprotein. *VirusDisease*, *31*(4), 399–407. https://doi.org/10.1007/s13337-020-00642-7

Figure 2 Reprinted by permission from [**Rightslink**]: [**Springer**] [Chasing COVID-19 through SARS-CoV-2 spike glycoprotein, Saxena, S. K., Kumar, S., Baxi, P., Srivastava, N., Puri, B., & Ratho, R. K.), [**2020**]

Huang, Y., Yang, C., Xu, X. F., Xu, W., & Liu, S. W. (2020). Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19. *Acta Pharmacologica Sinica*, *41*(9), 1141–1149. https://doi.org/10.1038/s41401-020-0485-4

Gasmalbari, E., & Abbadi, O. S. (2020). Infectious Disease and Tropical Medicine. *Non-Structural Proteins of SARS-CoV-2 as Potential Sources for Vaccine Synthesis*, *6*(667), 1–7. https://www.infectiousjournal.com/wp-content/uploads/sites/6/2020/10/e667.pdf

V'kovski, P., Kratzel, A., Steiner, S., Stalder, H., & Thiel, V. (2020). Coronavirus biology and replication: implications for SARS-CoV-2. *Nature Reviews Microbiology*, *19*(3), 155–170. https://doi.org/10.1038/s41579-020-00468-6

Chandra, A., Gurjar, V., Qamar, I., & Singh, N. (2020). Identification of potential inhibitors of SARS-COV-2 endoribonuclease (EndoU) from FDA approved drugs: a drug repurposing approach to find therapeutics for COVID-19. *Journal of Biomolecular Structure and Dynamics*, 1–11. https://doi.org/10.1080/07391102.2020.1775127

Baby, K., Maity, S., Mehta, C. H., Suresh, A., Nayak, U. Y., & Nayak, Y. (2021). Targeting SARS-CoV-2 Main Protease: A Computational Drug Repurposing Study. *Archives of Medical Research*, *52*(1), 38–47. https://doi.org/10.1016/j.arcmed.2020.09.013

Zhou, B., Thao, T.T.N., Hoffmann, D. *et al.* (2021). SARS-CoV-2 spike D614G change enhances replication and transmission. *Nature* 592, 122–127. https://doi.org/10.1038/s41586-021-03361-1

Khan, M. I., Khan, Z. A., Baig, M. H., Ahmad, I., Farouk, A. E., Song, Y. G., & Dong, J. J. (2020). Comparative genome analysis of novel coronavirus (SARS-CoV-2) from different geographical locations and the effect of mutations on major target proteins: An in silico insight. *PLOS ONE*, *15*(9), e0238344. https://doi.org/10.1371/journal.pone.0238344

Kim, Y., Jedrzejczak, R., Maltseva, N. I., Wilamowski, M., Endres, M., Godzik, A., Michalska, K., & Joachimiak, A. (2020). Crystal structure of Nsp15 endoribonuclease NendoU from SARS-CoV -2. *Protein Science*, *29*(7), 1596–1605. https://doi.org/10.1002/pro.3873

Cornillez-Ty, C. T., Liao, L., Yates, J. R., Kuhn, P., & Buchmeier, M. J. (2009). Severe Acute Respiratory Syndrome Coronavirus Nonstructural Protein 2 Interacts with a Host Protein Complex Involved in Mitochondrial Biogenesis and Intracellular Signaling. *Journal of Virology*, *83*(19), 10314–10318. https://doi.org/10.1128/jvi.00842-09

Sakai, Y., Kawachi, K., Terada, Y., Omori, H., Matsuura, Y., & Kamitani, W. (2017). Two-amino acids change in the nsp4 of SARS coronavirus abolishes viral replication. *Virology*, *510*, 165–174. https://doi.org/10.1016/j.virol.2017.07.019

Angelini, M. M., Akhlaghpour, M., Neuman, B. W., & Buchmeier, M. J. (2013). Severe Acute Respiratory Syndrome Coronavirus Nonstructural Proteins 3, 4, and 6 Induce Double-Membrane Vesicles. *MBio*, *4*(4). https://doi.org/10.1128/mbio.00524-13

Snijder, E., Decroly, E., & Ziebuhr, J. (2016). The Nonstructural Proteins Directing Coronavirus RNA Synthesis and Processing. *Coronaviruses*, 59–126. https://doi.org/10.1016/bs.aivir.2016.08.008

Ancar, R., Li, Y., Kindler, E., Cooper, D. A., Ransom, M., Thiel, V., Weiss, S. R., Hesselberth, J. R., & Barton, D. J. (2020). Physiologic RNA targets and refined sequence specificity of coronavirus EndoU. *RNA*, *26*(12), 1976–1999. https://doi.org/10.1261/rna.076604.120

Pillon, M. C., Frazier, M. N., Dillard, L. B., Williams, J. G., Kocaman, S., Krahn, J. M., Perera, L., Hayne, C. K., Gordon, J., Stewart, Z. D., Sobhany, M., Deterding, L. J., Hsu, A. L., Dandey, V. P., Borgnia, M. J., & Stanley, R. E. (2021). Cryo-EM structures of the SARS-CoV-2 endoribonuclease Nsp15 reveal insight into nuclease specificity and dynamics. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-020-20608-z

Notredame, C., Higgins, D. G., & Heringa, J. (2000). T-coffee: a novel method for fast and accurate multiple sequence alignment 1 1Edited by J. Thornton. *Journal of Molecular Biology*, *302*(1), 205–217. https://doi.org/10.1006/jmbi.2000.4042

Zhang, L., Lin, D., Sun, X., Curth, U., Drosten, C., Sauerhering, L., Becker, S., Rox, K., & Hilgenfeld, R. (2020). Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. *Science*, *368*(6489), 409–412. https://doi.org/10.1126/science.abb3405

Kneller, D. W., Phillips, G., O'Neill, H. M., Jedrzejczak, R., Stols, L., Langan, P., Joachimiak, A., Coates, L., & Kovalevsky, A. (2020). Structural plasticity of SARS-CoV-2 3CL Mpro active site cavity revealed by room temperature X-ray crystallography. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-16954-7

Figure Permission: https://creativecommons.org/licenses/by/4.0/legalcode

Kneller, D. W., Phillips, G., Weiss, K. L., Pant, S., Zhang, Q., O'Neill, H. M., Coates, L., & Kovalevsky, A. (2020). Unusual zwitterionic catalytic site of SARS–CoV-2 main protease revealed by neutron crystallography. *Journal of Biological Chemistry*, *295*(50), 17365–17373. https://doi.org/10.1074/jbc.ac120.016154

Verschueren, K. H., Pumpor, K., Anemüller, S., Chen, S., Mesters, J. R., & Hilgenfeld, R. (2008). A Structural View of the Inactivation of the SARS Coronavirus Main Proteinase by Benzotriazole Esters. *Chemistry & Biology*, *15*(6), 597–606. https://doi.org/10.1016/j.chembiol.2008.04.011

Mengist, H. M., Dilnessa, T., & Jin, T. (2021). Structural Basis of Potential Inhibitors Targeting SARS-CoV-2 Main Protease. *Frontiers in Chemistry*, *9*. https://doi.org/10.3389/fchem.2021.622898

Halford, B. (2020, September 17). *Pfizer's novel COVID-19 antiviral heads to clinical trial*. Chemical and Engineering News. https://cen.acs.org/pharmaceuticals/drug-discovery/Pfizers-novel-COVID-19-antiviral/98/web/2020/09

Zoete, V., Daina, A., Bovigny, C., & Michielin, O. (2016). SwissSimilarity: A Web Tool for Low to Ultra High Throughput Ligand-Based Virtual Screening. *Journal of Chemical Information and Modeling*, *56*(8), 1399–1404. https://doi.org/10.1021/acs.jcim.6b00174

Sterling, T., & Irwin, J. J. (2015). ZINC 15 – Ligand Discovery for Everyone. *Journal of Chemical Information and Modeling*, *55*(11), 2324–2337. https://doi.org/10.1021/acs.jcim.5b00559

Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shelley, M., Perry, J. K., Shaw, D. E., Francis, P., & 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. https://doi.org/10.1021/jm0306430

Madhavi Sastry, G., 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. https://doi.org/10.1007/s10822-013-9644-8

Liu, X., Zhang, B., Jin, Z., Yang, H., & Rao, Z. (2020). The crystal structure of COVID-19 main protease in complex with an inhibitor N3. *Nature*, 289–293. https://doi.org/10.2210/pdb6lu7/pdb

Mark, P., & Nilsson, L. (2001). Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *The Journal of Physical Chemistry A*, *105*(43), 9954–9960. https://doi.org/10.1021/jp003020w

Harder, E., Damm, W., Maple, J., Wu, C., Reboul, M., Xiang, J. Y., Wang, L., Lupyan, D., Dahlgren, M. K., Knight, J. L., Kaus, J. W., Cerutti, D. S., Krilov, G., Jorgensen, W. L., Abel, R., & Friesner, R. A. (2015). OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *Journal of Chemical Theory and Computation*, *12*(1), 281–296. https://doi.org/10.1021/acs.jctc.5b00864

Jorgensen, W. L., Maxwell, D. S., & Tirado-Rives, J. (1996). Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *Journal of the American Chemical Society*, *118*(45), 11225–11236. https://doi.org/10.1021/ja9621760

Li, J., Abel, R., Zhu, K., Cao, Y., Zhao, S., & Friesner, R. A. (2011). The VSGB 2.0 model: A next generation energy model for high resolution protein structure modeling. *Proteins: Structure, Function, and Bioinformatics*, *79*(10), 2794–2812. https://doi.org/10.1002/prot.23106

Cournia, Z., Allen, B., & Sherman, W. (2017). Relative Binding Free Energy Calculations in Drug Discovery: Recent Advances and Practical Considerations. *Journal of Chemical Information and Modeling*, *57*(12), 2911–2937. https://doi.org/10.1021/acs.jcim.7b00564

Daina, A., Michielin, O., & Zoete, V. (2017). SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Scientific Reports*, *7*(1). https://doi.org/10.1038/srep42717

Daina, A., & Zoete, V. (2016). A BOILED-Egg To Predict Gastrointestinal Absorption and Brain Penetration of Small Molecules. *ChemMedChem*, *11*(11), 1117–1121. https://doi.org/10.1002/cmdc.201600182

Trypsteen, W., Van Cleemput, J., Snippenberg, W. V., Gerlo, S., & Vandekerckhove, L. (2020). On the whereabouts of SARS-CoV-2 in the human body: A systematic review. *PLOS Pathogens*, *16*(10), e1009037. https://doi.org/10.1371/journal.ppat.1009037

Genheden, S., & Ryde, U. (2015). The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opinion on Drug Discovery*, *10*(5), 449–461. https://doi.org/10.1517/17460441.2015.1032936

Singh, N., & Warshel, A. (2010). A comprehensive examination of the contributions to the binding entropy of protein-ligand complexes. *Proteins: Structure, Function, and Bioinformatics*, *78*(7), 1724–1735. https://doi.org/10.1002/prot.22689

Li, L., Wang, L., & Alexov, E. (2015). On the energy components governing molecular recognition in the framework of continuum approaches. *Frontiers in Molecular Biosciences*, *2*. https://doi.org/10.3389/fmolb.2015.00005

Wermuth, C. G., Aldous, D., Raboisson, P., & Rognan, D. (2015). *The Practice of Medicinal Chemistry* (4th ed.) [E-book]. Academic Press.

Talley, K., Ng, C., Shoppell, M., Kundrotas, P., & Alexov, E. (2008). On the electrostatic component of protein-protein binding free energy. *PMC Biophysics*, *1*(1), 2. https://doi.org/10.1186/1757-5036-1-2

Friesner, R. A., Murphy, R. B., Repasky, M. P., Frye, L. L., Greenwood, J. R., Halgren, T. A., Sanschagrin, P. C., & Mainz, D. T. (2006). Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein−Ligand Complexes. *Journal of Medicinal Chemistry*, *49*(21), 6177–6196. https://doi.org/10.1021/jm051256o

Zhang, J., Zhang, H., Wu, T., Wang, Q., & van der Spoel, D. (2017). Comparison of Implicit and Explicit Solvent Models for the Calculation of Solvation Free Energy in Organic Solvents. *Journal of Chemical Theory and Computation*, *13*(3), 1034–1043. https://doi.org/10.1021/acs.jctc.7b00169

Onufriev, A. V., & Case, D. A. (2019). Generalized Born Implicit Solvent Models for Biomolecules. *Annual Review of Biophysics*, *48*(1), 275–296. https://doi.org/10.1146/annurev-biophys-052118-115325

McDonnell, PharmD, BCOP, A. M., & Dang, PharmD, BCPS, C. H. (2013). Basic Review of the Cytochrome P450 System. *Journal of the Advanced Practitioner in Oncology*, *4*(4). https://doi.org/10.6004/jadpro.2013.4.4.7

Benet, L. Z., Hosey, C. M., Ursu, O., & Oprea, T. I. (2016). BDDCS, the Rule of 5 and drugability. *Advanced Drug Delivery Reviews*, *101*, 89–98. https://doi.org/10.1016/j.addr.2016.05.007

Baell, J. B., & Holloway, G. A. (2010). New Substructure Filters for Removal of Pan Assay Interference Compounds (PAINS) from Screening Libraries and for Their Exclusion in Bioassays. *Journal of Medicinal Chemistry*, *53*(7), 2719–2740. https://doi.org/10.1021/jm901137j

Forouzesh, N., & Mishra, N. (2021). An Effective MM/GBSA Protocol for Absolute Binding Free Energy Calculations: A Case Study on SARS-CoV-2 Spike Protein and the Human ACE2 Receptor. *Molecules*, *26*(8), 2383. https://doi.org/10.3390/molecules26082383

Alexandrov, V. (2005). Normal modes for predicting protein motions: A comprehensive database assessment and associated Web tool. *Protein Science*, *14*(3), 633–643. https://doi.org/10.1110/ps.04882105

Altincekic, N., Korn, S. M., Qureshi, N. S., Dujardin, M., Ninot-Pedrosa, M., Abele, R., Abi Saad, M. J., Alfano, C., Almeida, F. C. L., Alshamleh, I., de Amorim, G. C., Anderson, T. K., Anobom, C. D., Anorma, C., Bains, J. K., Bax, A., Blackledge, M., Blechar, J., Böckmann, A., . . . Schlundt, A. (2021). Large-Scale Recombinant Production of the SARS-CoV-2 Proteome for High-Throughput and Structural Biology Applications. *Frontiers in Molecular Biosciences*, *8*. https://doi.org/10.3389/fmolb.2021.653148

Shao, J., Tanner, S., Thompson, N., Cheatham, T. (2007) Clustering Molecular Dynamics Trajectories :1. Characterizing the Performance of Different Clustering Algorithms. *Journal of Chemical Theory and Computation* 10.1021/ct700119m

Appendix

Figure A1

Protein Information for Nsp15 Including the Amino Acid Sequence and Secondary Structure Assignment

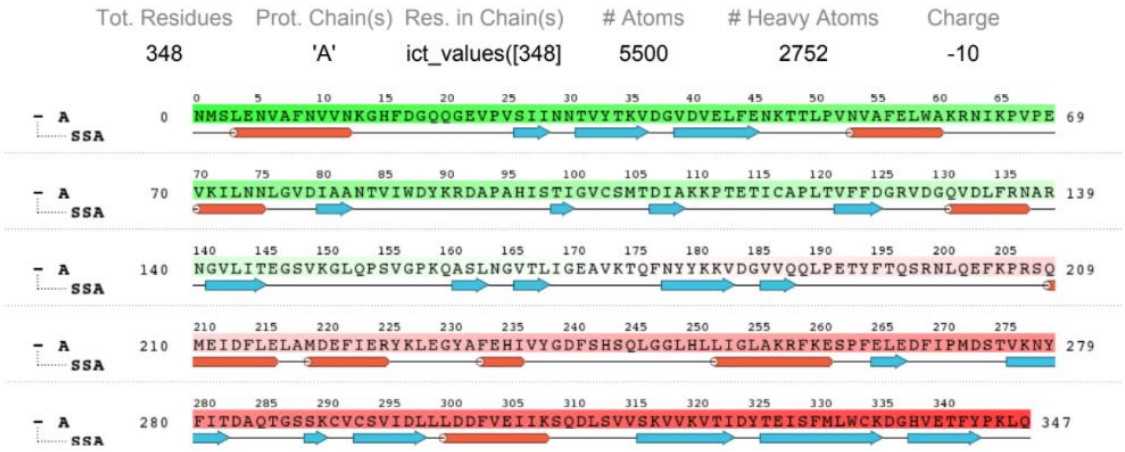


Figure A2

Protein Information for Mpro Including the Amino Acid Sequence and Secondary Structure Assignment