Fall 12-2023

Computational Molecular Docking Studies of Small Molecule Inhibitors With the SARS-CoV-2 Spike Protein Variants: In-Silico Drug Discovery Using Virtual Screening and Drug Repurposing Approaches

Grace Gupta
Chapman University, grgupta@chapman.edu

Computational Molecular Docking Studies of Small Molecule Inhibitors With the SARS-CoV-2 Spike Protein Variants: In-Silico Drug Discovery Using Virtual Screening and Drug Repurposing Approaches

A Thesis by

Grace Gupta

Chapman University

Orange, CA

Schmid College of Science and Technology

Submitted in partial fulfillment of the requirements for the degree of

Master of Science in Computational and Data Sciences

December 2023

Committee in charge:

Gennady Verkhivker, Ph.D., Chair

Cyril Rakovski, Ph.D.

Mohamed Allali, Ph.D.

CHAPMAN UNIVERSITY
SCHMID COLLEGE OF SCIENCE AND TECHNOLOGY
Computational and Data Sciences

The thesis of Grace Gupta is approved.

Gennady Verkhivker, Ph.D., Chair

Cyril Rakovski, Ph.D.

Mohamed Allali, Ph.D.

November 2023

Computational Molecular Docking Studies of Small Molecule Inhibitors With the SARS-CoV-2 Spike Protein Variants: In-Silico Drug Discovery Using Virtual Screening and Drug Repurposing Approaches

ABSTRACT

Computational Molecular Docking Studies of Small Molecule Inhibitors With the SARS-CoV-2 Spike Protein Variants: In-Silico Drug Discovery Using Virtual Screening and Drug Repurposing Approaches

by Grace Gupta

The pandemic caused by the emergence of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in 2019 has caused a global public health crisis of nearly unprecedented scale. In the years following the outbreak, the scientific community has mobilized to develop several vaccines and treatments. Drug repurposing as a strategy for drug development has produced many of the current therapeutic options. The greatest challenge to designing a therapeutic inhibitor of SARS-CoV-2 is the shifting mutational landscape of the virus as it evolves. In this study, we focus on the spike protein as a target for potential inhibitors. We explore two methods of inhibiting spike function, allosteric inhibition and direct inhibition. In the study of allosteric inhibition, we screened two compound libraries against two allosteric sites. In the study of direct inhibition, several top-performing direct inhibitors of the wildtype spike were evaluated against five variants, B.1.1.7, B.1.351, P.1, B.1.617.2, and B.1.1.529. In summary, we identify four potential allosteric inhibitors that warrant further in-vitro study. We also find that the direct potential inhibitors of the wildtype spike had the most similar performance against the B.1.617.2 and B.1.1.7 variants.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

1. An Overview of COVID-19

1.1 Introduction

The novel coronavirus disease (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has been confirmed responsible for over 6 million deaths worldwide since its emergence in Wuhan, China in 2019 (Williamson, 2020). The virus is transmitted through the inhalation of viral particles inside respiratory droplets from coughing or sneezing (Lotfi, 2020). Clinical presentation ranges from asymptomatic or mild flu-like symptoms, accounting for approximately 80% of cases, to severe and critical symptoms including pneumonia, dyspnea, hypoxia, respiratory failure, and multiorgan failure (Brouqui, 2021). Death occurs in ~1% of all cases, but the severity of illness and lethality are highly correlated with age. Individuals over 65 years old are considered the highest risk demographic, accounting for ~80% of all deaths (Bonanad, 2020).

The disease burden is worsened by the potential to experience complications post-infection. Post-COVID symptoms, termed “Long COVID”, occur in nearly one in five of adults who have contracted the illness (Raveendran, 2021). Symptoms of Long COVID are persistent fatigue, shortness of breath, and chronic sinusitis/congestion among other cardiopulmonary and naso-oropharyngeal complications, but also include disruptions to normal neuro-psychological functioning, such as memory loss and cognitive impairment (Aiyegbusi, 2021). The primary defense of governments against COVID-19 has been vaccination. Several vaccines were quickly developed in the years following the outbreak. The Pfizer-BioNTech vaccine was approved in August 2021 by the FDA for use in individuals 16 years or older. The Johnson and Johnson and

Moderna vaccines were authorized for distribution in the U.S. in February 2021, and January 2022 respectively. In addition to vaccination efforts, social distancing and self-quarantine protocols are recommended in the case of a positive COVID-19 diagnosis. Mass vaccination has been effective in limiting the spread of the virus, but vaccines need to be updated as the virus mutates. Researchers have identified numerous variants of the SARS-CoV-2 virus circulating worldwide. As of May 2023, there are over ten variants being monitored by the CDC in the U.S., including sub-variants of the most recent Omicron lineage. In addition to ensuring the continued effectiveness of vaccines, efforts to discover therapeutics must also consider the shifting mutational landscape. Therefore, the development of novel efficacious drugs against the SARS-CoV-2 virus remains a critical and ongoing challenge.

1.2 Coronavirus Taxonomy

The SARS-CoV-2 virus belongs to the genus Coronavirus in virus taxonomy (Walls, 2020). Coronaviruses were first identified in the 1960s and belong to the *Orthocoronavirinae* subfamily of the *Coronaviridae* family in the *Nidovirales* order (Grellet, 2022). Coronavirus genomes consist of positive-sense single-stranded RNA and are the largest among RNA-viruses (Grellet, 2022). The natural hosts of coronaviruses are typically wild animals, such as bats, rodents, and birds, however coronaviruses are able to species-jump to infect humans. Human coronaviruses typically cause mild upper respiratory infections, such as in the case of the common cold. An estimated 15% of common colds in the U.S. are caused by coronaviruses (Mesel-Lemoine, 2012). However, the history of coronaviruses infecting humans has been punctuated by strains that have a distinct pathogenicity. The deadly SARS-CoV pandemic in 2003 led to heightened research interest in elucidating the emergence and structural mechanisms of human coronaviruses (Yang, 2020). Since the 2003 outbreak, four additional coronaviruses have been

found to infect humans, including human coronavirus NL63 (HCoV-NL63) (Van Der Hoek, 2004), human coronavirus HKU1 (HCoV-HKU1) (Woo, 2005), Middle East Respiratory Syndrome (MERS-CoV) (Zaki, 2012), and SARS-CoV-2 (Zhu, 2020).

1.3 Coronavirus Structure Overview

Coronaviruses are named for the appearance of their protein-studded spherical outer membrane, which appears crown-like. The outer membrane of all coronaviruses is a permeable lipid membrane bilayer that transports and protects the genetic material. The genome of the SARS-CoV-2 virus consists of positive-sense single-stranded RNA (Grellet, 2022). The nature of the coronavirus genome simplifies the translation and replication of the virus inside the host cell. In molecular biology, the translation of a genome into the proteins it encodes typically involves an intermediary stage between gene and protein, in the form of messenger RNA (mRNA). Due to its single-stranded RNA structure, a coronavirus genome itself acts as mRNA and therefore can be directly translated by the host cell. Coronavirus genomes are among some of the largest known viral genomes (~30 kb). The coronavirus genome encodes four main structural proteins in addition to several non-structural proteins (nsps) and accessory proteins (Forni, 2017) (Figure 1).

The four structural proteins encoded by the coronavirus genome are the envelope (E), membrane (M), spike (S), and nucleocapsid (N) proteins. Of these four proteins, the envelope, membrane, and spike proteins are embedded within the outer lipid membrane. The envelope and membrane proteins are heavily involved in viral assembly (Li, 2016). The spike protein functions to mediate cell entry by initiating interaction with the host cell. The nucleocapsid protein encapsulates the viral genetic material inside the viral coating. The coronavirus genome also encodes several non-structural proteins (nsps) and accessory proteins that are vital for its ability to replicate, including

the nsp12 RNA-dependent RNA polymerase (RdRp) and nsp5 protease (3CLpro; Mpro) (Stobart, 2013). These non-structural proteins create the machinery that allows the translation and replication of the virus inside the host cell.

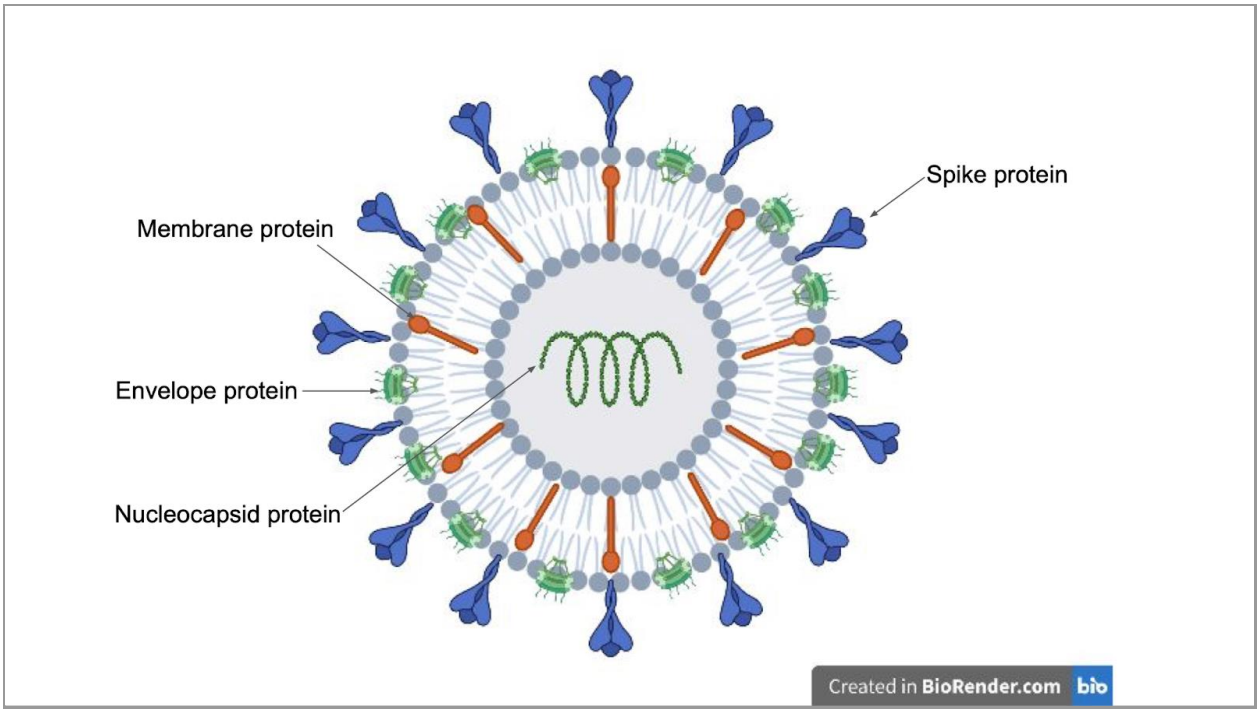

Figure 1. Structure of SARS-CoV-2.

The envelope protein consists of lipids and proteins that are partially derived from the host cell during virus maturation. Other proteins within the envelope are virus-encoded, including the spike glycoprotein. The virus-encoded proteins are associated with various activities including binding to host receptors, facilitating fusion, and receptor modification (Harrison, 2005). The envelope protein interacts with the membrane protein and other accessory proteins including ORF3a, ORF7a, and host cell proteins during the replication cycle (Rahman, 2021). Envelope proteins are also involved in the budding and release of new viruses from the host cell.

The membrane glycoprotein is a transmembrane protein which consists of an N-terminal ectodomain and a C-terminal endodomain (Lenard, 2008). The primary function of the membrane protein is to protect the nucleocapsid. The membrane consists of a continuous lipid bilayer, in which various transmembrane proteins are embedded. These transmembrane proteins allow the virus to both attach and fuse to the host cell membrane (Lenard, 2008). In addition to its role in viral assembly and budding, studies have demonstrated that the membrane protein cooperates with the spike protein during cell entry (Bianchi, 2020).

The nucleocapsid protein binds the viral genome into a compact structure for transport. It is composed of two main domains capable of binding RNA connected by a linker region. The nucleocapsid protein helps promote genome packaging through its interaction with RNA chaperone proteins, which aid in the folding of the genomic material (Gao, 2021). The nucleocapsid protein also plays a role in intracellular protein transport and DNA degradation. Additionally, the expression of these proteins has been shown to produce a strong host immune response. The nucleocapsid protein has been shown to be well-conserved among coronaviruses (Gao, 2021).

The spike protein is a transmembrane protein which projects from the viral surface. The primary function of the spike protein is to initiate fusion of the viral and host cell membranes. The spike protein is critical in all stages of fusion, including receptor recognition, viral attachment, and entry. The SARS-CoV-2 spike is highly conserved among coronaviruses (Huang, 2020). In our research, we will explore the spike protein as a target for drug development.

1.4 Spike Protein Structure

The spike protein sequence is 1273 aa and consists of two main components, the S1 subunit (14-685 residues) and the S2 subunit (686-1273 residues) (Huang, 2020) (Figure 2). The spike protein is trimeric, meaning that each subunit is composed of three identical monomers. The S1 subunit is the globular head of the spike protein, which contacts the host cell receptor.

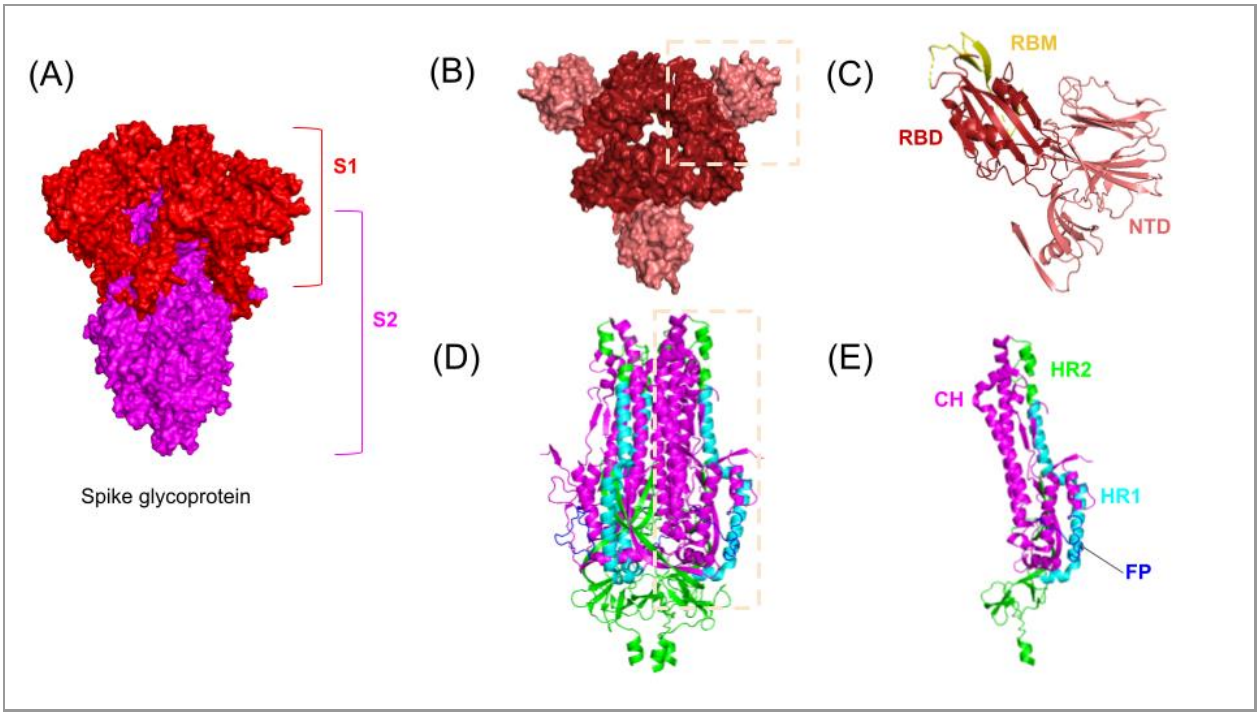

Figure 2. (A) Spike glycoprotein (PDB ID: 6VXX). S1 subunit colored red; S2 subunit colored magenta. (B) S1 trimer (top-down view). (C) Sub-structures of S1 monomer. (D) S2 trimer (side-view). (E) Sub-structures of S2 monomer.

The part of S1 that binds the host receptor is called the receptor-binding domain (RBD) (333-526 residues) (Lan, 2020). The RBD can be further subdivided into two regions, the RBD core and the receptor-binding motif (RBM). The RBM (443-503 residues) is a highly polar region of the RBD containing the contact residues which bind to the receptor (Candido, 2022). The N-terminal

domain (NTD) of the S1 subunit is located adjacent to the RBD. The function of the NTD is not currently known, though it is speculated to also assist in binding the host receptor (Cheng, 2019).

The S2 subunit is the stalk of the spike protein which maintains the attachment to the host receptor and facilitates fusion. The S2 subunit contains an internal fusion peptide (FP), central helix, (CH), and two hydrophobic heptad repeat regions, HR1 and HR2. During fusion, the FP and heptad regions of the S2 subunit shift in conformation to come into close contact with the host transmembrane anchor, permitting the fusion of the viral and host membranes (Bosch, 2003).

1.5 Activation of the Spike Receptor Binding Domain

The spike RBD has been discovered experimentally to take two distinct conformations, an ‘up’ state and a ‘down’ state (Wrapp, 2020) (Figure 3). In the ‘down’ conformation, the RBD is segregated away from the extracellular matrix, tucked towards the rest of the S1 subunit. In the ‘up’ conformation, the RBD protrudes outward, exposing it to contact the ACE-2 receptor. It is speculated that the ‘down’ conformation of the RBD is its natural state (Wrapp, 2020). In this segregated ‘down’ state, the spike protein is more likely to evade immune response. The trigger of the switch between conformations is not fully understood, although the ‘up’ state has been correlated with the interaction of the spike protein with several host proteases.

1.6 Entering Mechanism of SARS-CoV-2

Entry of the SARS-CoV-2 virus into host cells through the fusion of the viral and host cell membranes can be divided into three main stages. In the pre-fusion stage, the spike RBD

contained in the S1 subunit recognizes and binds the host cell receptor (Figure 4). The SARS-CoV-2 receptor in human cells is the angiotensin-converting enzyme 2 (ACE2). The formation of the spike-ACE2 complex causes a cleavage site between the S1 and S2 subunits, called the S2' site, to be exposed (Guo, 2023).

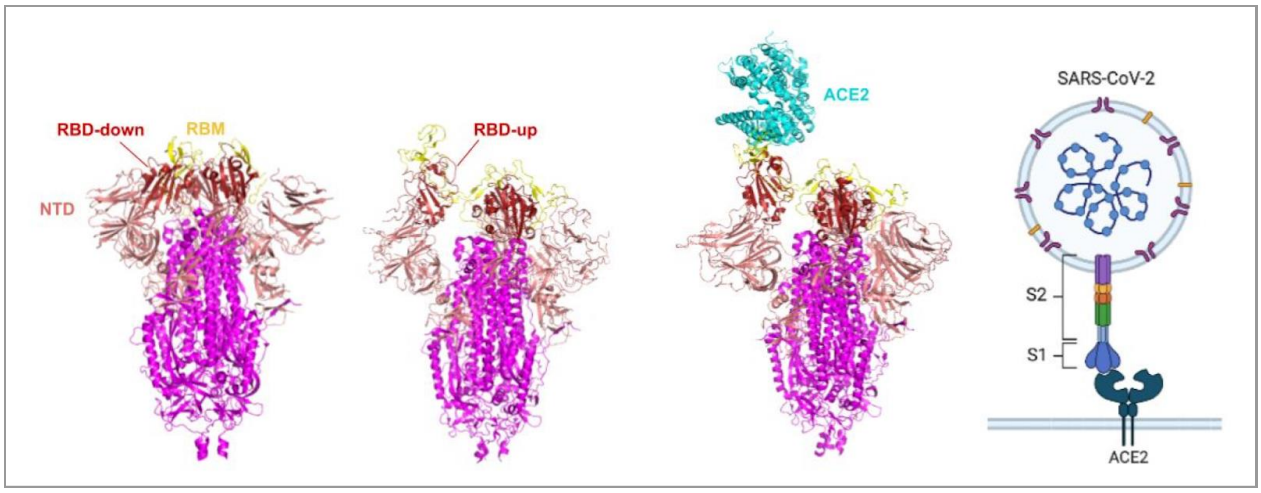

Figure 3. RBD-down and RBD-up conformations of the SARS-CoV-2 spike protein. Also depicted: spike RBD-up bound to human receptor ACE2.

Binding to ACE2 causes conformational changes in the spike S1 subunit to expose the S2' cleavage site. SARS-CoV-2 is distinguished from the SARS-CoV virus by the presence of a cleavage site at the S1-S2 junction, which is cleaved by furin (Guo, 2023). Following cleavage of the S1-S2 subunit boundary, the S2' site is cleaved by a type II transmembrane protein (TMPRSS2) or a cathepsin protein expressed by the host cell. The S1-S2 boundary cleavage causes the dissociation of S1, exposing the S2' cleavage site. The dissociation of S1 propels HR1 of the S2 subunit towards the host membrane. The fusion peptide (FP) of S2 can insert itself into the host membrane. Lastly, the HR1 and HR2 regions of the S2 subunit form a tight bundle allowing the fusion of the viral membrane and the host cell membrane (Guo, 2023).

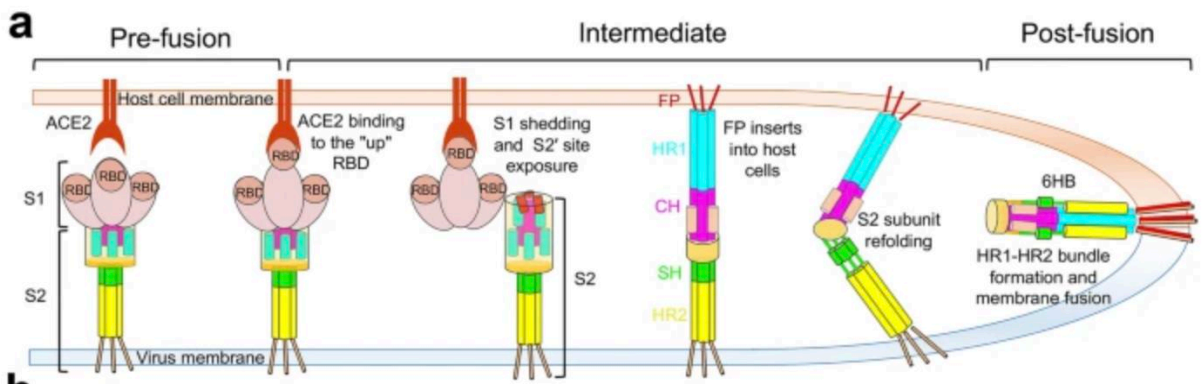

Figure 4. Image credit: Guo, 2023. Illustration of the attachment and fusion of SARS-CoV-2 spike to the host cell.

1.7 Antibody Binding to SARS-CoV-2 Spike

Several antibodies have been developed that target the spike protein. The region of the antigen recognized by an antibody is referred to as its epitope. The epitope of all the current antibodies designed to bind to the spike protein is the spike receptor-binding domain (RBD). In other words, these antibodies are specific to the RBD, which is located on the S1 subunit.

The antibody class is used to denote the specific region of the RBD recognized by the antibody. Class 1 antibodies, such as COVOX-222, compete with the human receptor ACE2 for RBD binding. These antibodies only recognize the RBD when it is in the ‘up’ position. Class 2 antibodies, such as LY-CoV555, also compete with ACE2 to bind the RBD, however they can recognize the RBD in both its conformational states. Class 3 residues bind a region close to residue N343. LY-CoV1404, a Class 3 antibody, is an effective neutralizer of variants of SARS-CoV-2, including the B.1.1.7, B.1.351, and B.1.617.2 (Westendorf, 2022). The contact residues of LY-CoV1404 within the Class 3 epitope are highly conserved. Lastly, the epitope of Class 4

anti-RBD antibodies target a less accessible region facing the interior of the protein on RBD's in the 'up' conformation.

Sotrovimab	RBD class 3 antibody with non-RBM epitope.
Bebtelovimab	RBD class 3 antibody with non-RBM epitope.
Tixagevimab+Cilgavimab	RBD class 3 antibodies with non-RBM epitopes.
Bamlanibimab+Etesevimab	RBD class 2 antibodies with non-RBM and epitopes.
Casirivimab+Imdevimab	RBD class 1 antibody with overlapping RBM epitope.

Table 1. Current antibody therapies with emergency use approval.

A few of these antibody therapies have received emergency use authorizations by the FDA (Table 1) (Shrestha, 2021). The receptor-binding motif is the area of the RBD that directly binds the ACE-2 receptor. Most of the antibodies target regions of the RBD that do not overlap with the receptor-binding motif (RBM) (Shrestha, 2021). The SARS-CoV-2 virus has a high rate of mutation. Parts of the viral genome, such as the RBM, mutate more quickly than others (Amicone, 2022). The receptor-binding motif of the spike protein is subject to higher selection pressures due to its crucial role in interfacing with the host cell and is therefore a region likely to accrue mutations.

The antibody therapies approved for emergency use are relatively recent and continue to be vetted against newer variants, such as omicron. The omicron variant has many mutations in the spike RBD, within or affecting the epitope of antibodies such as Sotrovimab and Bebtelovimab

(Mahase, 2022). The mutability of new variants poses a challenge to the current development strategy of antibody-based therapeutics.

1.8 Allosteric Binding of SARS-CoV-2 Spike

Several studies have tested the ability of compounds to bind the spike protein to indirectly inhibit spike function. Indirect inhibition is referred to as allosteric inhibition. An example of a promising allosteric compound targeting the spike protein is linoleic acid. The shape of a hydrophobic binding pocket within the spike RBD on the S1 subunit was predicted to fit the shape of linoleic acid (Toelzer, 2020). Since the binding pocket does not overlap with the active site, linoleic acid is considered a potential allosteric inhibitor.

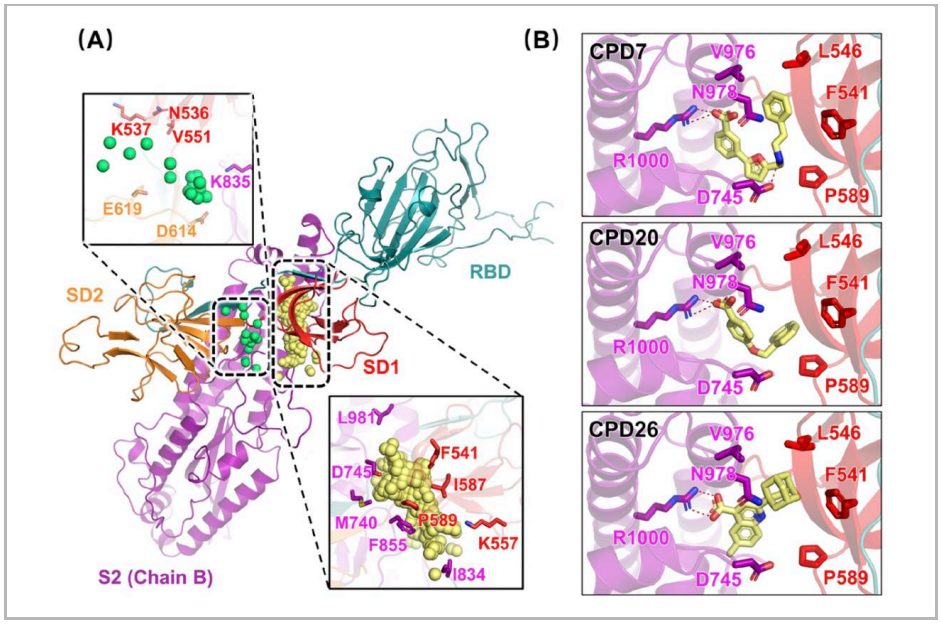

Figure 5. Image credit: Wang, Qian. et al. 2021. (A) Spike protein and two main binding regions of the binding pocket identified at the S1/S2 subunit junction. (B) Binding conformations of CPD7, CPD20, and CPD26.

Another study discovered an allosteric binding pocket at the junction of the S1 and S2 subunits (Wang, 2021) (Figure 5). This binding pocket is further removed than the previous hydrophobic

pocket, as it does not overlap with the spike RBD. The junction between the S1 and S2 subunits creates a shallow pocket. Researchers suggest that binding of this region could allosterically inhibit spike function by stabilizing the closed state of the spike RBD (Wang, 2021). The results of this study indicated that compounds with high binding affinity for this region (denoted CPD7, CPD20, and CPD26) tended to cluster in two main areas inside the binding pocket.

In the first part of our research, we are interested in both validating the results of this study in addition to testing the ability of novel ligands to bind this region. Further details are described in Chapter 4.

2. Computer Aided Drug Discovery and Molecular Docking Algorithms

2.1 Introduction

The evolution of drug discovery can be traced back to the nineteenth century, when advances in chemistry and biology created the conditions for the serendipitous discovery of medicinal compounds. Chloral hydrate, for example, is one of the oldest sedatives that was synthesized as the byproduct of an unrelated experiment (Ban, 2022). Although the development of many of the first medicines was accidental, novel therapeutics were intentionally formulated from the extracts of traditional remedies. Aspirin, or acetylsalicylic acid, was developed by identifying salicylic acid as the active ingredient in willow bark, a natural fever reducer (Ugurlucan, 2012). Classical pharmacology began to take shape based on this general principle: first observe a therapeutic effect, then find the compound sufficient to reproduce it. Only after an active compound was identified was there an effort to determine the mechanism of action.

The modern drug discovery paradigm reverses these steps by starting with a specific target biomolecule, typically a protein (Figure 6). The reverse approach to pharmacology became popular after the complete sequencing of the human genome, which has made the rapid sequencing of purified proteins possible (Rao, 2011). Researchers begin with an understanding of the disease etiology, drawing from scientific literature, publicly available protein and chemical databases, and direct biochemical experimentation to identify a viable drug target. A viable drug target must play a crucial role in producing the disease phenotype. For example, the target protein may be an enzyme that translates the genetic material of the virus inside the host cell.

The target protein is isolated, purified, and crystalized in order for its structure to be resolved using nuclear magnetic resonance (NMR) and X-Ray diffraction (Pellecchia, 2002). Next, chemical libraries containing drug-like compounds are assessed experimentally to identify compounds with high affinity for the target protein. Promising compounds that progress to clinical trials are referred to as lead compounds. Lead compounds are optimized to improve their therapeutic effect and reduce adverse effects before they are finally assessed for their *in vivo* therapeutic effect and efficacy during clinical trials.

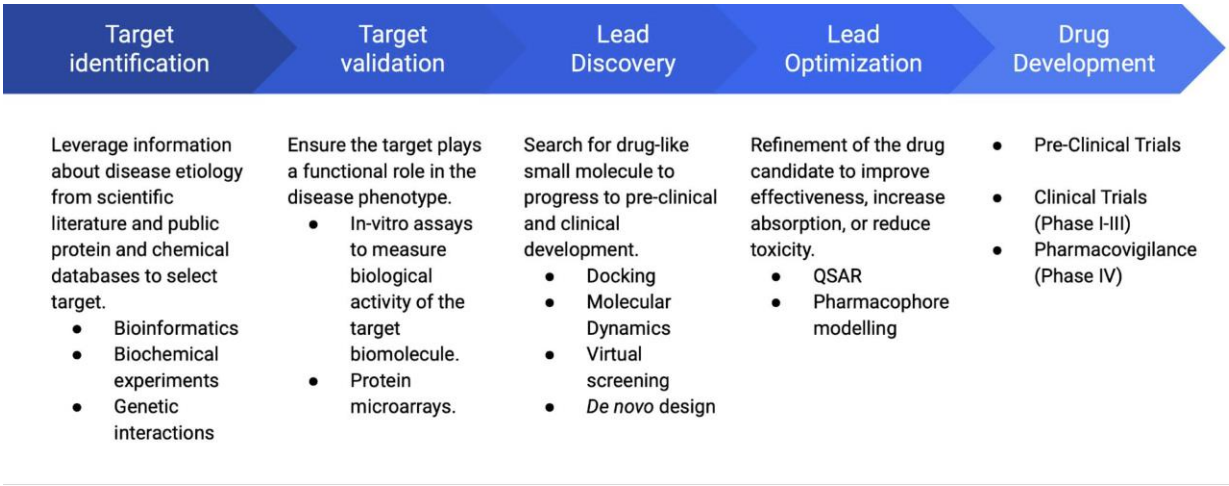

Figure 6. Stages of modern drug development.

Computational techniques play an important supplemental role to experimental methods in lead discovery and optimization. Compound libraries are not only screened *in vitro*, but also virtually through virtual screening software (Parvathaneni, 2019). Virtual screening can be used to efficiently approximate the binding affinity of many thousands of compounds to a specific target protein. All potential drugs must be experimentally validated regardless of how they perform in simulation. However, virtual screening is a time-efficient and cost-effective method to quickly filter large compound libraries for the candidates that are most likely to perform well experimentally.

Computational methods also have useful applications in drug optimization, for example, quantitative structure-activity relationship (QSAR) models are commonly used to predict the performance of lead candidates. QSAR models are regression-based models designed to predict the chemical activity of a compound based on its physico-chemical properties and molecular description (Neves, 2018). Lastly, software has been developed to permit the design of compounds from scratch, in a process called *de novo* drug design. *De novo* design relies on the information available on the structure of the target and any known active binder molecules to construct a novel compound (Hartenfeller, 2011). Broadly, the major roles of computational approaches in drug discovery are the following: the virtual screening of large compound libraries for promising compounds to be tested experimentally, the optimization of lead compounds, and the *de novo* design of novel compounds.

2.2 Molecular Docking and Virtual Screening

Computational drug design relies on simulation to approximate the fit of a target protein and candidate compounds based on their molecular structures and the strength of physico-chemical interactions. The simulation process is referred to as molecular docking, in which a compound is "docked" into the target protein (Figure 7).

Virtual screening in drug development describes the overall process of running many iterations of molecular docking experiments to filter a set of candidate drug compounds. The concept of molecular docking evolved from the "Lock and Key" model of understanding receptor-ligand binding, introduced by Emil Fischer in 1894, who was inspired by the geometry of complementary shapes (Jorgensen, 1991). In pharmacology, a compound that forms a complex

with a biomolecule, such as a protein, is referred to as a ligand (Ekins, 2007). The target protein is referred to as the receptor. According to this model of ligand-receptor dynamics, the ideal ligand is one that serves as the unique key to its respective lock (the binding site of the receptor). Most ligands, however, do not achieve a perfect fit to the receptor. Therefore, the model was revised to view the ligand and receptor as dynamic entities which modulate the other's shape.

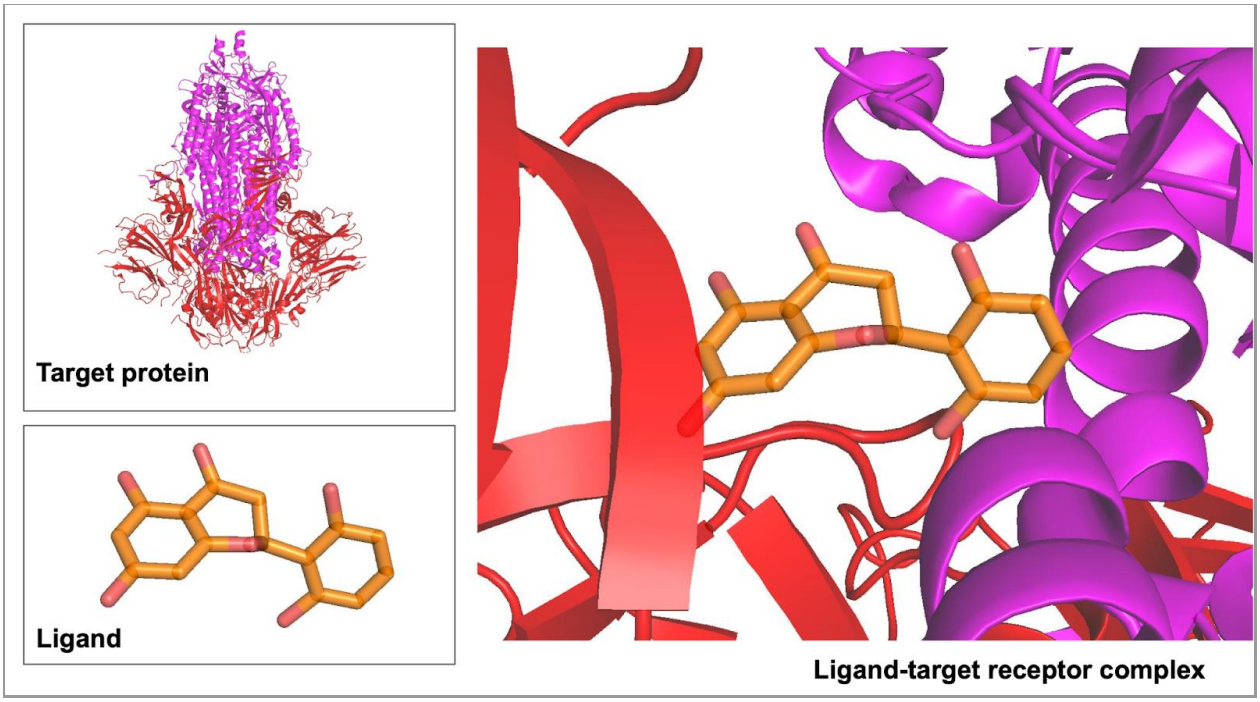

Figure 7. Illustration of the simulated binding of a ligand with a specified target protein receptor during molecular docking.

Docking algorithms that model the ligand with conformational flexibility are known as flexible docking algorithms. Flexible docking is more CPU-intensive than previous rigid docking algorithms, since it is more computationally expensive to consider a larger search space. However, flexible docking is nonetheless the standard due to its improved accuracy (Bonvin, A. M. 2006).

2.3 Molecular Docking Search Algorithms

All docking programs use a search algorithm to assess the conformational space for the most stable ligand-receptor complex. The stability of the ligand-receptor complex is correlated with its free energy. A complex with a high degree of free energy is more likely to dissociate. A search algorithm attempts to converge on the global minimum energy conformation. An exhaustive search of the entire conformational space is costly and increases with the number of rotatable bonds in the ligand-receptor complex.

Search algorithms can either be designed to search a limited portion of the conformational space, termed systemic search, or they may sample the entire conformational space, referred to as stochastic search.

2.4 Autodock Vina Search Algorithm

Autodock Vina, a popular virtual screening software used in our research, employs three types of stochastic search algorithms to find the global minimum energy ligand-receptor conformation. These search algorithms are simulated annealing (SA), traditional genetic algorithm (GA), and Lamarckian genetic algorithm (LGA) (Trott, 2010).

Simulated annealing is named after the physical process of "annealing" in metallurgy, which is the controlled heating and cooling of a metal. At high temperatures, metallic atoms move with a high degree of stochasticity, which allows the elimination of impurities before the metal is cooled to a pure crystalline structure. Simulated annealing borrows this principle of introducing a controlled degree of stochasticity before allowing the algorithm to converge, to avoid becoming stuck at a local minimum (Aarts, 2005). The introduced stochasticity may have the adverse effect

of shifting convergence from a better solution to a worse solution. However, this tradeoff is often acceptable for optimization problems that are computationally intractable with a large number of variables and many potential local minima.

The genetic algorithm (GA) and Lamarckian genetic algorithm (LGA) are optimization methods inspired by natural selection. The driving principle is to repeatedly modify a "population" or set of solutions by applying a fitness function, then stochastically select the most fit individual solutions. Each solution's properties are modified to form a new generation used in the next iteration of the algorithm. The algorithm may terminate when either a maximum generation threshold has been met or the fitness of the population of solutions has reached a sufficient level (Fuhrmann, 2010). The Lamarckian genetic algorithm differentiates itself from the traditional genetic algorithm by allowing the next generation to "inherit" the local search adaptations of their parents. These algorithms are sensitive to initial conditions and may also produce local minima solutions, which is why it is often used in conjunction with other search algorithms.

2.5 Docking Scoring Functions

Once the lowest energy conformation has been identified, it is assigned a score based on the fit of the ligand to the receptor, and the resulting stability of the complex. The scoring function evaluates a given ligand-receptor conformation based on the affinity the ligand has for the receptor. A ligand is assumed to have greater affinity if the complex is unlikely to dissociate.

The binding affinity of a complex is typically measured as the equilibrium dissociation constant (*KD*), as given below, where [L] and [P] are the molar concentrations of the ligand, receptor protein, and ligand-receptor protein complexes respectively (Belitzky, 2022). A lower binding

affinity indicates a stronger attraction, while a higher binding affinity could be caused by a weak attraction or a poor fit. Since the search function may converge on any conformation that is theoretically possible in a simulation, a good scoring function must be able to distinguish the conformations that are likely to appear experimentally.

$$KD = \frac{[L][P]}{[LP]}$$

An ideal scoring function must also assign a binding score that correctly predicts and ranks its potency. Although significant progress has been made in improving the accuracy of scoring functions, standard benchmarks indicate that these functions unilaterally struggle to accurately predict and rank receptor-ligand binding affinities. Therefore, many virtual screening programs use a consensus approach to score the ligand-receptor complexes.

2.6 Autodock Vina Scoring Function

Autodock Vina uses a hybrid function that combines two of the most common scoring functions: force-field and empirical-based scoring. Force-field based functions estimate binding affinity by the summation of intermolecular forces, including van der Waals and electrostatic forces (Hill, 2015). These functions consider the interaction energies between non-bonded atoms as well as the internal ligand energy. The Autodock Vina scoring function uses atom pairs and evaluates for factors relating to dispersion and repulsion, as well as accounting for hydrogen bonding, electrostatic forces, and desolvation (Trott, 2010).

$$E = \sum_i \sum_j \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_y^6} + \frac{q_i q_j}{\varepsilon(r_{ij}) r_{ij}} \right)$$

Empirical scoring makes use of experimental data to assign a binding score. The scoring function can be linear or nonlinear, though typically a multiple linear regression model is used. Each variable in the function represents a key factor contributing to the total free energy of the bound ligand-receptor complex (Warren, 2006). Experimental datasets containing three-dimensional protein-ligand complexes with their respective binding affinities are used to weight the coefficients of the regression (Guedes, 2018).

An example of the empirical scoring function used by the GOLD docking software rewards favorable interactions ("S") that increase stability, such as the formation of hydrogen bonds and coordinate bonds with metal ions. The function also penalizes ("P") a conformation for internal strain and the unnatural overlap of atoms (steric clashes) (Liebeschuetz, 2012).

$$Score = S_{H-bond} + S_{metal} + S_{lipophilic} + p_{rotor} + p_{strain} + p_{clash} + [p_{covalent} + p_{constraint}]$$

Knowledge-based scoring functions assume that two atoms that make contact more frequently are likely to have an energetically favorable interaction. A statistical analysis of structural data, usually retrieved from protein databases such as the Protein Data Bank (PDB), is used to estimate pairwise atomic potentials (Gohlke, 2000). A training set is created based on the structural data, which contains every atomic pair observed in the structure and its occurrence frequency. In comparison to empirical scoring, knowledge-based scoring functions assess all contacts between protein and ligand atoms for their energetic contribution, as opposed to prioritizing select interactions.

Autodock Vina uses a hybrid empirical and knowledge-based scoring function. The conformational-dependent part of the scoring function is given below (Trott, 2010).

$$c = \sum_{i<j} f_{t_i t_j}(r_{ij})$$

A symmetric set of interaction functions between atom pair (*ti,tj*) is related to their interatomic distance *rij*. This function can also be written as the summation of all intermolecular and intramolecular interactions between the protein and ligand atoms (Trott, 2010).

$$c = c_{inter} + c_{intra}$$

The Autodock Vina hybrid scoring function determines the interaction functions and interatomic distances using information from both knowledge-based structural data and empirical-based experimental affinity data.

2.7 Virtual Screening Protocol Using Vina

A typical virtual screening pipeline begins with the ligand molecules, which can be downloaded from public databases such as PubChem or the ZINC database (Figure 8). These molecules are structure-data files (sdf) which encode chemical structure properties. These files are then converted to Protein Data Bank (PDB) format, which allows the molecules to be rendered in 3D. PDB files are then converted to PDBQT format, which includes additional information about atom charges and types that is necessary to simulate the most accurate interactions between ligand and receptor. These conversions can be performed using any translation software, such as OpenBabel, which can convert a variety of file formats used in bioinformatics (N M O'Boyle, 2011). The receptor molecule can typically be downloaded as a PDB format directly from the Protein Data Bank.

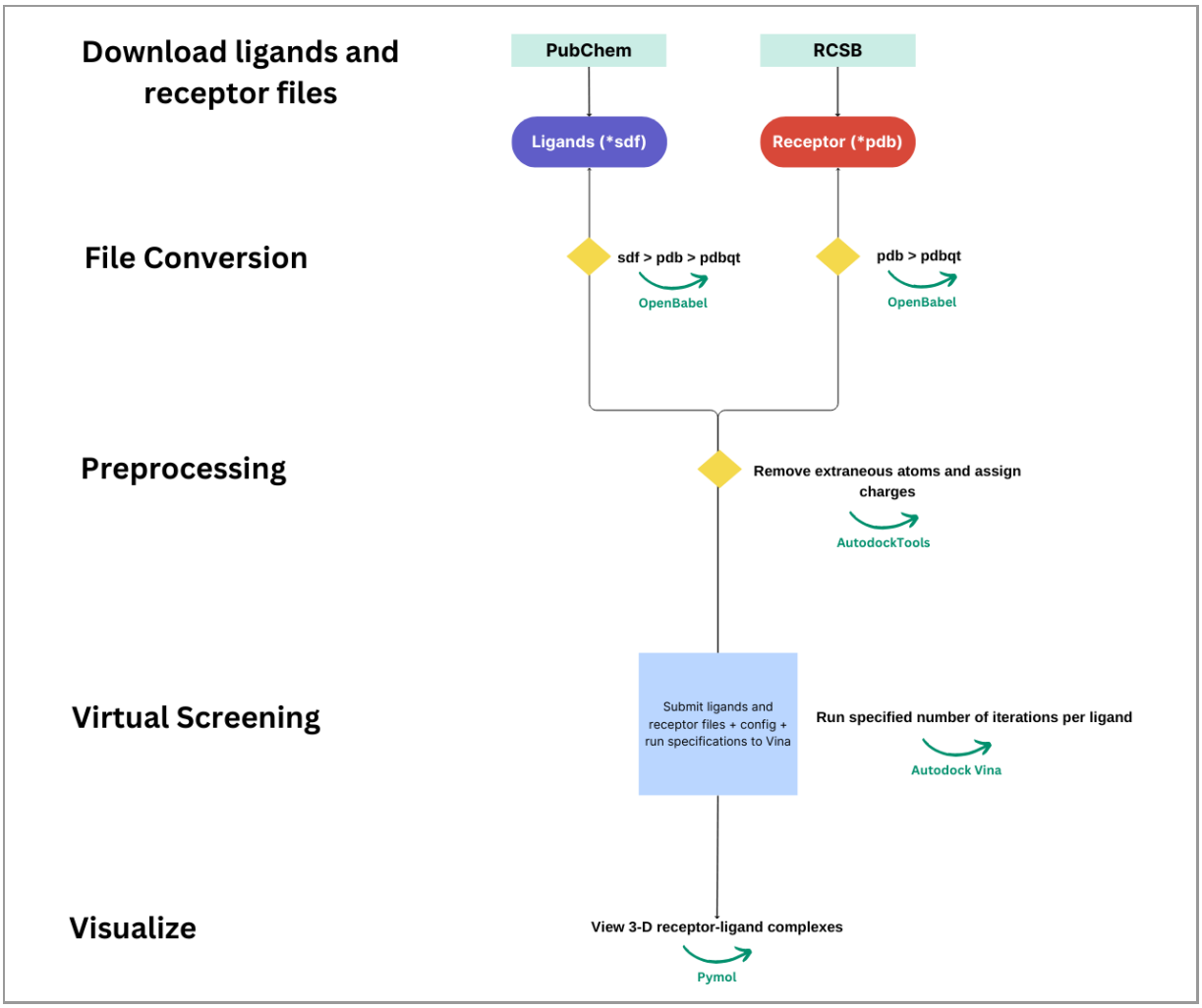

Figure 8. Illustration of the virtual screening methodology used in this research. Molecular docking conducted using Autodock Vina.

The ligand and receptor molecules must also be preprocessed before docking. Docking algorithms rely on the physico-chemical properties of the receptor and compounds to accurately predict their interaction (Cho, 2005). Therefore, extraneous atoms in the crystal structure of the receptor pdbqt file must be removed and appropriate electronic charges must be assigned to the atoms of the receptor and ligands. AutodockTools is a popular software used for pre-processing (Morris, 2009).

To perform virtual screening, the pdbqt files of all ligands and the receptor are submitted as input to Vina, along with a configuration file. The configuration file specifies the exact coordinates of the docking site as a grid box. The configuration file references a script that specifies the number of molecular docking runs per ligand (Appendix A). All ligand-receptor conformations will be confined to this search space. The output of Vina includes a text file summarizing the top nine ligand-receptor conformations and their corresponding binding affinities. Vina also outputs each conformation as a pdbqt file for visualization and analysis. PyMol is a popular python-based tool that can be used to visualize Vina output files (Schrodinger, 2015).

Autodock Vina is a popular tool for conducting the virtual screening of many thousands of compounds against a target of interest. Molecular docking methods and their associated searching and scoring algorithms are the foundation of the computational approach to drug discovery. In conjunction with the public accessibility of chemical databases containing FDA-approved drugs in file formats compatible for virtual screening, screening software such as Autodock Vina becomes a powerful tool in the discovery process of repurposing drugs against a novel disease.

2.8 Vina Virtual Screening as Applied to SARS-CoV-2

Previous studies have used Autodock Vina to conduct virtual screening against several targets within SARS-CoV-2, such as the main protease and non-structural proteins. Several studies have used this virtual screening software to explore inhibition of the spike protein. These studies differ slightly in their methodologies; however, the general protocol of file conversion, preprocessing, virtual screening, and visualization is foundational. One study tested the binding of chloroquine, remdesivir, ribavirin and luteolin against several target proteins, including the spike protein (Yu,

2020). The authors of this study did not target a specific site or substructure, instead they were interested in finding any possible binding pockets on the spike protein. Docking to the whole surface of a protein is referred to as blind docking and is preferred when the purpose of the experiment is to discover previously untested binding sites or when the target site is unknown.

In the first part of our study, we will perform blind docking of the spike protein using a different suite of compounds. We will also be testing these compounds against several variants of SARS-CoV-2. A study performed in 2022 tested several organic acids and phenolic compounds (e.g. astragalin, 4-p-Coumaroylquinic acid, and 3-p-Coumaroylquinic acid) against the spike S2 subunit. These compounds were found to directly bind the S2 subunit with high affinity (Jose, 2022). Since the S2 subunit is not considered the active site of the spike protein (the S1 subunit is the part of the spike protein that binds the human ACE2 receptor), the compounds in this study do not directly inhibit the spike. Instead, they have the potential to disrupt function indirectly. In the second part of our study, we will explore indirect inhibition of the spike protein.

3. Drug Repurposing and Applications for SARS-CoV-2

3.1 Drug Repurposing

Drug repurposing is the investigation of existing market drugs for their therapeutic effect outside of their original purpose (Pushpakom, 2019). The traditional *de novo* approach to drug discovery often takes many years to take a drug from conception to market. Drug repurposing is the preferred approach to drug discovery during public health emergencies, when there is an urgent demand for an effective therapeutic. Repurposed drugs also have the benefit of having passed preclinical safety trials and are less likely to have unknown deleterious effects. The overall timeline for drug development can therefore be shortened.

In the beginning stages of the pandemic, we saw the repurposing of the antiviral remdesivir to treat COVID-19. Prior to the pandemic, remdesivir was in the process of being evaluated for its effect against the Ebola virus and had not yet received FDA approval. However, in 2022, Remdesivir became the first drug to be issued an Emergency Use Authorization (EUA) by the FDA for the treatment of COVID-19 in May 2020 (Simonis, 2021). Other repurposed drugs include the antimalarial drugs Chloroquine and Hydroxychloroquine, which were used in the early stages of the pandemic to treat patients with severe illness (Lei, 2020). A 2019 review found that approximately 30% of all newly approved drugs in the U.S. are repurposed (Park, 2019).

Drug repurposing can be performed experimentally or computationally. Virtual screening is a type of computational repurposing, also referred to as “in silico” drug repurposing (Wadood,

2013). Generally, there are three major strategies for drug repurposing: knowledge-based, signature-based, and phenotype-based repurposing.

Knowledge-based drug repurposing is the most common strategy, which leverages what is known about the disease targets and pathways to select a drug that is likely to interact to produce a therapeutic effect (Reaume, 2011). Virtual screening is a type of knowledge-based drug repurposing. Virtual screening involves the in-silico screening of known drug-like compounds against a specific protein target. These compounds are retrieved from publicly available chemical libraries, such as PubChem, which provide information on chemical properties, known biological activity, safety and toxicity information, and molecular identifiers (Li Q. C., 2010). A popular compound library used for virtual screening is the ZINC compound database, which is a publicly available database containing over 230 million compounds for virtual screening purposes (Irwin, 2012).

Signature based drug repurposing compares gene expression profiles, called gene signatures. Instead of relying on knowledge of the specific disease pathways, the genetic expression of a cell can be compared before and after disease exposure to create a disease signature (Shukla, 2021). This signature can be compared with a drug-induced signature through differential expression analysis. If the gene expression in the disease signature is absent in the drug signature, this may be an indication that the drug is downregulating a gene that produces the disease phenotype. In this manner, a comparison of the drug and disease signatures may be used to support a hypothesis that the drug produces a therapeutic effect. Signature based drug repurposing has

been successfully used to repurpose topiramate, a drug used to treat epilepsy, for the treatment of inflammatory bowel disease (Hodos, 2016).

The last strategy is phenotype-based drug repurposing, which also avoids the need to identify the mechanism of action by screening drugs based solely on their ability to reduce the disease phenotype (Pham, 2021). Knowledge-based virtual screening methods in which a target must be identified are limited by the existing knowledge base. In comparison, phenotype-based approaches can uncover unexpected connections, such as off-target effects within a known disease pathway, that can be exploited for therapeutic use. Researching phenotypic approaches may also further elucidate disease-causing pathways, which expands the disease knowledge base.

3.2 Approved Therapeutics for COVID-19

Several therapeutics have been approved to treat COVID-19. The current CDC guidelines indicate three treatments as the most administered: Remdesivir (Veklury), Nirmatrelvir with Ritonavir (Paxlovid), and Molnupiravir (Lagevrio).

Remdesivir is a repurposed antiviral, previously used to treat Hepatitis C, that was authorized for emergency use in the early stages of the pandemic. Remdesivir became an FDA approved COVID-19 treatment in April, 2022 (Beigel, 2020). Remdesivir acts as an inhibitor of the RNA-dependent, RNA polymerase (RdRp), which is a viral enzyme crucial for replication inside the host cell. A recent clinical study that conducted randomized controlled trials in hospitalized adults with Covid-19 found that remdesivir has a high probability of reducing mortality for patients requiring ventilation (McCreary, 2020). Remdesivir has also been reported to shorten the overall time to recovery (Lin, 2021).

Nirmatrelvir with Ritonavir (Paxlovid) was granted an Emergency Use Authorization in December 2021 as an oral treatment for COVID-19 (Lamb, 2022). Paxlovid received full FDA approval in May 2023 for use in high-risk adult patients. Nirmatrelvir-ritonavir is also an antiviral developed by Pfizer which inhibits the viral 3CLpro or main protease (Lamb, 2022). This protease performs proteolytic cleavages at various stages in the viral life cycle. A recent study preceding its FDA approval found that treatment of symptomatic Covid-19 with nirmatrelvir-ritonavir reduced the risk of progression to severe illness by 89% (Dryden-Peterson, 2023).

Molnupiravir (Lagevrio) is an anti-viral that was also authorized for emergency use in December, 2021 (Singh, 2021). Molnupiravir is used to treat high-risk adults presenting with mild to moderate COVID-19. The mechanism of action of molnupiravir is not a direct structural interference like remdesivir and nirmatrelvir-ritonavir. Molnupiravir is a derivative of a small-molecule ribonucleoside (NHC), that can be incorporated into the viral genome (Imran, 2021). NHC uptake has been shown to cause mutations to occur during viral replication that significantly interfere with the progression of illness (Gordon, 2021). Molnupiravir has been reported to reduce the rate of mortality and in-hospital disease progression by 50% (Mahase, 2021).

Remdesivir, nirmatrelvir-ritonavir, and molnupiravir target parts of the viral machinery involved primarily with replication, namely the viral RNA-dependent, RNA-polymerase (RdRp) and the main protease (Mpro). However, other viral proteins may be suitable drug targets and are being

explored computationally and experimentally. The spike protein is being investigated as a third drug target.

A concern with any novel therapeutic is its ability to have consistently high binding affinity for the target site against new variants. Therefore, most therapeutics are designed to target conserved regions on the target protein, since they are unlikely to be significantly changed between variants. The newest variant lineage, Omicron, poses a particular challenge to designing inhibitors against the spike protein. The Omicron variant and sub-variants contain the most number of mutations in the spike RBD, with 10 mutations present in the RBM alone (Mannar, 2022). In comparison, the first variant of concern, the B.1.1.7 variant originating in the U.K., contained only one mutation in the entire RBD (Yang T. J., 2021). Although the approved spike inhibitors bind to regions that are the most conserved on the receptor binding domain, the mutability of the S-RBD remains a challenge to designing novel spike inhibitors targeting this region.

4. Identifying Candidate Allosteric Inhibitors of SARS-CoV-2 from FDA-Approved Drugs Using Molecular Docking Tools with Interaction Analysis

4.1 Introduction

The spike protein (S) is one of four structural proteins that make up the SARS-CoV-2 virus. The S protein contains two subunits, S1 and S2, which are involved in host cell attachment and cellular membrane fusion respectively (Huang Y. Y., 2020). Interaction with the host angiotensin-converting enzyme 2 (ACE-2) receptor is mediated by the receptor-binding domain (S-RBD) of the S1 subunit (Prajapat, 2020). This interaction is critical for establishing the pre-fusion state of the virus before its fusion and entry to the host cell. Specifically, S-RBD residues C480-C488 interact with the ACE-2 N-terminal helix (Kalhor, 2022). Several studies have tested the ability of small molecules to disrupt the formation of the spike-ACE2 complex by competitively binding the receptor binding domain (Wang G. Y., 2021). There are currently a few S-RBD inhibitors that have been approved by the FDA for emergency use, though none have yet received full approval.

The indirect disruption of the spike-ACE2 complex formation is also possible and less explored. Allosteric inhibition describes the indirect inhibition of a protein's functional activity through the binding of an inhibitor to sites on the target protein other than its active site. In this case, the active site of the spike protein is the S-RBD. Protein function is determined by its shape, also called its conformation, which can be modified through the binding of an inhibitor. The function of the spike protein, namely the ability of the S-RBD to interact with hACE2, is controlled by its ability to shift from a 'down' conformation to an 'up' conformation (Gur, 2020). In the 'down'

conformation, the S-RBD is tucked away from the hACE2 receptor, rendering it inaccessible (Wang, 2021). An allosteric inhibitor could modify the spike protein such that the S-RBD remains in the 'down' conformation, preventing the crucial transition from taking place without binding the S-RBD directly. Previous studies provide evidence to support the existence of allosteric sites on the spike protein. Certain residues have been indicated to be heavily involved in the mobility of the S1 subunit containing the S-RBD (Olotu, 2020). Therefore, these residues are of interest for exploring allostery. Two allosteric sites have been well-mapped and are promising targets for therapeutics.

The first potential allosteric site indicated by the literature is located at the junction of SD1 and SD2 (Wang, 2021) (Figure 9). In this study, we will refer to this site as Site A. The authors of the original study detailing its location showed that fluctuation in rigidity at this site interfered with the ability of the S-RBD to shift conformations. This study conducted virtual screening of drug-like compounds against the site and reported several compounds as having high binding affinity, referred to as CPD1-7, CPD20, and CPD26. In our research, we are interested in validating the performance of CPD1-7, CPD20, and CPD26 against this site and in newly evaluating these compounds against a second allosteric site.

The second potential allosteric site detailed by the literature is located on the S1 subunit, underneath the S-RBD and receptor binding motif (RBM) (Li B. W., 2021) (Figure 9). In this study, we will refer to this site as Site B. The original study identified this site as one of several potential allosteric sites on the spike protein using SiteMap (Schrodinger, Inc.), a method used to identify binding sites (Li B. W., 2021). The sites were filtered by cross-referencing them against

the results of a different binding-site identification software, FTMap. The remaining candidate sites were further filtered based on the stability of their interaction with drug-like compounds in molecular dynamics simulations. Among the remaining sites, Site B was reported to be highly stable. We also chose to explore this site because it does not overlap with Site A.

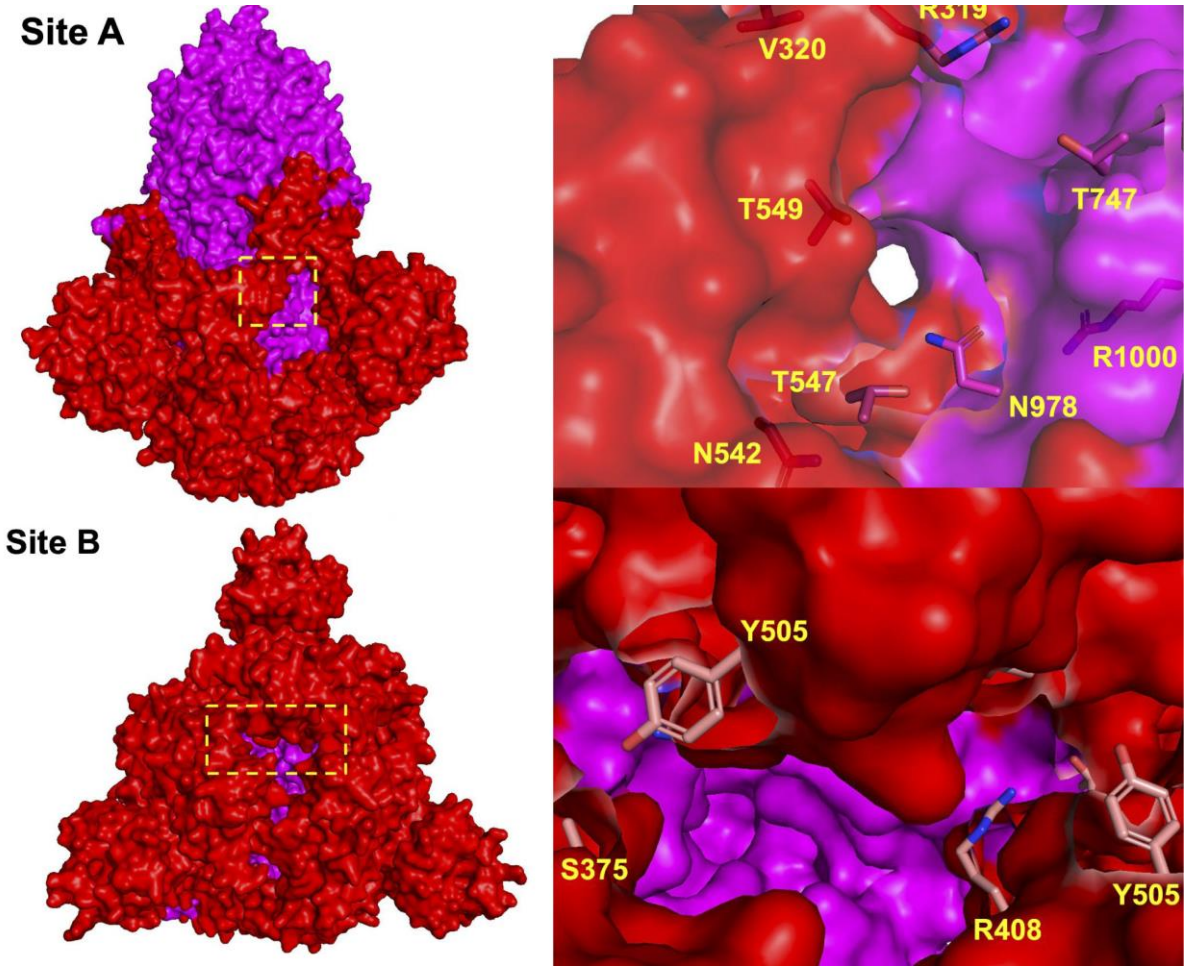

Figure 9. Site A and Site B residues depicted on the spike protein (PDB: 6VXX).

We are interested in assessing the performance of FDA-approved drug-like compounds and previously tested compounds (CPD 1-7, CPD20, and CPD26) against Site A (Wang, 2021) and Site B (Li, 2021). The compounds we will be testing are divided into two compound libraries, as shown in the diagram below (Figure 10). The first set of compounds, which we will refer to as Library 1, consists of 200 compounds. These compounds were selected based on the results of a

previous study unrelated to exploring allostery at Site A or Site B (Zhai, 2022). This study tested several thousand FDA-approved druglike compounds against seven sites on the S1 spike subunit. The authors found that two of the seven sites on S1 were highly druggable. The top 100 performing compounds at each of these two highly druggable sites were used to construct Library 1. We expect these compounds to perform well against Site A and Site B based on their affinities for other locations on the S1 subunit.

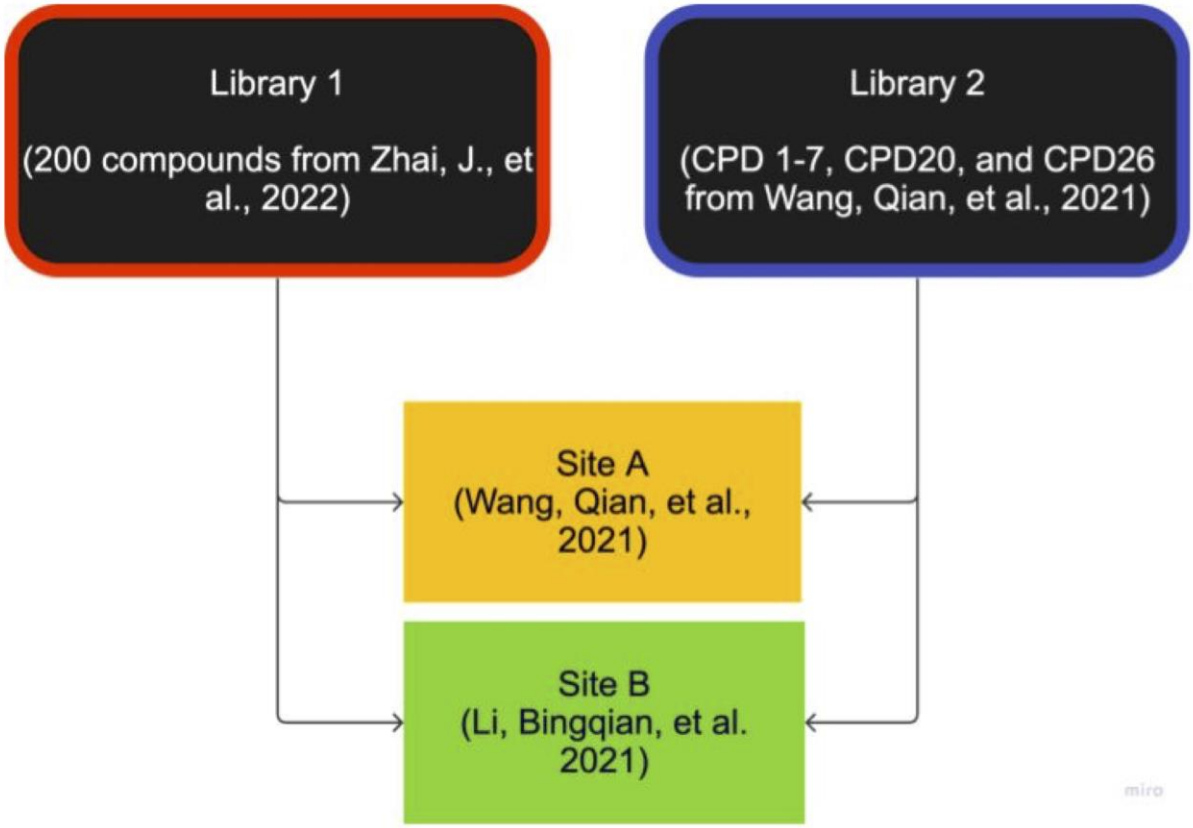

Figure 10. Diagram of the study methodology depicting the two compound libraries and two allosteric sites tested.

The second compound library we will be testing consists of the most potent compounds found to bind Site A as reported by the original study, namely CPD1-7, CPD20, and CPD26 (Wang, 2021). We are interested in validating the performance of these compounds against Site A and in

newly evaluating their performance against Site B. In summary, two libraries of druglike compounds will be tested against two allosteric sites identified in previous studies.

4.2 Method

The 3-D crystal structure of the spike glycoprotein at a resolution of 2.80 Å (PDB ID: 6VXX) was downloaded from the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (Walls, A. et al. 2020). Site A is located at the S1-S2 (Chain B) junction, with grid parameters (Å) x= 40 y= 40 z= 40 and Center (Å) x= 225.178, y= 174.656, z= 213.746. Site B is located between Chain B and Chain C, with grid parameters (Å) x= 40 y= 40 z= 40 and Center (Å) x= 196.983, y= 213.878, z= 256.849.

Two compound libraries were compiled for screening against both sites. Library 1 consists of 200 drug-like compounds. Library 1 was constructed based on the results of Zhai, J., et al., 2022. A supplementary data file (d1cp04736a1.xlsx) reports the compounds with the highest affinity for the two most druggable sites on the S1 subunit, referenced in the study as Clusters 1 and 4 (Zhai, 2022). The top 100 compounds from Clusters 1 and 4 were downloaded from the ZINC database in sdf (structure data file) format and converted to pdbqt format to be compatible with the requirements for docking software.

Library 2 consists of compounds CPD1-7, CPD20, and CPD26, tested by Wang, Qian, et al., 2021. These compounds were not directly downloadable from a chemical database. OpenBabel (N M O'Boyle, 2011), a chemical file format conversion software, was used to translate the chemical formulas of these compounds to pdbqt format.

Pre-processing for virtual screening was conducted using AutodockTools (Morris, 2009). AutodockTools was used to remove waters and inorganic atoms from the pdbqt files, as well as to calculate appropriate charges. Docking was performed using Autodock Vina (Trott, 2010). Each compound was docked for 100 iterations at both Site A and Site B.

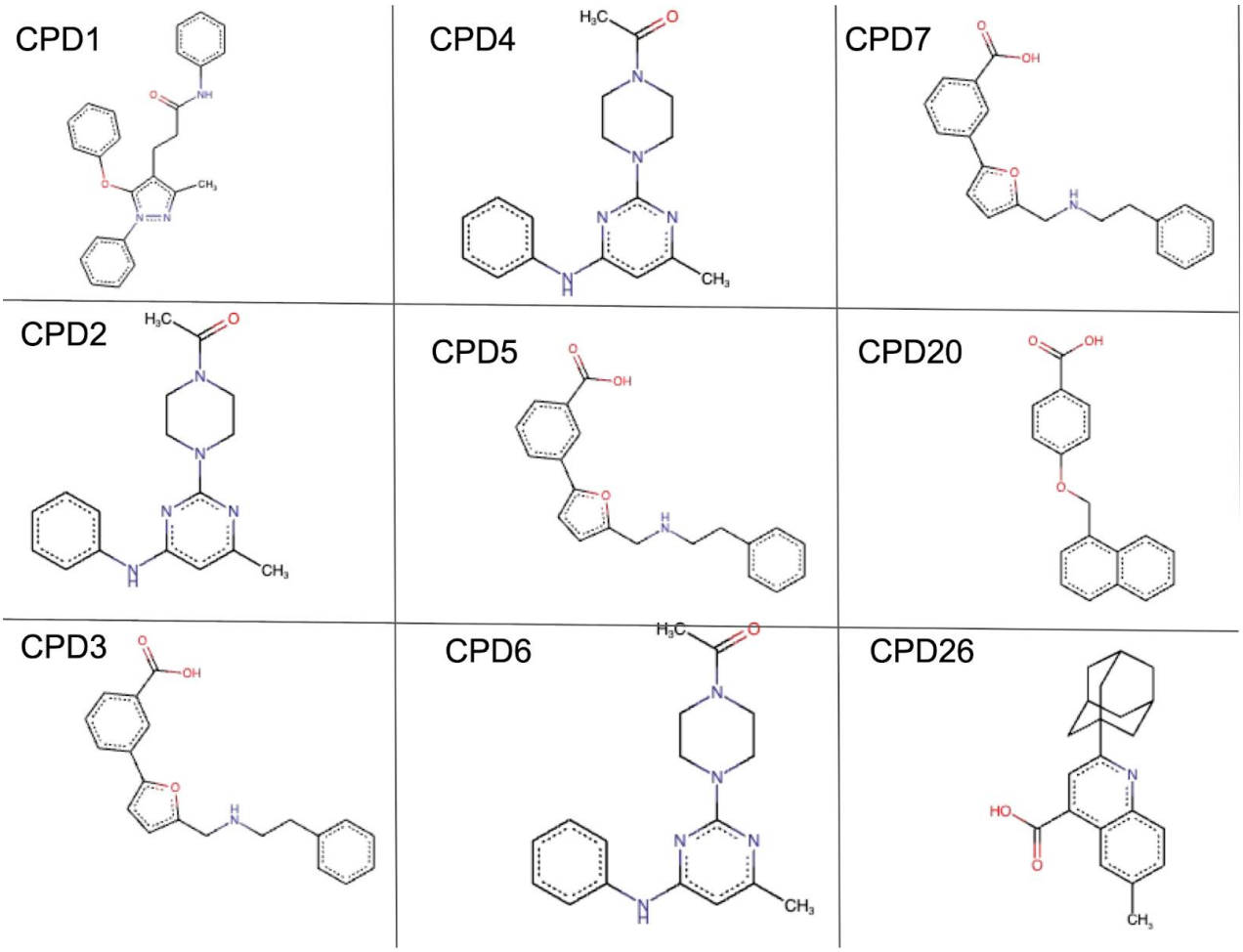

Table 2. Identifying information for CPD1-7, CPD20, and CPD26.

The top 10 compounds with the highest mean binding affinity at each site were filtered for further visual analysis using PyMol (Schrödinger, LLC). All BASH scripts created to automate virtual screening and preprocessing are included in Appendix A.

4.3 Results and Discussion

Our results will be divided into two parts, as depicted in Figure 11. In the first section, we will discuss the performance of Library 1 at Site A and Site B. The second section will discuss the performance of Library 2 and will include a comparison of our results at Site A to the results of the original study which detailed its location (Wang, 2021).

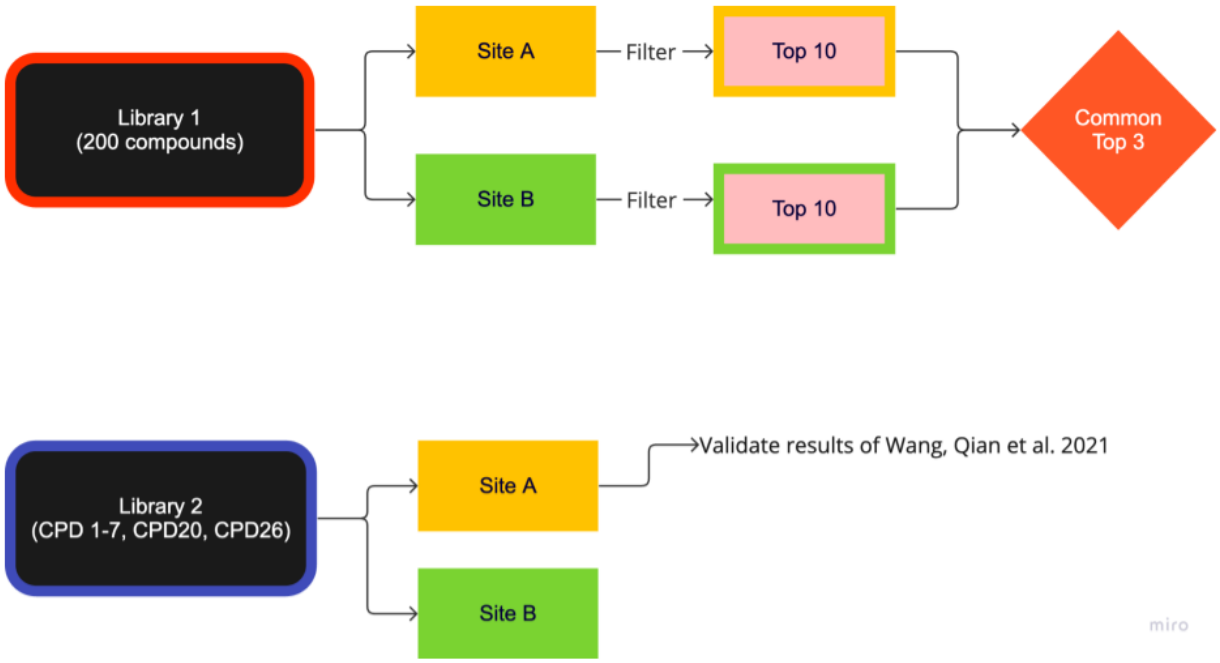

Figure 11. Illustration of the process used to analyze each of the compound libraries tested in this study.

We will begin by discussing the distribution of binding affinity for all iterations of the 200 compounds in Library 1 at Site A and Site B (Figure 12). Library 1 had an average binding affinity of -7 kcal/mol across all tested compounds against both sites. The distribution of Site B shows a slight skew towards higher binding affinity compared to the distribution of Site A.

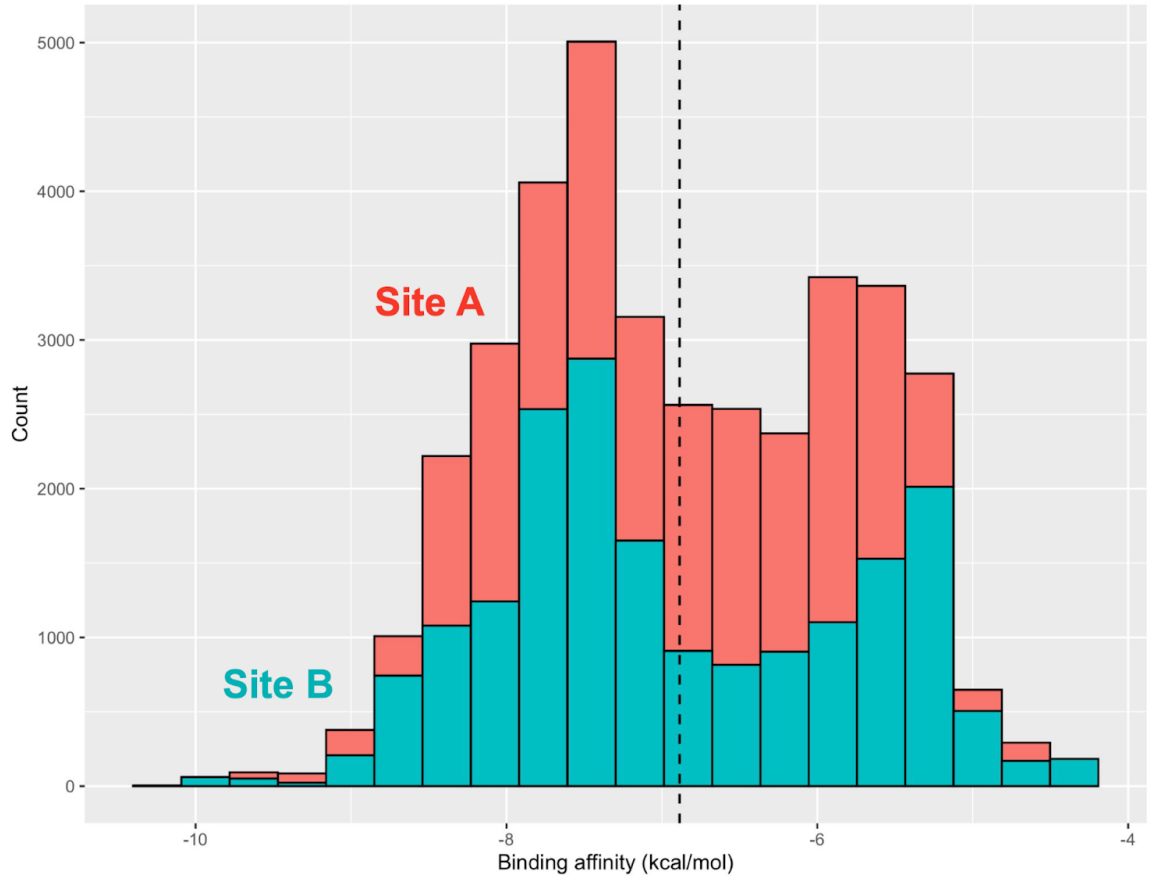

Figure 12. Comparison of binding affinity distributions for all 200 compounds in Library 1 (all 100 iterations included) tested against Site A and Site B.

Next, the top 10 lead compounds were filtered from both sites. Compounds were selected based on their median binding affinity. The best binding affinity scores of the lead compounds are -9.7 kcal/mol for Site A and -9.0 kcal/mol for Site B (Figure 13). In general, the median binding affinity of the top 10 lead compounds was similar for both sites, ranging between -8 to -9 kcal/mol. However, the lead compounds from Site B displayed decreased variance in binding affinity compared to those from Site A, suggesting a higher degree of stability of the receptor-ligand complex at Site B.

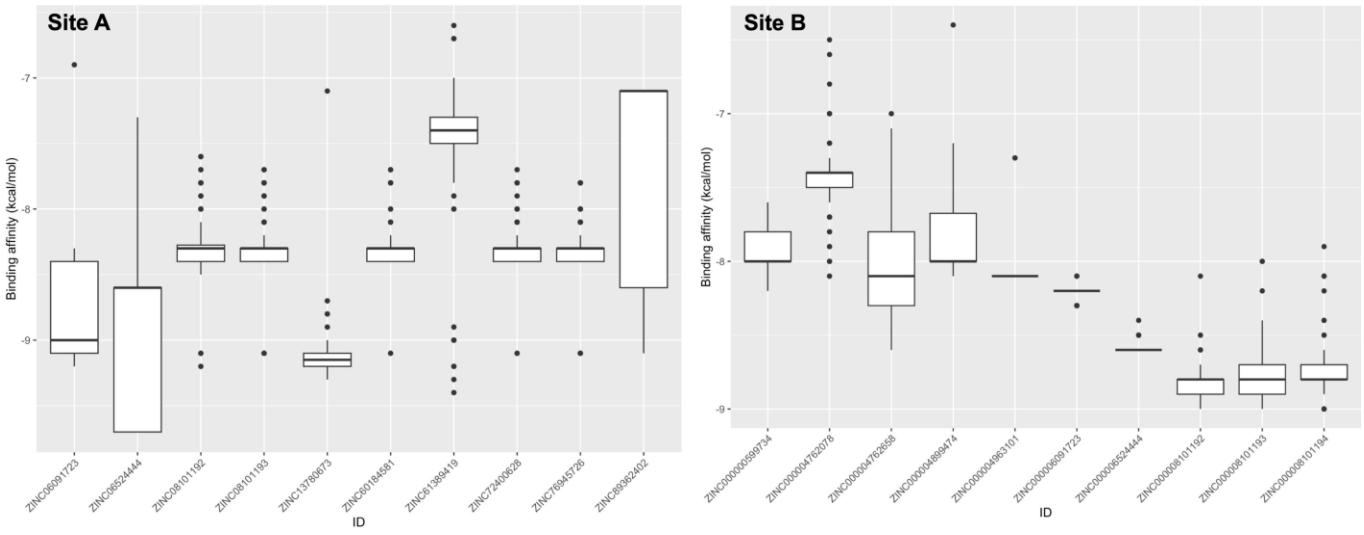

Figure 13. Binding affinity distributions for the top 10 compounds with the best binding affinity scores for Site A and Site B.

Three compounds were found to be common to the top 10 compound list for both Site A and Site B: ZINC06524444, ZINC08101192, and ZINC08101193 (Figure 14). These three compounds had increased binding affinity and decreased variance at Site B. ZINC06524444 showed the most notable difference in variance between the sites. These compounds were selected for further individual analysis, which we will discuss in a later section.

In addition to viewing the binding affinity distributions of the top 10 lead compounds at each site, we are interested in understanding the interactions these compounds make when bound to the spike protein. To visualize the binding conformations of the lead compounds at each site, the most common ligand-receptor complexes of the top 10 lead compounds were overlaid in Pymol in order to visualize common modes of binding at Site A and Site B. At Site A, two clusters of binding at Site A were identified (Figure 15-16).

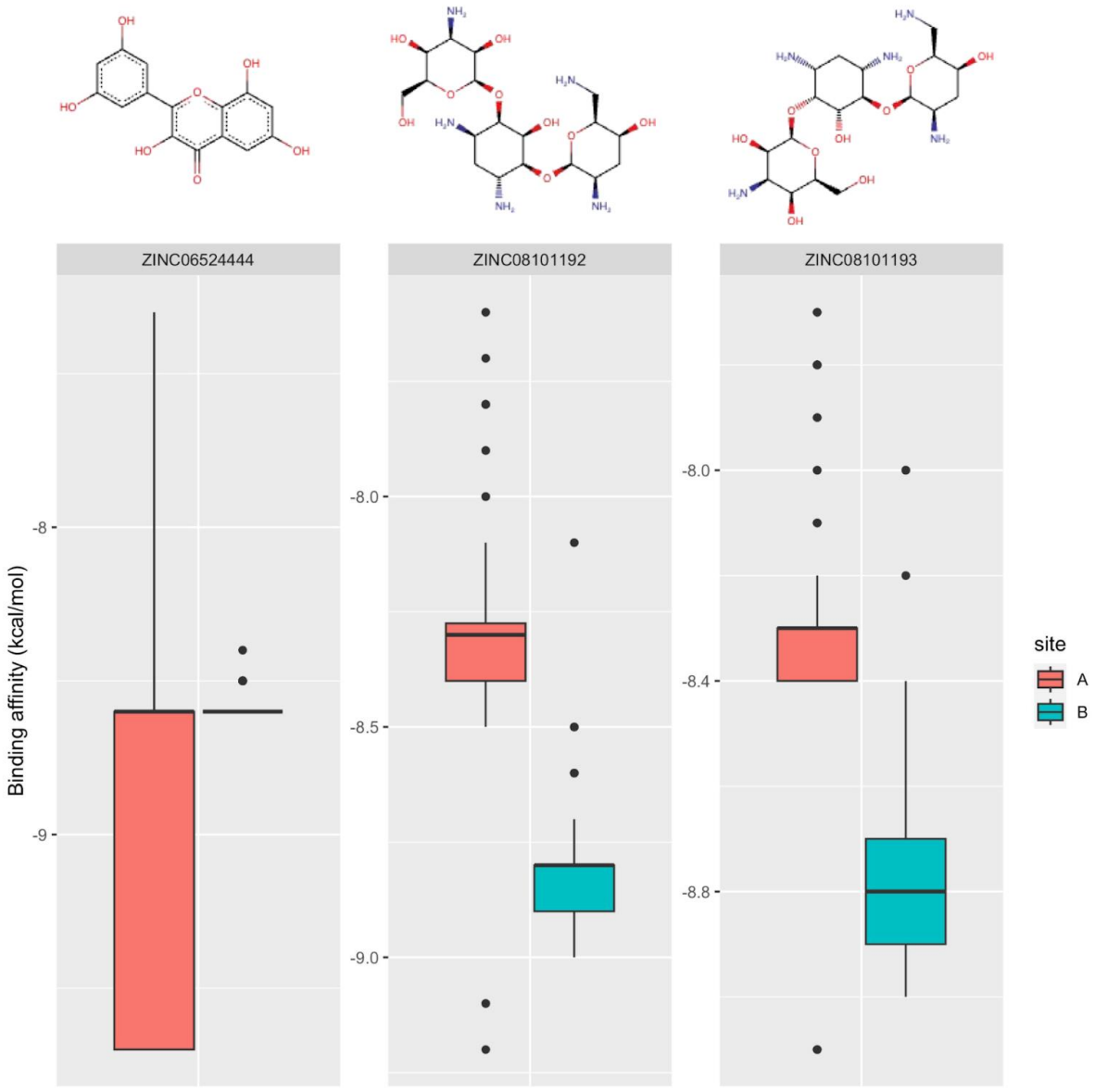

Figure 14. Comparison of binding affinity distributions for the three compounds common to the top 10 lists at both Site A and Site B.

The most common binding pocket at Site A is the least exposed, denoted Cluster 1 (Figure 15). Cluster 1 is tucked inside the S protein closer to HR1 and is covered by CTD1 and CTD2. The positioning of Cluster 1 may indicate its ability to inhibit the function of CTD1 and CTD2. The change in conformation the CTDs undergo during fusion is crucial for cleaving the spike

glycoprotein into the S1 and S2 subunits during fusion (Jackson, C. B. et al. 2022). The binding conformations of these compounds are also positioned to potentially disrupt the movement of HR1. HR1 interacts with HR2 during fusion, resulting in shortening the distance between the viral proteins and the host cell enough to allow for entry (Xia, S. et al. 2020). HR1 has been suggested as a potential target for fusion inhibitors, as the HR1 and HR2 regions of the S2 protein tend to be conserved (Xia, S. et al. 2020).

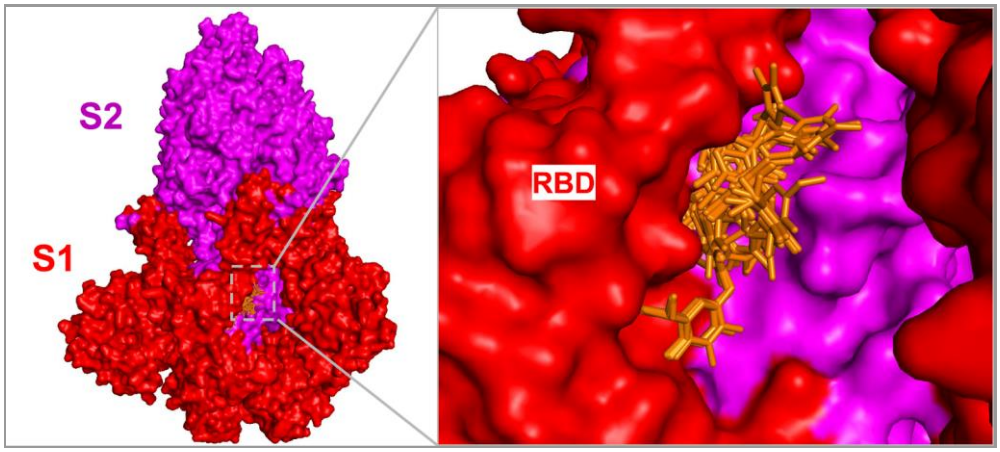

Figure 15. The binding pocket at Site A is situated between the S1 (red) and S2 (magenta) subunits. Bound ligands depicted in orange.

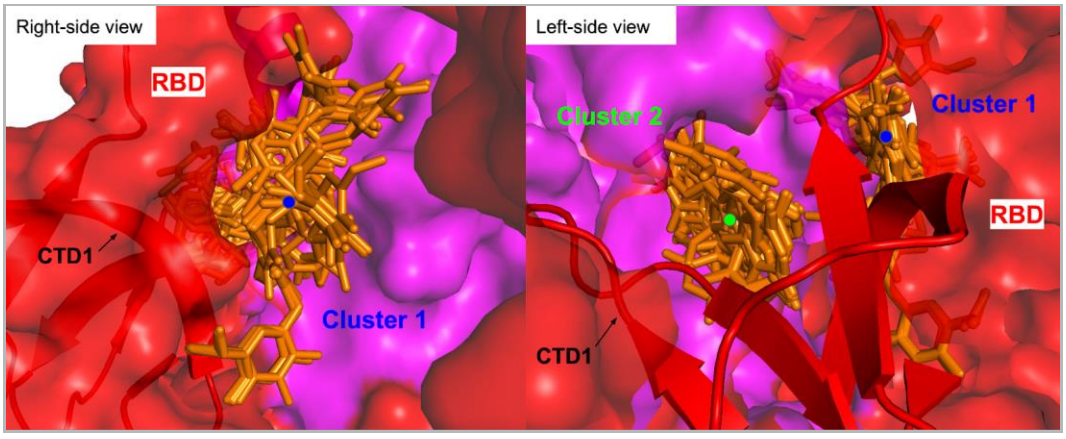

Figure 16. Close-up of the two most common binding clusters at Site A.

At Site B, overlaying the most common binding modes of the top 10 compounds revealed two clusters of binding that are situated between the S2 core and the S1 RBD and RBM (Figure 17-18). The compounds interact with residues R408, D405, R403, and Y505 of the RBM. Residues R408 and D405 are involved in shifting the RBD from the 'down' to 'up' state. Residue R403 has been shown to affect binding to ACE-2, (Zech, 2021) and Y505 is a significant mutation found in variants, such as Omicron (Alkhatib, 2021).

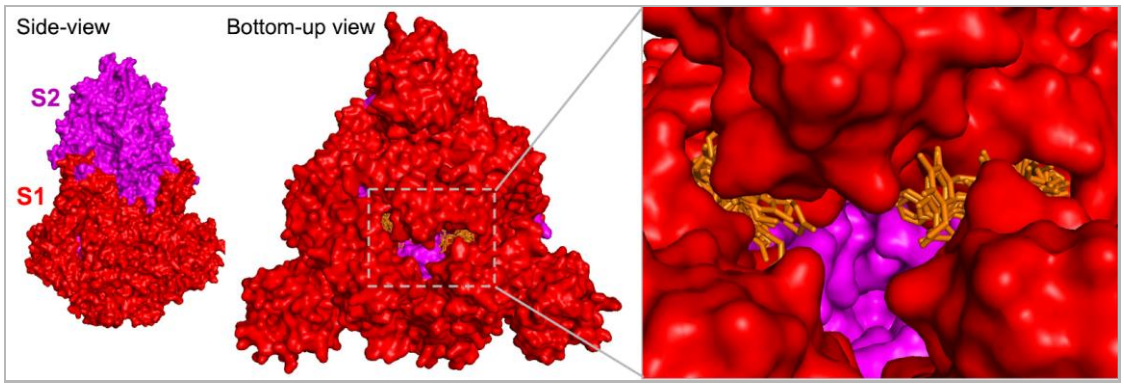

Figure 17. Common binding clusters at Site B (orange) situated in the junction of the S1 (red) and S2 (magenta) subunits.

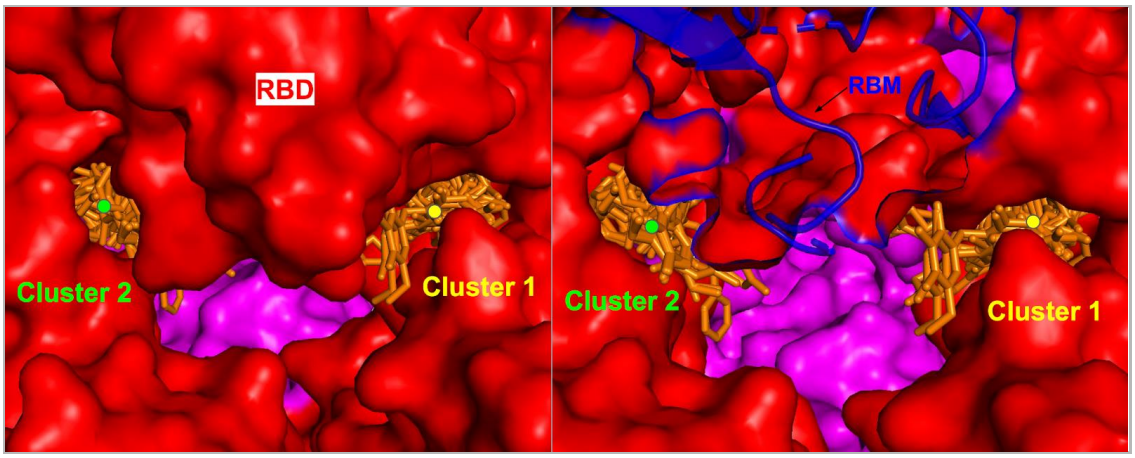

Figure 18. Close-up of the two most common binding clusters at Site B (orange).

We will now discuss the binding conformations of the three lead compounds common to both allosteric sites. The first compound is ZINC06524444.

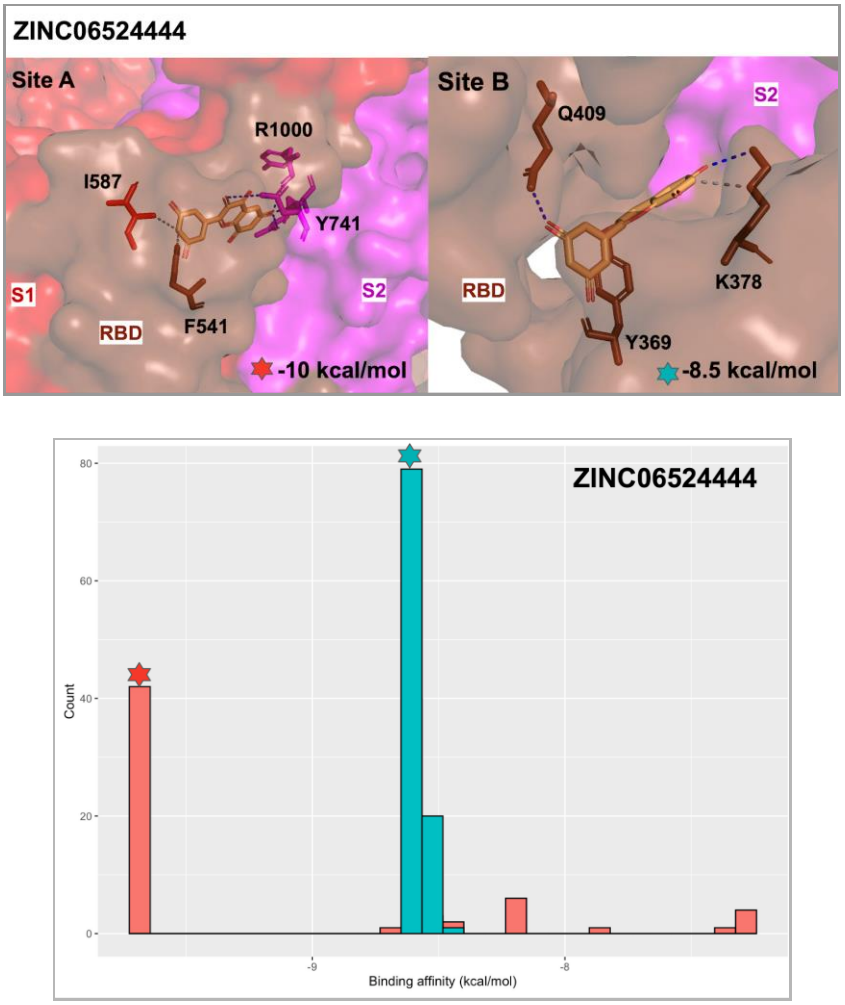

Figure 19. Most common binding pose and binding distributions of ZINC06524444 at Site A (red) and Site B (blue).

ZINC06524444 made a higher number of interactions with residues at Site A, including an additional hydrogen bond with residues on the S2 side of the interface, with D745 and Y741 of the α1/α2 helix region and R1000 of the central helix (CH). This region has been suspected to interact with host protease TMPRSS2 during fusion and entry (Vankadari, 2022).

ZINC06524444 has a slightly higher binding affinity for Site A, however this compound binds Site B with more consistency (Figure 19).

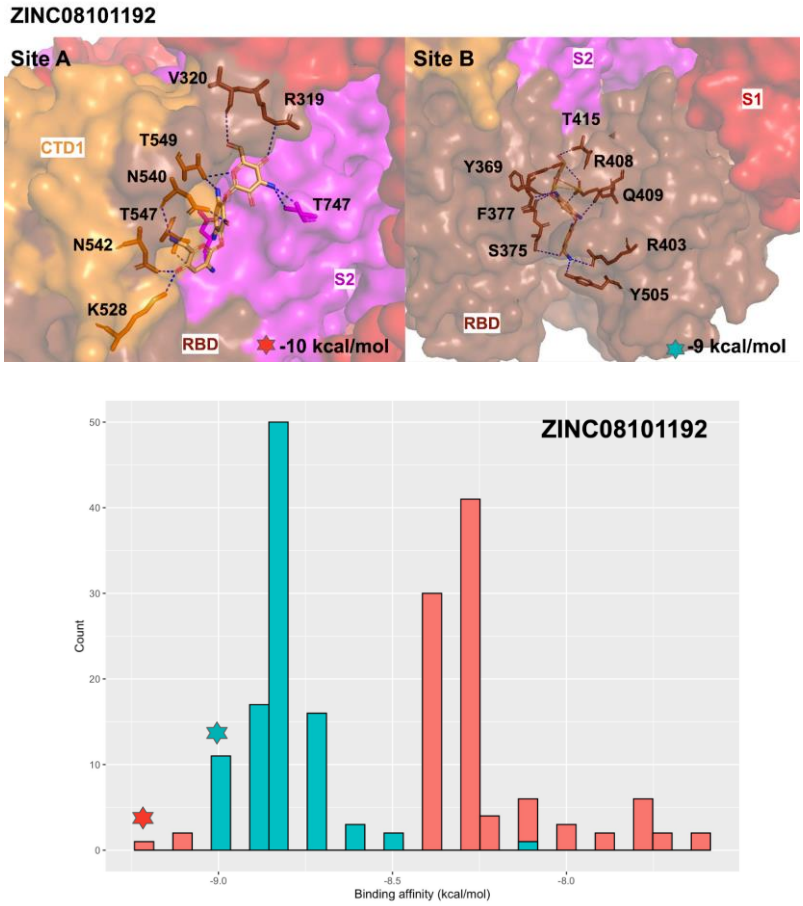

Figure 20. Most common binding pose and binding distributions of ZINC08101192 at Site A (red) and Site B (blue).

The second lead compound common to both sites is ZINC08101192. ZINC08101192 had approximately the same mean binding affinity at Site A and Site B, with a slight skew in the distribution towards higher binding affinity at Site B (Figure 20). The conformation at Site A is stabilized by nine hydrogen bonds. Both sites form a salt bridge (Site A, R319 and Site B, R408), but the conformation at Site B forms an additional pi-cation bridge with R408.

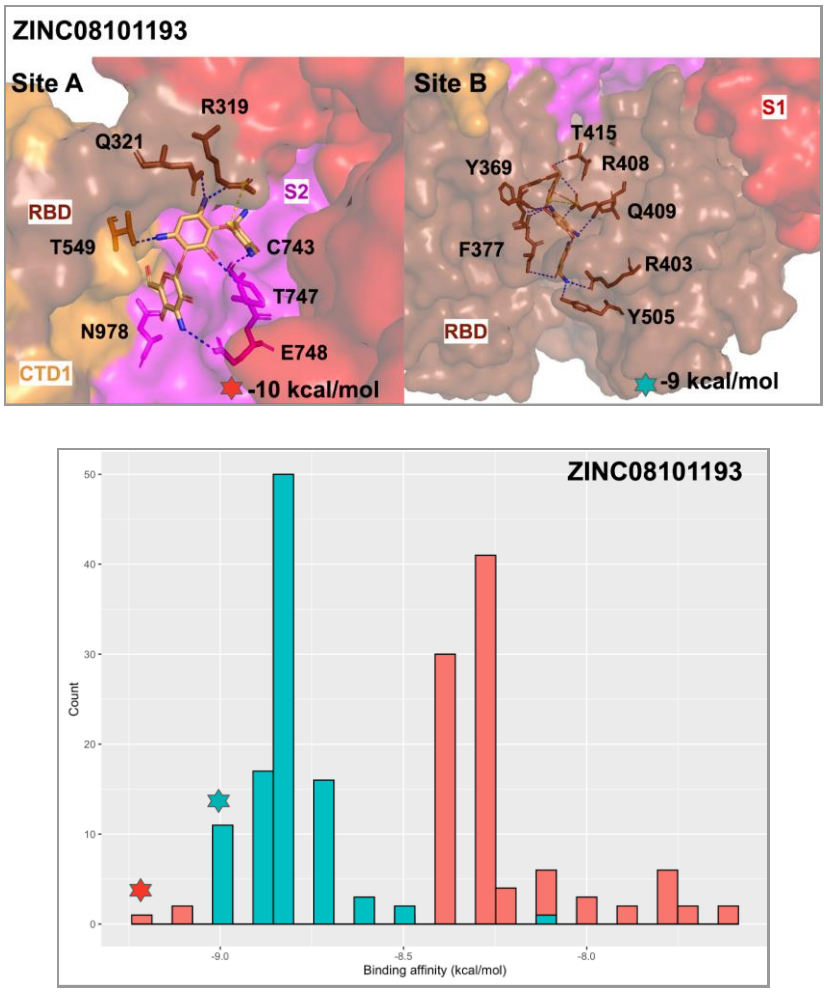

Figure 21. Most common binding pose and binding distributions of ZINC08101193 at Site A (red) and Site B (blue).

The increased number and strength of bonds could be responsible for the better performance at Site B. The binding pocket at Site B also has the ligand more closely packed in the S1 subunit, which may allow a greater number of interactions to form, compared to the density of the pocket at Site A, which straddles the gap between the subunits.

The final common lead compound is ZINC08101193. Similar to ZINC08101192, ZINC08101193 had slightly better performance at Site B. The number of hydrogen bonds

formed at Site B was higher than the number at Site A (Figure 21). However, since the residues at Site B are more closely packed together, there may be more stable conformations possible within the binding pocket.

We will now transition to discuss the performance of Library 2 against both allosteric sites, beginning with Site A. Most of the compounds in Library 2 had significantly different binding distributions at this site. CPD7, CPD3, and CPD4 did not differ significantly from each other, as well as CPD7 and CPD1 to a lesser degree, likely due to these ligands having a larger amount of variance in their binding distributions. CPD2 appears to have performed the best at Site A, followed by CPD20 (Figure 22).

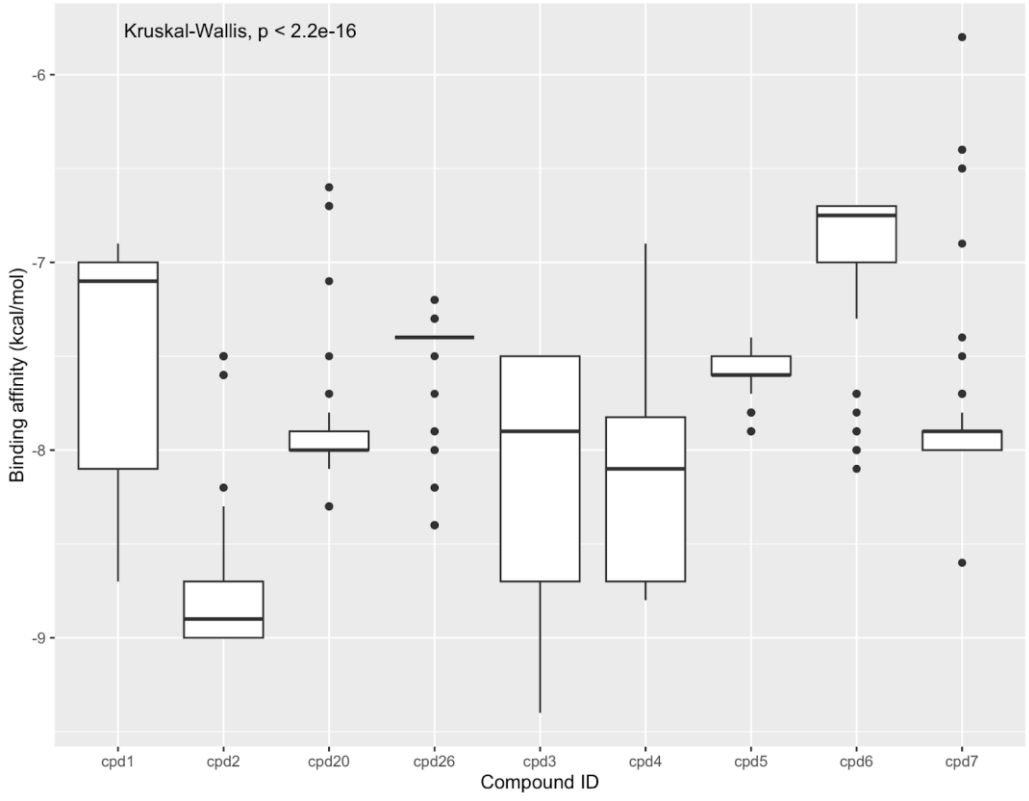

Figure 22. Binding affinity for ligands CPD1-7, CPD20, and CPD26 at Site A.

We will look individually at the performance of CPD7, CPD20, and CPD26 at Site A, as these were the compounds tested in the original study with expected binding poses reported (Wang, Qian, et al., 2021). We will compare the expected results from this study with our results for each of these three compounds, beginning with CPD7.

The two most common binding conformations of CPD7 accounted for 75% of the total number of runs (Conf1 - 39%, Conf2 - 36%) (Figure 23). The exact expected predicted binding pose was not observed in any of the runs, however, Conf2 is similar in orientation to the expected binding pose. In the first conformation, the benzene ring faces the S1 subunit and the carbonyl group interacts with residues from the S2 subunit, however in this confirmation, CPD7 forms a salt bridge instead of hydrogen bonds with Arg1000. The structure is additionally stabilized on the S2 side by hydrogen bonding with Met740 and Gly744. Instead of the nitrogen group bonding with Asp745, it forms a hydrogen bond with Thr573, which likely orients the benzene ring to hydrophobically interact with Pro589.

In the second conformation, the benzene group faces the S2 subunit and the carbonyl groups interact with both subunits, forming two hydrogen bonds with S1 residue Thr573 and one hydrogen bond with S2 residue Asn978. The nitrogen group hydrogen bonds with Asn856 of the S2 subunit and the benzene ring hydrophobically interacts with S2 residues Leu977 and Val976. Overall, the two conformations of CPD7 performed similarly in terms of frequency and binding affinity.

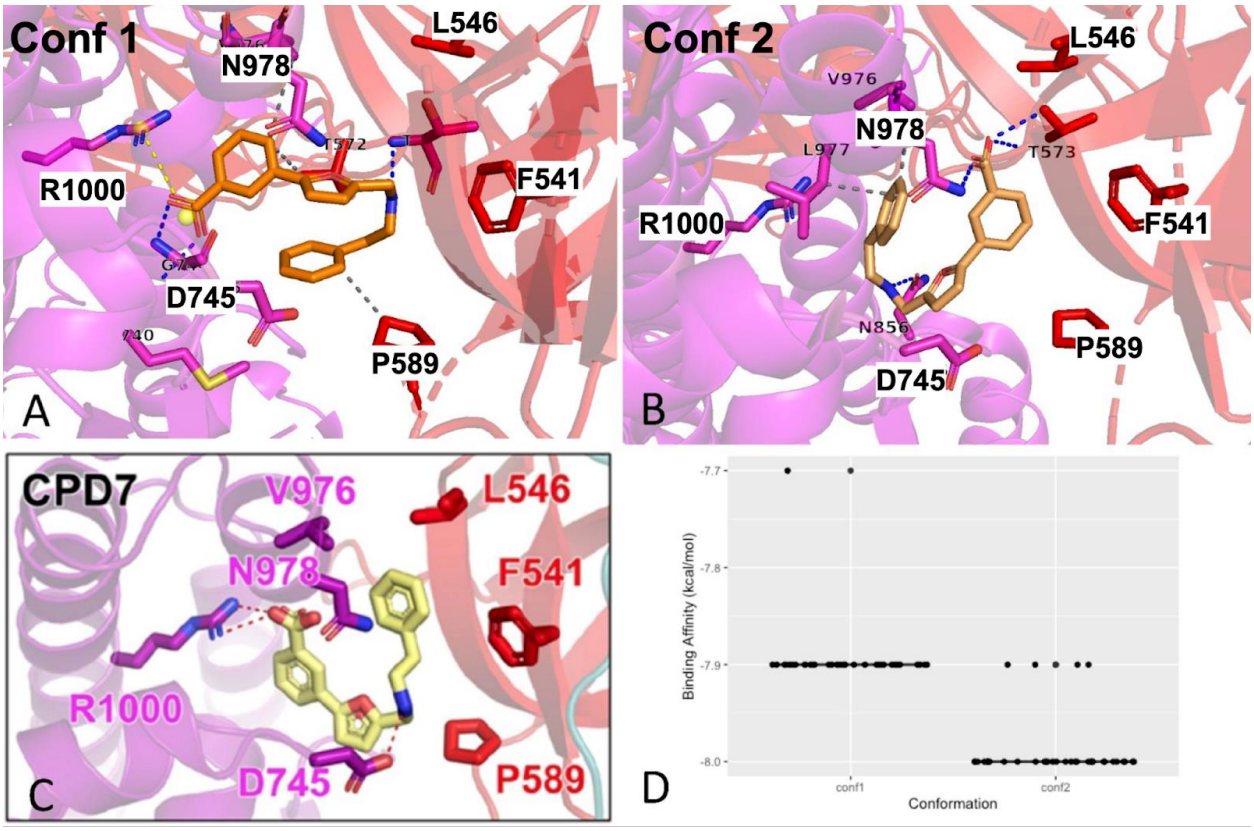

Figure 23. Most common conformations of CPD7 (A: Conf1, B: Conf2). (C) Expected binding pose for CPD7 (image from Wang, 2021). (D) Distribution of binding affinity for conformations 1 and 2.

CPD20 was expected to bind with its aromatic rings facing the S1 subunit and its carbonyl groups interacting with S2 residue Arg1000. The two most common conformations account for 76% of the total runs, with Conf2 occurring over twice as frequently as Conf1 (Conf1 - 20%, Conf2 - 56%) (Figure 24). Conf1 is similar to the expected conformation, oriented in the predicted direction, but forming a salt bridge with Arg1000 instead of hydrogen bonds, and interacting additionally with Ile742 and Gly744 of the S2 subunit.

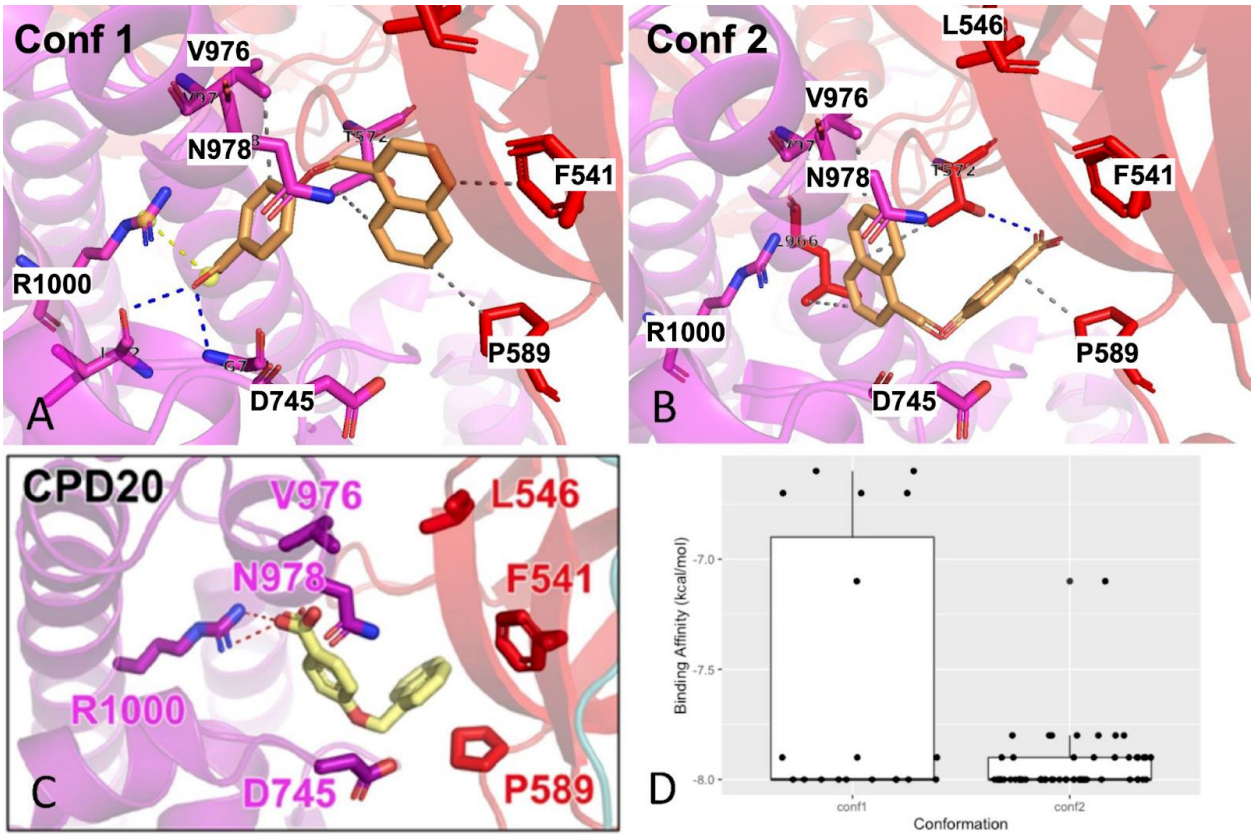

Figure 24. Most common conformations of CPD20 (A: Conf1, B: Conf2). (C) Expected binding pose for CPD20 (image from Wang, 2021). (D) Distribution of binding affinity for conformations 1 and 2.

The aromatic rings were primarily stabilized through hydrophobic interactions with S1 residues Phe541, Pro589, and Thr572. In the second more common conformation, the aromatic rings face the S2 subunit and the carbonyl groups only make one hydrogen bond with Thr572. The aromatic rings make hydrophobic interactions with S2 residues Leu966 and Val976. The binding affinity distribution is not significantly different between the two conformations, although Conf1 has a larger range of binding affinity.

CPD26 was expected to form two hydrogen bonds with Arg1000. The two most common conformations account for 17% of the total runs; most of the remaining poses did not bind within

the predicted pocket (Figure 25). Conf2 was observed in only four runs. The carbonyl groups of this conformation did not interact with S2 residue Arg1000, but with Gly744, Leu977, and Asn978.

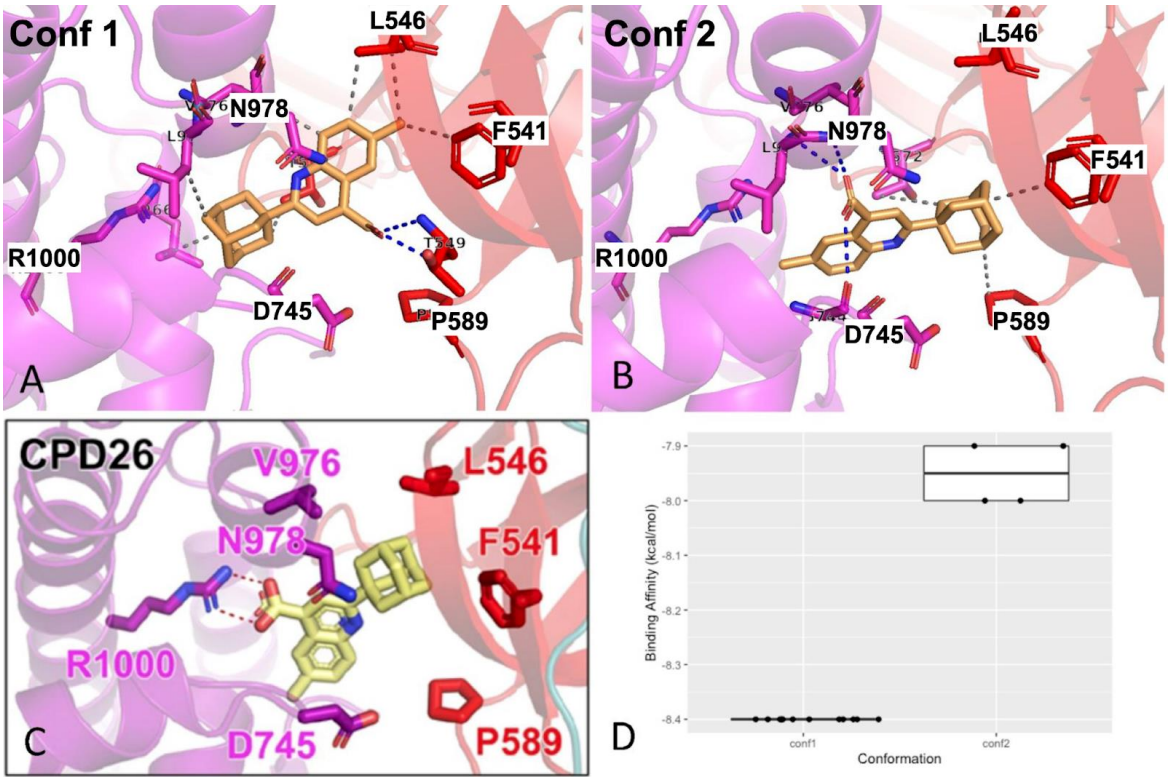

Figure 25. Most common conformations of CPD26 (A: Conf1, B: Conf2). (C) Expected binding pose for CPD26 (image from Wang, 2021). (D) Distribution of binding affinity for conformations 1 and 2.

The more common conformation, Conf1, most like the expected pose, was observed in 13% of the runs. This structure is supported by two hydrogen bonds with S1 residue Thr549 and hydrophobic interactions on both the S1 and S2 subunits.

Lastly, we will discuss the performance of Library 2 at Site B. At Site B, all compounds performed significantly differently from each other in terms of binding affinity.

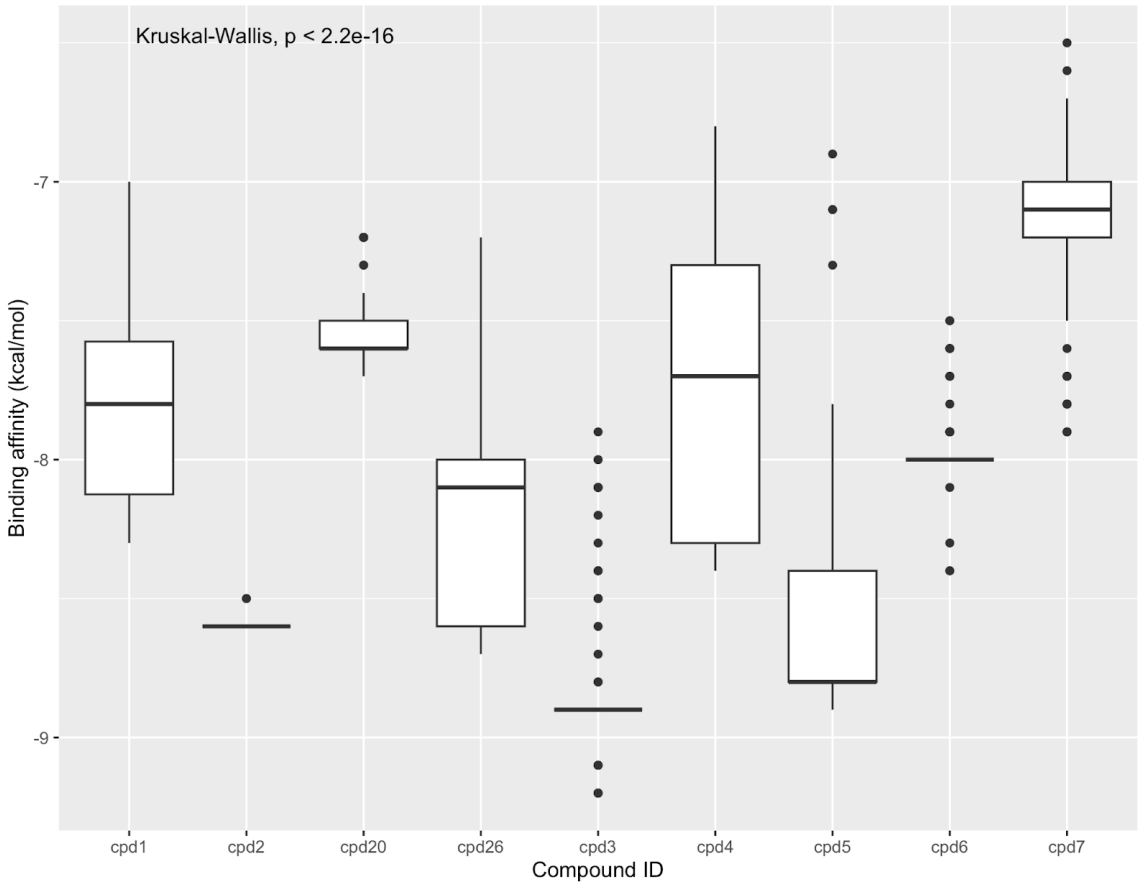

Figure 26. Binding affinity distributions for ligands CPD1-7, CPD20, CPD26 at Site B.

Although CPD3 has the highest average binding affinity, CPD2 is overall the best performing lead, in terms of binding affinity and variance. CPD1, CPD3, and CPD4 also again showed high amounts of variance relative to the group. CPD7 performed more poorly at this site, while CPD3, CPD5, and CPD2 had the best median binding affinities (Figure 26).

A comparison of the performance of Library 2 at Site A and Site B revealed notable differences. All compounds performed significantly differently at the two tested binding sites (Figure 27). All distributions fell within a range of -7.0 to -9.0 kcal/mol, the exception being CPD7 which fell below -6 kcal/mol at Site A.

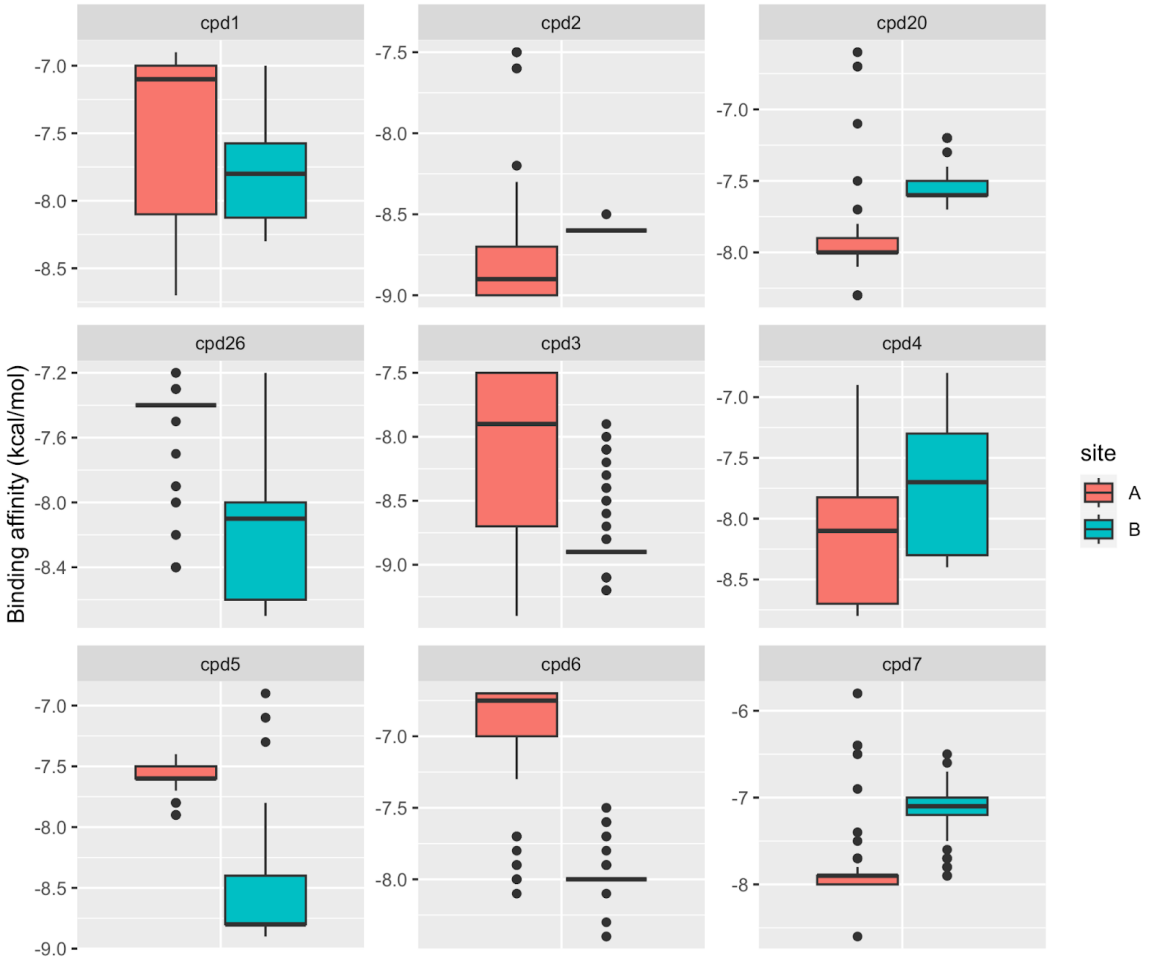

Figure 27. Binding affinity comparisons for ligands CPD1-7, CPD20, and CPD26 at both binding sites.

CPD1, CPD2, CPD4, and CPD20 had the most similar performance between sites, with an approximate difference in average binding affinity of 0.5 kcal/mol. CPD26, CPD3, CPD5, and CPD6 showed a preference for Site B, with a difference in average binding affinity larger than 0.5 kcal/mol. The only compound that showed better performance at Site A was CPD7.

4.4 Conclusion

In this research, we tested the ability of two libraries of druglike compounds to effectively bind allosteric sites, to weaken the interaction between the Spike protein and hACE2 protein.

Interactions with residues that make up the CTD and HR proteins, at Site A, and residues R408 and D405 at Site B, are indicated to affect the pre-fusion state of the spike glycoprotein. Compounds from Library 1 bound both Site A and Site B effectively, with a slight preference for Site B, likely due to the increased number of contacts possible within the binding pocket. Three candidates had high performance at both sites, ZINC06524444, ZINC08101192, and ZINC08101193, with the latter two showing preference for binding Site B.

Library 2, consisting of compounds CPD1-7, CPD20, and CPD26, bound both sites with moderate binding affinity. The binding affinities generally ranged from -7.0 to -9.0 kcal/mol. CPD7 performed the least well at both sites, with some conformations having a binding affinity less than -6 kcal/mol. CPD1, CPD3, and CPD4 consistently had high variation in performance at both sites. In general, the tested compounds had a higher affinity for Site B than for Site A, particularly CPD26, CPD5, and CPD6. The compound that performed the best overall with the least variation between sites was CPD2.

In examining the most common conformations of CPD7, CPD20, and CPD26 against their expected binding poses, none of these compounds bound Site A exactly as predicted in the original study (Wang, Qian, et al. 2021). This is expected behavior given the stochasticity inherent in molecular docking. Similar binding poses to the expected were observed, in terms of orientation of the ligand and interacting residues, however these conformations were not the most frequent for CPD20 and CPD26. However, the most frequently observed pose for CPD7 was also the conformation most similar to the predicted pose. The three compounds were expected to form hydrogen bonds with S2 residue Arg1000, and this interaction was observed in

at least one conformation for CPD7 and CPD20. The ligands commonly formed bonds with more than one residue, particularly S2 residue Gly744. In general, the results support the conformations found in the original study (Wang, 2021).

In this study, we identified ten FDA-approved drug-like compounds that bind each of the specified allosteric sites with the high affinity and stability. In addition, we identify three compounds that are top performers at both sites and characterize their common binding conformations and binding locations on the spike protein. We were also able to compare our results at Site A against those of the original study. Lastly, we showed that between the two allosteric sites, the compounds with highest binding affinity and least amount of variability tended to show a preference for Site B.

5. Identifying Inhibitors of the Wildtype Spike Receptor Binding Domain from FDA-Approved Drugs and Evaluating Comparative Performance in SARS-CoV-2 Variants

5.1 Introduction

Since its emergence in 2019, the SARS-CoV-2 virus has evolved into many distinct lineages. The virus mutates at a high rate, approximated to 0.44 substitutions per week (Amicone, 2022). Although the majority of the mutations are neutral or mildly deleterious, the small percentage of high-impact mutations that increase the pathogenicity of the virus are cause for urgent concern. These high-impact mutations cause changes to the virus phenotype that increase its evolutionary fitness, such improved transmissibility or antigenicity (Harvey, SARS-CoV-2 variants, spike mutations and immune escape. , 2021). An effective drug candidate will need to have a high binding affinity and stability of binding against all variant spike proteins. Drug candidates need to be continually evaluated against new variants as the virus evolves.

The active site of the spike protein has a particularly high mutation rate in the genome, making it a challenging target for drug design (Liu, 2022). The spike protein is a promising drug target due to its critical role in modulating viral entry through its interaction with the human host cell receptor ACE-2 (Day, 2021). However, the crucial nature of the spike protein to viral function is also what causes it to be under high evolutionary selective pressure. Variants possess mutations to the receptor binding domain of the S1 spike subunit that have been experimentally shown to increase affinity for the ACE-2 receptor (Starr, 2022). Five variants of SARS-CoV-2 have proved to be the most dominant and concerning across the globe, each containing high impact mutations in the spike RBD. These variants are B.1.1.7 (UK), B.1.351 (South Africa), P.1.

(Brazil), B.1.617.2 (Delta), and B.1.1.529 (Omicron). The area of the receptor binding domain that makes direct contact with ACE-2, is called the receptor-binding motif (RBM) (Figure 28). This area has been speculated to be under the most selective pressure. The frequency of mutation in the RBM of the most recent variant, the Omicron variant, has been reported to be approximately 760 folds higher than the mutation rate of the entire spike genome (Kim, 2021).

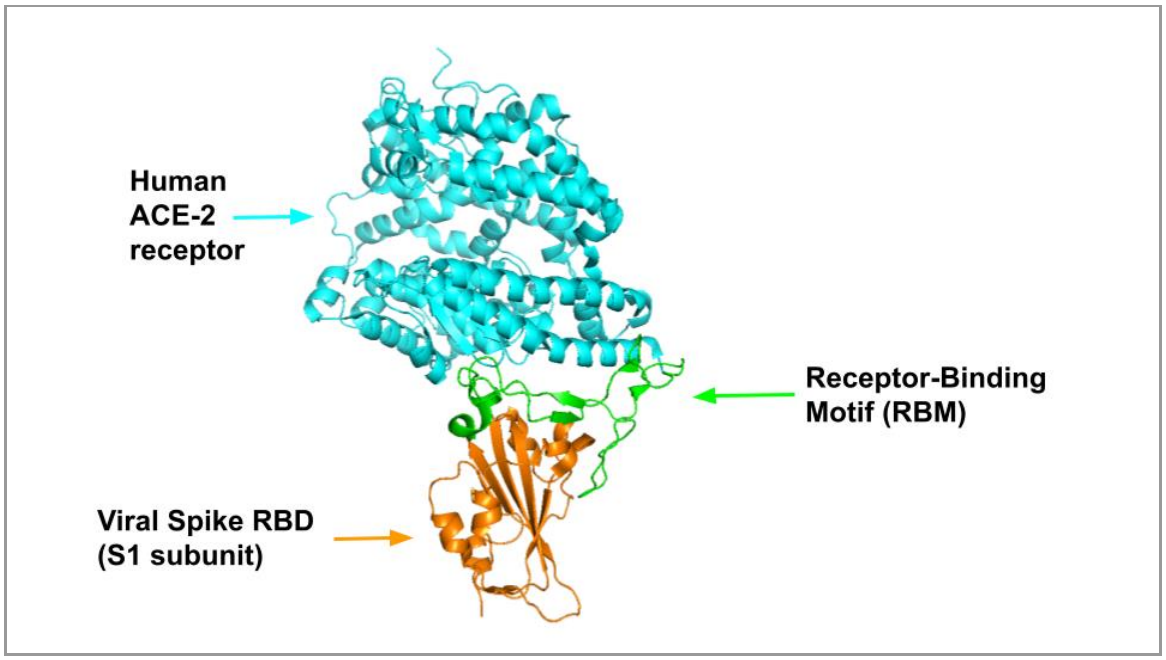

Figure 28. Illustration depicting the binding of the spike protein receptor binding domain (orange), which includes the receptor-binding motif (green), to the host ACE-2 receptor (cyan) (PDB id: 6M0J).

The first significant strain to branch from the wildtype was the alpha variant, or B.1.1.7, which was first in the U.K in September 2020. The B.1.1.7 variant has one significant mutation in the RBD region, N501Y (Liu Y. L., 2021) (Figure 29). The subsequent variants, the beta variant B.1.351, originating in South Africa, and the gamma variant P.1. originating in Brazil, contain two additional mutations in the RBD, E484K and K417T (South Africa)/K417N (Brazil). The N501Y mutation evolved convergently in the UK, Brazil, and South Africa (Martin, D. P., 2021), and improves the affinity of the spike protein for ACE-2, likely by interacting with ACE2

residue 353 (Socher, 2021). A study comparing the ability of mutant spike RBDs to bind ACE-2 in vitro found that RBDs containing the N501Y mutation showed a four-fold increase in binding affinity. The N501Y mutation was also correlated with an increased number of strong bonds between RBD and ACE-2 residues, specifically in the form of pi and cation bonds (Tian, 2021).

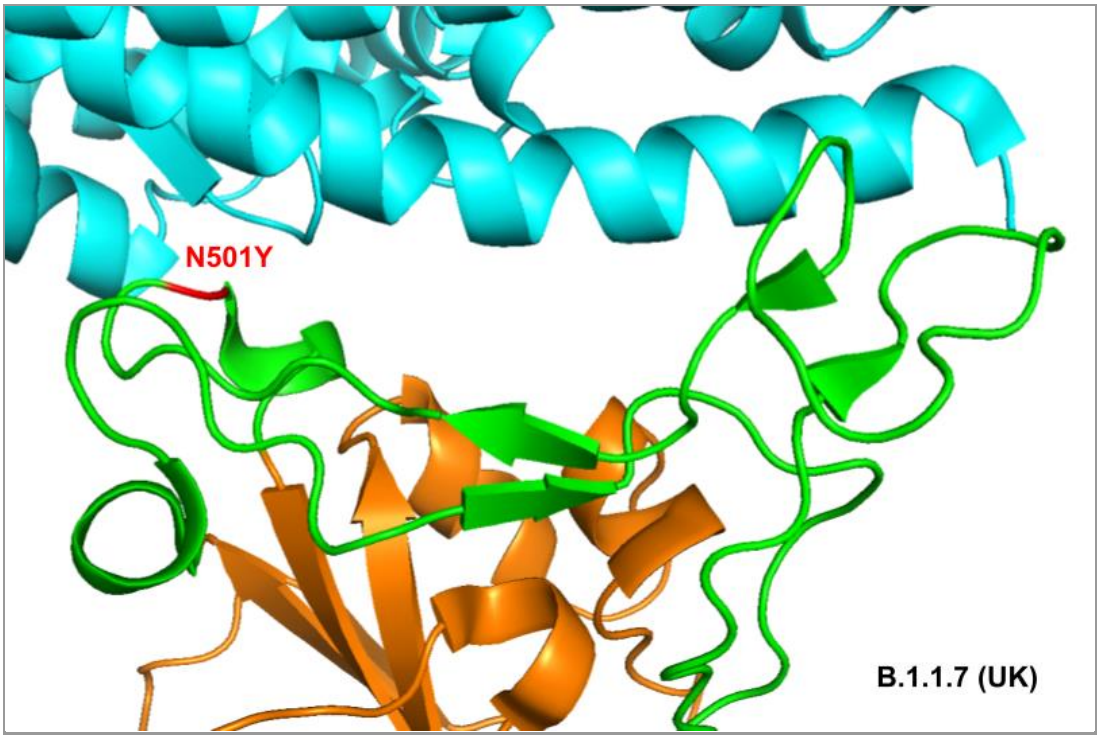

Figure 29. Illustration of the N501Y mutation in the RBD of the B.1.1.7 (U.K.) variant. ACE-2 receptor colored blue; RBD colored orange; RBM colored green.

The B.1.351 variant was first detected in South Africa in March 2020. This variant shares two of its three mutated RBD residues with the P.1. variant. The P.1. variant originated in Brazil and was first detected in February 2020. The B.1.351 and P.1. variants both contain the N501Y mutation, similar to the B.1.1.7 strain (Figure 30). These two variants also share the E484K mutation in the RBM. The E484K substitution has been indicated to confer resistance to monoclonal antibodies (Chen, 2021). The increased immune resistance conferred by this mutation has been suggested to increase the risk of reinfection. The third mutation in the RBD

region of the B.1.351 and P.1. variants is at residue 417. There is a slight difference in this mutation between the two variants. The B.1.351 variant substitutes lysine (K) for thymine (T), while the P.1. variant substitutes it for asparagine (N). Residue K417 of the RBD and ACE-2 residue D30 commonly interact, and the electrostatic properties of this interaction are highly influential in limiting the mobility of the RBD-ACE2 complex (Fantini, 2021). Mutations in this residue, namely the conversion of lysine to threonine (B.1.351) or to asparagine (P.1) may change the energy of binding to ACE-2. The B.1.351 and P.1. have proved to be more lethal than the B.1.1.7 variant and have higher infectivity (Khan, 2021).

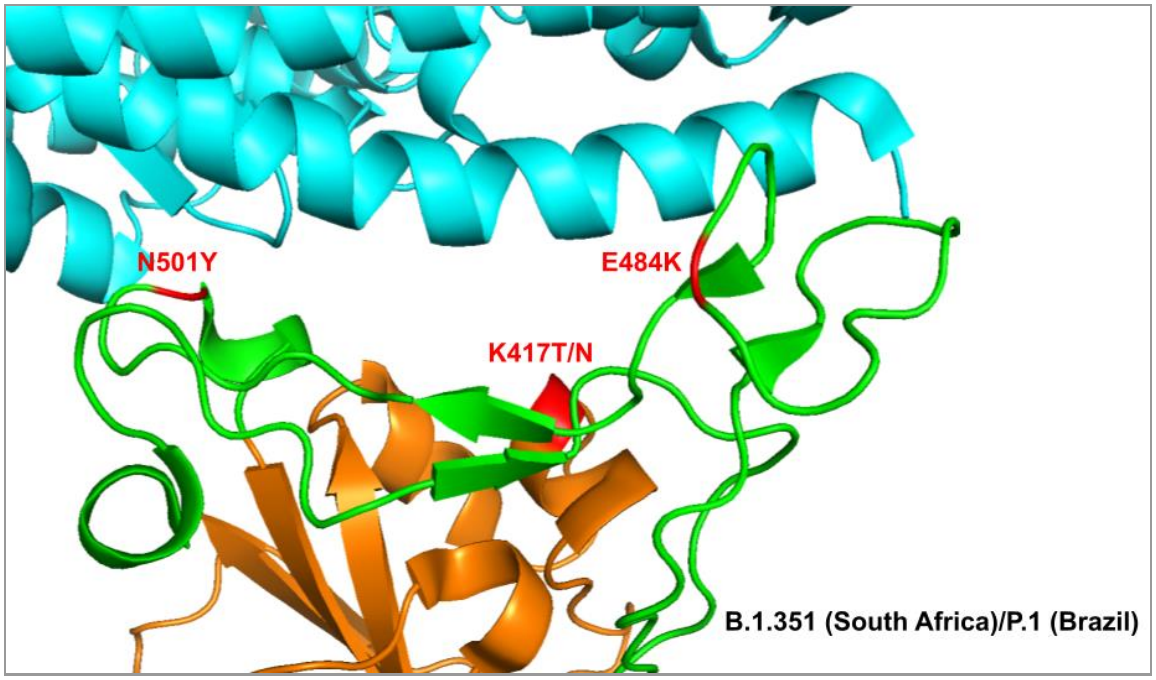

Figure 30. Illustration of the N501Y, E484K, and K417T/N mutations in the RBD of the B.1.351 (South Africa) and P.1 (Brazil) variants. ACE-2 receptor colored blue; RBD colored orange; RBM colored green.

The B.1.617.2 variant, or the Delta variant, first emerged in India in December, 2020. This variant was brought to worldwide attention during the second wave of the pandemic, when it was found to have higher transmissibility compared to the wildtype strain (Shiehzadegan, 2021). The

delta strain also has an increased ability to evade the host immune response as compared to previous strains. The B.1.617.2 variant contains two mutations in the RBD region, L452R and T478K (Wu, 2022) (Figure 31). The L452R mutation is dangerous in that it increases the stability of the spike complex with ACE-2, increases its infectivity, promotes viral replication, and is involved in evading the host cell immune response (Motozono, 2021). The L452R and T478K mutations have been shown to strengthen interactions between ACE-2 residue E37 and RBD residue 403 (Motozono, 2021).

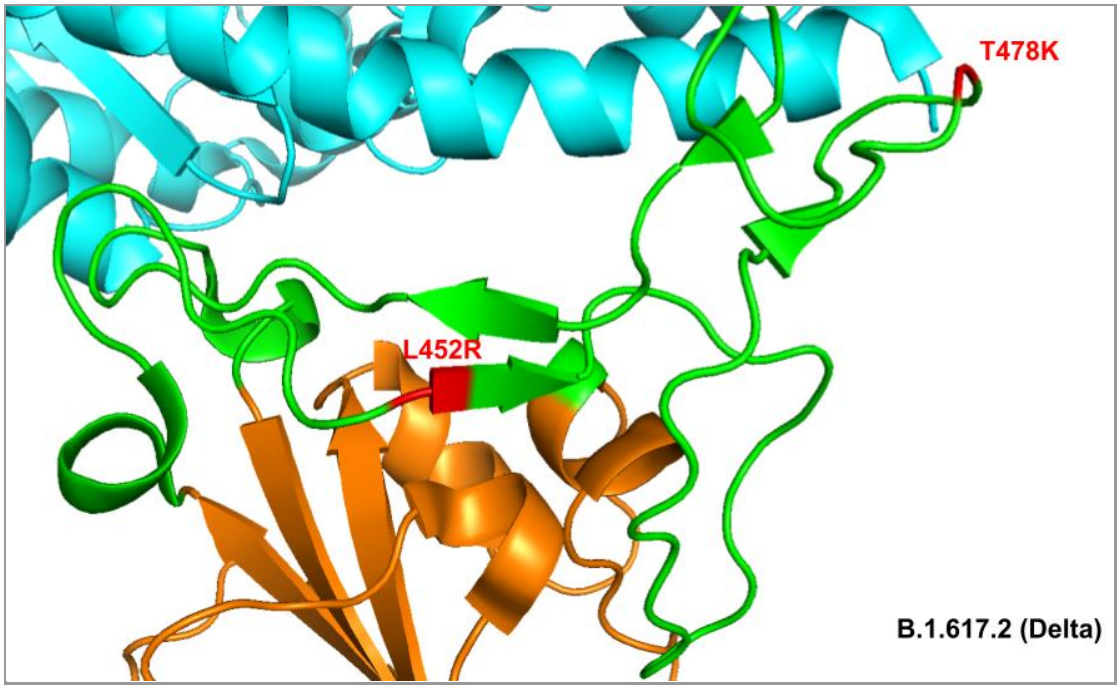

Figure 31. Illustration of the T478K and L452R mutations in the RBD of the B.1.617.2 (Delta) variant. ACE-2 receptor colored blue; RBD colored orange; RBM colored green.

The last variant of concern that is currently being monitored by the CDC is the B.1.1.529 variant, or the Omicron variant. The B.1.1.529 variant is the most unlike the wildtype spike, containing 15 mutations in the spike RBD (Wu, 2022) (Figure 32). One hypothesis for the reason behind the unusual number of mutations in this variant is that this strain may have evolved unseen in animal

hosts before reinfecting humans (Mallapaty, 2022). These mutations have been shown experimentally to cause it to be able to more easily evade host immunity, but overall, the Omicron spike has weaker interaction with ACE-2 (Wu, 2022). It should be noted that weaker binding with ACE-2 does not necessarily guarantee lower transmissibility, and that the increase in ability to evade immunity may make up for its deficiencies in binding to the host receptor (Wu, 2022).

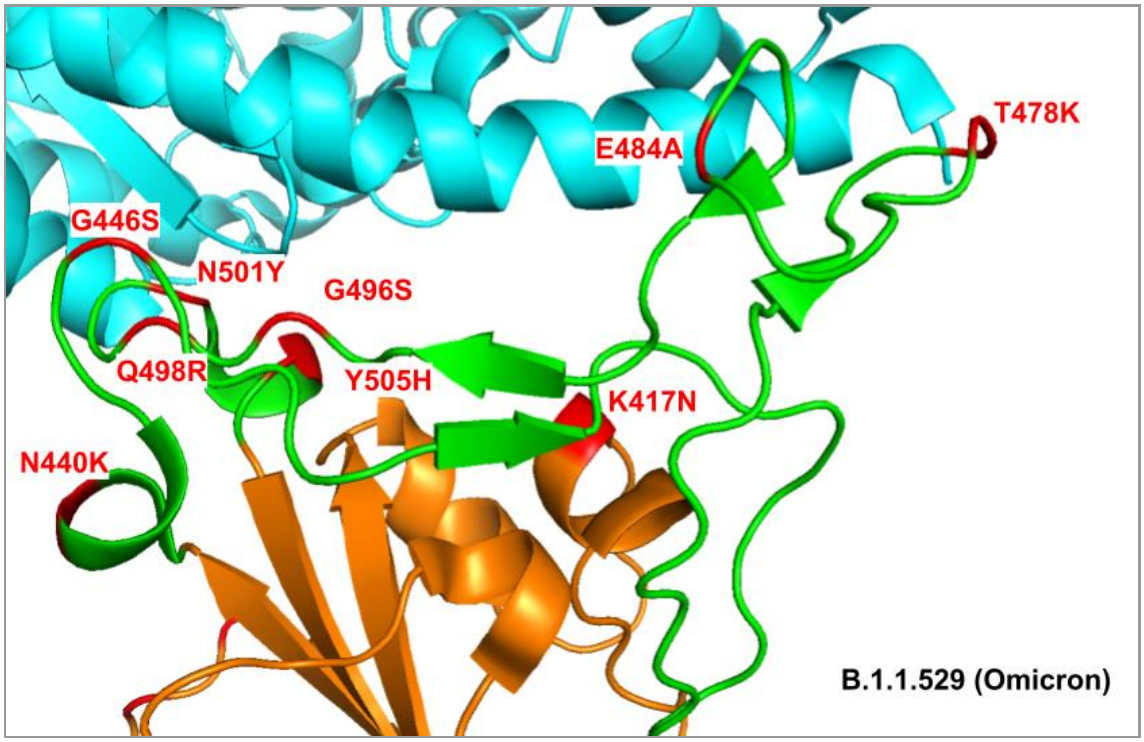

Figure 32. Illustration of the mutations in the RBM of the B.1.1.529 (Omicron) variant. ACE-2 receptor colored blue; RBD colored orange; RBM colored green.

The present study will identify the top inhibitors against the wildtype spike from an FDA-approved compound library and evaluate the top performing drugs against the variant spikes: B.1.1.7 (U.K.), B.1.351 (Africa), P.1. (Brazil), B.1.617.2 (Delta) and B.1.1.529 (Omicron) variants.

5.2 Method

The 3-D crystal structure of the spike receptor binding-domain bound to human angiotensin-converting enzyme ACE-2 (PDB ID: 6M0J) at a resolution of 2.45 Å was downloaded from the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (Lan, Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor, 2020). The ZINC database containing 1750 FDA approved molecules were docked against the receptor 6M0J using Autodock Vina (Trott, 2010).

B.1.1.7 (UK)	Asn501 → Tyr (N501Y)
B.1.351 (South Africa)	Lys417 → Thr (K417T) Glu484 → Lys (E484K) Asn501 → Tyr (N501Y)
P.1 (Brazil)	Lys417 → Asn (K417N) Glu484 → Lys (E484K) Asn501 → Tyr (N501Y)
B.1.617.2 (Delta)	Leu452 → Arg (L452R) Thr478 → Lys (T478K)
B.1.1.529 (Omicron)	G339D, S371L, S373P, S375F, K417N, N440K, G446S, S477N, T478K, E484A, Q493R, G496S, Q498R, N501Y, and Y505H

Table 3. SARS-CoV-2 variants and respective mutated residues in the spike RBM.

A midsize box was used to specify the grid coordinates for docking. The coordinates were chosen to contain the SARS-CoV-2 receptor-binding motif, specifically the β6 sheet that directly interfaces with ACE-2. The grid box mapping parameters along x, y and z coordinates are: (Å) x= 38 y= 78 z= 40 and Center (Å) x= -36.139 center_y= 33.528 center_z= 3.444. Pre-processing for virtual screening was conducted using AutodockTools (Morris, 2009). All BASH scripts created to automate virtual screening and preprocessing are included in Appendix A.

Next, the top four compounds with the highest median binding affinity against the wildtype spike were filtered. These compounds were then tested against the five variant strains: B.1.1.7 (UK), B.1.135 (South Africa), P.1 (Brazil), B.1.617.2 (Delta), and B.1.1.529 (Omicron). Variants were created from 6M0J using the PyMol mutagenesis tool. Each compound was docked against each variant 100 times. For each drug-variant complex, the most common conformation was picked for interaction analysis and visualized in PyMol (Schrödinger, LLC).

5.3 Results and Discussion

The following FDA approved drugs were assessed for their performance against the wildtype and B.1.1.7, B.1.351, P.1, B.1.617.2 and B.1.1.529 variants. We will discuss the binding affinity distribution and conformations of the top four candidates against the wildtype and variant spikes. These compounds are paritaprevir, sirolimus, voxilaprevir, and cabazitaxel.

Paritaprevir is an antiviral approved for the treatment of chronic Hepatitis C that inhibits the NS3/4A serine protease of Hepatitis C Virus, preventing the cleavage of mature viral particles. Previous studies found paritaprevir to bind the S-RBD/ACE-2 interface consistently and with high binding affinity (de Oliveira, 2021). In accordance with previous docking results, we find paritaprevir has the second highest median binding affinity for the wildtype spike. Paritaprevir binds the variant spike proteins with significantly lower affinities than it does the wildtype, but none show poor binding (less than -8 kcal/mol) (Figure 33).

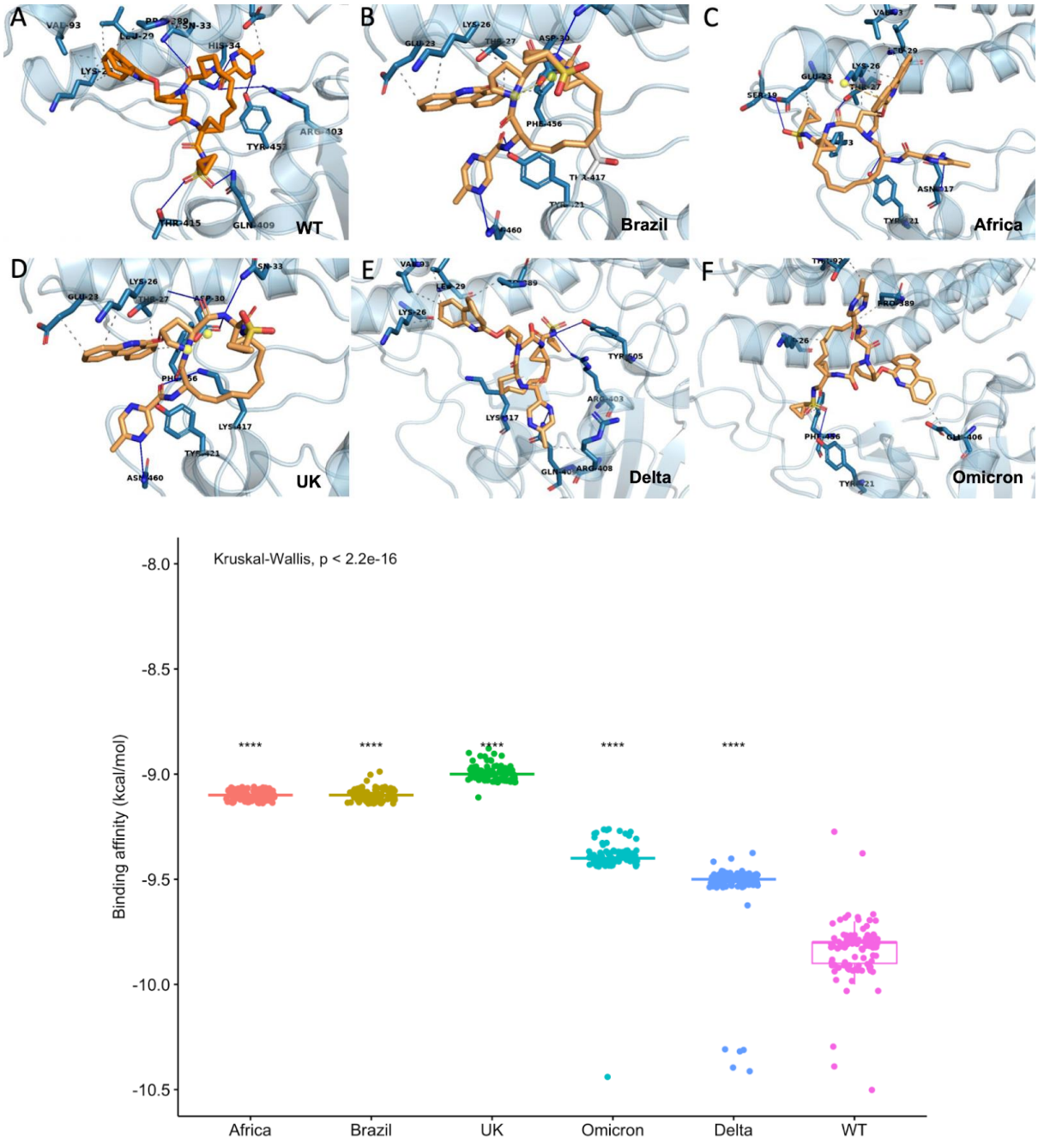

Figure 33. Most common binding conformation of Paritaprevir bound to the RBD/ACE-2 interface for wildtype and variant spike proteins. Binding affinity distributions of Paritaprevir bound to wildtype and variants.

The Delta and Omicron variants have second and third best binding energy respectively after the wildtype, but also the highest degree of variation. In both of these variant complexes, paritaprevir is stabilized only through hydrophobic interactions on the ACE-2 side. In comparison, complexed with the wildtype, paritaprevir forms two hydrogen bonds with ACE-2 residues His34 and Asn33. The additional support on the receptor side may be responsible for the higher binding affinity of paritaprevir with the wildtype. However, the total number of bonds formed on either side of the interface does not explain the variant binding affinities. The Brazil and African variants have the highest number of hydrogen bonds formed with paritaprevir after the wildtype, with an additional salt bridge or pi-cation bond respectively but are outperformed by the Delta and Omicron variants.

It is possible that interaction with specific residues is a better predictor of binding energy than the total number of bonds formed. For example, paritaprevir only bonds with Arg403 of the S-RBD in the wildtype and the Delta complexes. Previous studies report that Arg403 contributes significantly to the strength of the S-RBD interaction with the ACE-2 receptor (Zech, Fabian, et al. 2021; Pavlova, Anna, et al., 2021). Paritaprevir makes a hydrogen bond with Delta S-RBD residue Tyr505, which may change its position such that interaction with His34 and Asn33 is less favorable. The Omicron variant has a mutated His505, which may be less likely to interact with paritaprevir than tyrosine. This alters its position away from Arg403, which forms a hydrogen bond with paritaprevir in the wildtype and Delta complexes. It is possible that additional support on the ACE-2 side of the interface is responsible for the higher binding affinity in the wildtype.

All types of viral spike protein bind strongly with paritaprevir and have significantly different binding compared amongst each other, with the African and Brazilian variants being the most similar (Figure 33). In both cases, paritaprevir forms hydrogen bonds with Tyr421 and is stabilized by two hydrogen bonds on either side of the interface with an additional stronger bond (salt bridge or pi-cation) on the ACE-2 receptor side. Both variants have a mutated S-RBD residue 417, but asparagine is more likely to form bonds than tyrosine. Notably, although these are the only complexes to form bonds stronger than hydrogen bonds on the ACE-2 side, this does not translate to better binding affinity.

The next compound we will discuss is Sirolimus. Sirolimus is an immunosuppressive drug and mTOR inhibitor approved for treatment of patients with H1N1 pneumonia. (Zhou, Yadi, et al. 2020) Sirolimus has the highest median binding affinity for the wildtype spike. Similar to paritaprevir, sirolimus binds the variant spike proteins with significantly less affinity than it does the wildtype, and the Delta, Omicron, and UK variants show the highest variation in affinity (Figure 34). The Brazilian and African spike proteins are again similar in terms of conformation and are not significantly different in terms of binding energy (Figure 34).

The binding pose observed for the wildtype is similar to that observed for the Delta variant, both forming a salt bridge with His34 of the ACE-2 receptor and a hydrogen bond with Arg403 on the S-RBD side. The T478K mutation shared by the Omicron and Delta variant has been shown to strengthen the interaction between S-RBD and ACE-2 (da Costa, et al. 2022). The UK variant slightly but significantly outperforms the Delta and Omicron variants with less variability, even though sirolimus makes one fewer hydrogen bond in this complex.

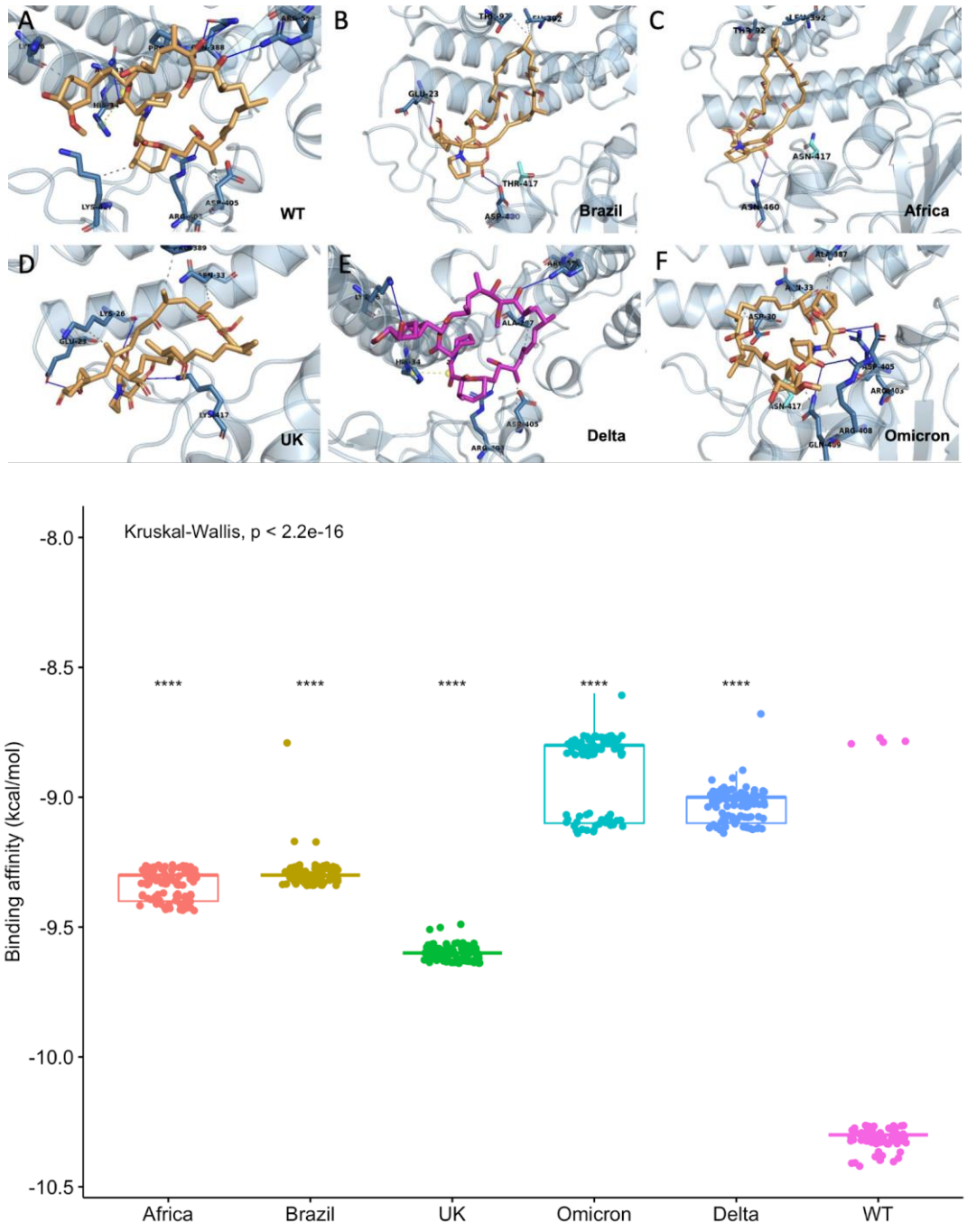

Figure 34. Most common binding conformation of Sirolimus bound to the RBD/ACE-2 interface for wildtype and variant spike proteins. Binding affinity distributions of Sirolimus bound to wildtype and variants.

Sirolimus makes a hydrogen bond with Lys417 only with the UK S-RBD. Previous studies show Lys417 to be a hotspot (Ali and Vijayan. 2020). It is possible that although the conformations that interact with ACE-2 residues Asn33 and His34, as well as Arg403 of the S-RBD, have increased binding affinity, they are also less stable.

The next compound to discuss is Voxilaprevir. Voxilaprevir is an antiviral approved for the treatment of chronic Hepatitis C that inhibits the NS3/4A serine protease of Hepatitis C Virus, preventing the cleavage of mature viral particles. Voxilaprevir has the fourth highest median binding affinity for the wildtype spike. Voxilaprevir shows significantly different binding affinities for the wildtype and all variant spike proteins (Figure 35). Again, the Brazilian and African variants have very similar binding affinities with the lowest variability of the group.

Similar to sirolimus, voxilaprevir complexed with the UK variant has the second highest binding affinity and with less variation than Delta and Omicron. The complex with the UK S-RBD is not unique in its interactions with ACE-2 residues Asn33 and His34 and S-RBD residue Arg403, observed also in the Brazilian and African complexes.

The African complex has three additional hydrogen bonds compared to the UK variant, and the Brazilian variant forms two pi-cation bonds, yet we observe that these variants have significantly lower binding affinity (Figure 35). The UK variant is unique in that voxilaprevir forms a hydrogen bond with Lys417, all other variants only show hydrophobic interactions with this residue.

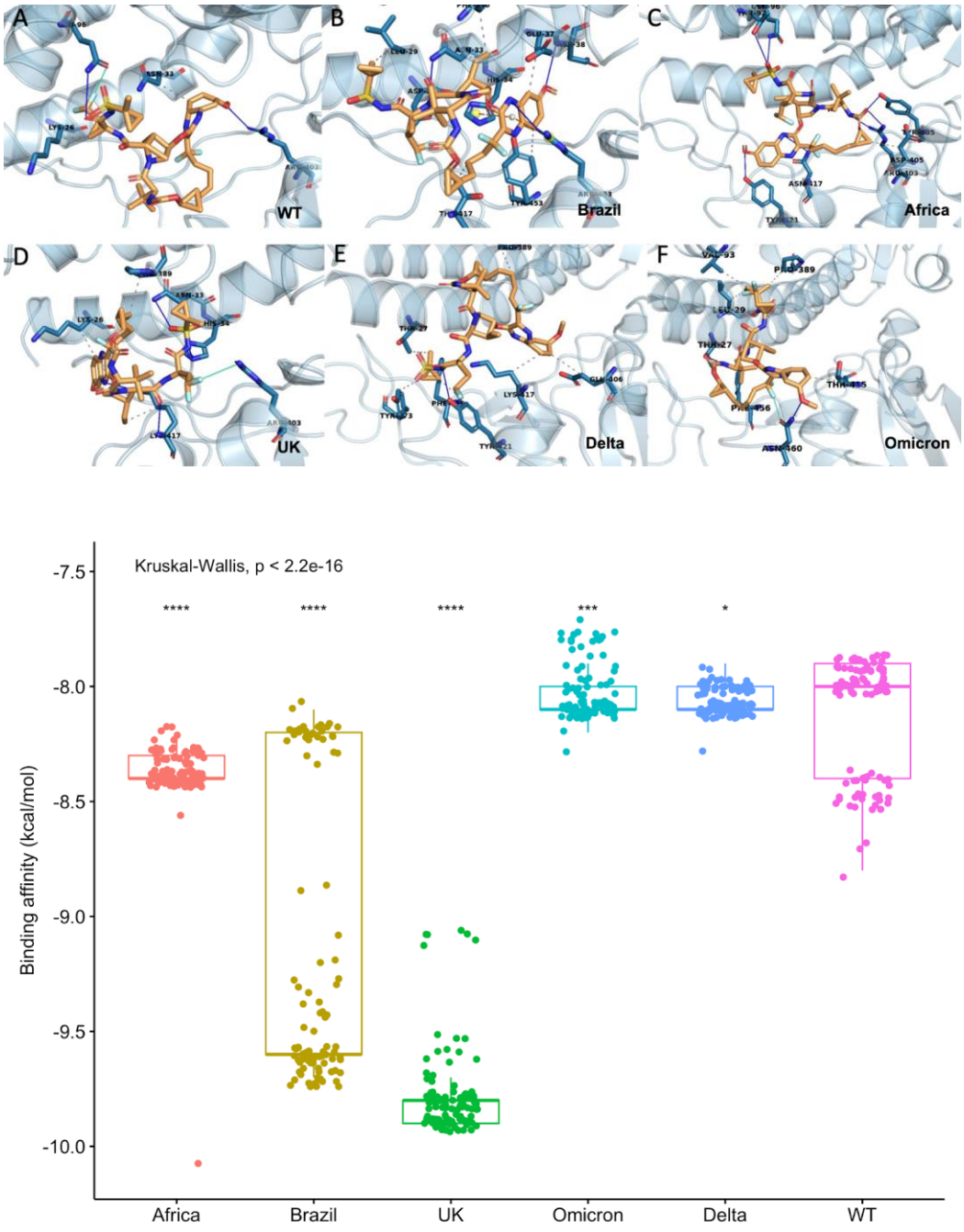

Figure 35. Most common binding conformation of Voxilaprevir bound to the RBD/ACE-2 interface for wildtype and variant spike proteins. Binding affinity distributions of Voxilaprevir bound to wildtype and variants.

The last compound to discuss is Cabazitaxel. Cabazitaxel is a chemotherapeutic agent used to treat prostate cancer. A study found that cancer patients being treated with cabazitaxel exhibited lower rates of SARS-CoV-2 infection (Foote, Michael B., et al. 2021).

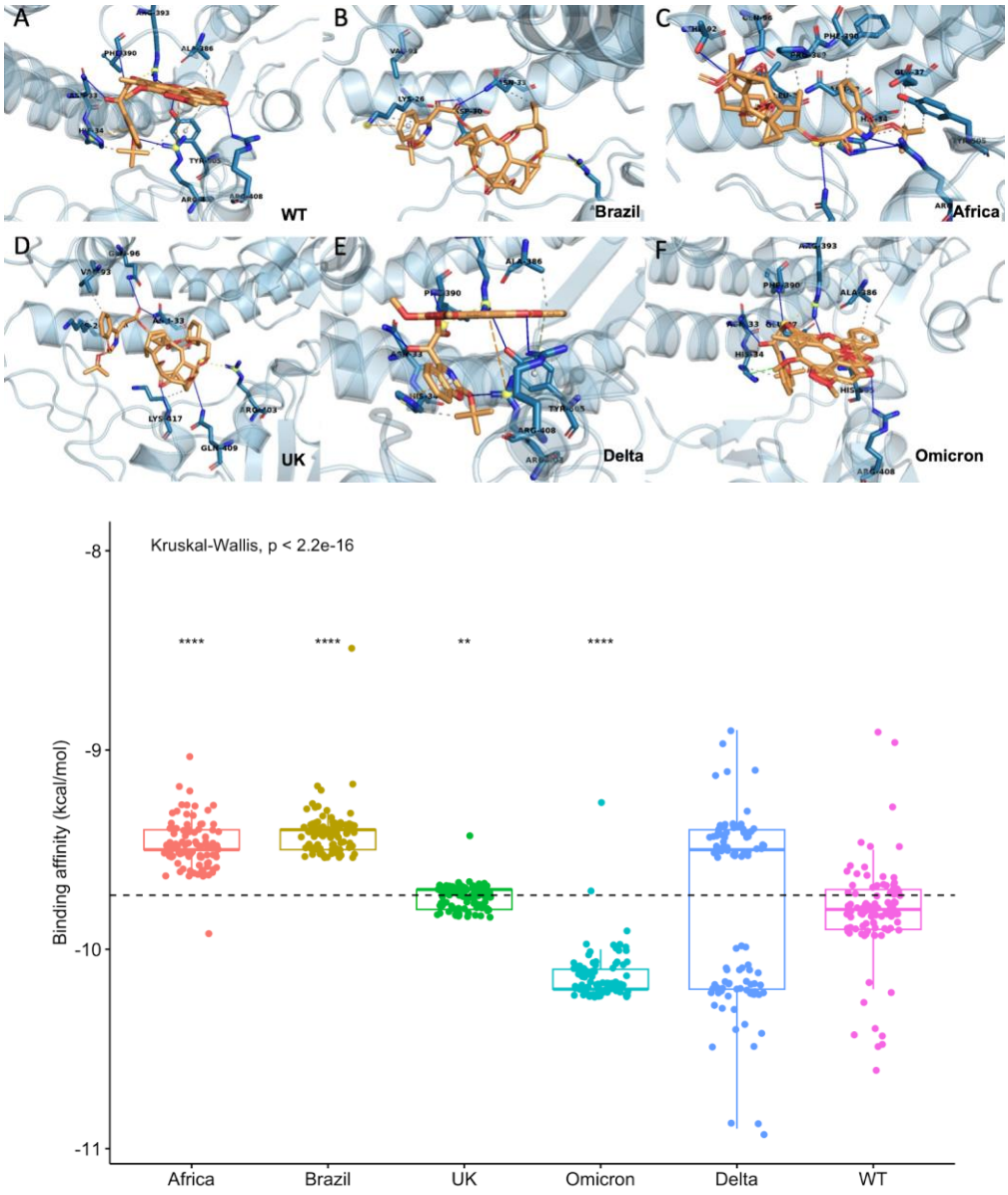

Figure 36. Most common binding conformation of Cabazitaxel bound to the RBD/ACE-2 interface for wildtype and variant spike proteins. Binding affinity distributions of Cabazitaxel bound to wildtype and variants.

It binds the S-RBD of the wildtype consistently and has the third highest binding affinity for the wildtype spike. Cabazitaxel binds significantly differently to all spike-protein types (Figure 36). All spike-protein types except for Omicron interacted with S-RBD residue Arg403.

Other common interactions were with ACE-2 residues Asn33 and His34. Notably, cabazitaxel interacted strongly with Tyr505, forming pi-cation bonds with the wildtype, pi-stacking bonds with the Delta variant, and a hydrogen bond with the Omicron variant (Figure 36). Unlike the previous variants, the Brazilian and African variants are quite different in terms of binding affinity and conformation. Again, although the African variant makes several additional bonds compared to Brazil and Omicron, this does not translate to better binding affinity.

5.4 Conclusion

The top candidate drugs exhibited a few notable similarities in interaction with the wildtype and variant spike protein complexes. In all cases, the drugs bound the variant spike proteins with significantly lower binding affinity. Following the wildtype, the Delta and UK variants had the next best binding affinities and tended to perform similarly in terms of average binding affinity and variance. These variant spike proteins are the most structurally similar to the wildtype spike, each containing only one mutation in the RBD region. However, the overall structural similarity or number of mutations is not necessarily a predictor of binding affinity. The Omicron variant consistently outperformed the African and Brazilian variants, even though it has the largest total number of mutations: it contains the Delta and UK mutations (T478K and N501Y), as well as mutated residues at positions 417 and 484 as seen in the African and Brazilian variants. Omicron contains the same K417N mutation as the Brazil variant, but has an E484A mutation, instead of the African and Brazilian E484K mutation. It should be noted that only

certain mutations fall within the specified binding site, notably K417N, E484A, N501Y, and T478K, but the Omicron variant is nonetheless the most structurally dissimilar.

Visually, the Omicron-bound complexes tend to most closely resemble the Delta-bound complexes, interacting with many of the same residues. It is possible then that the T478K mutation most influences drug binding to the Omicron spike for this binding region. The T478K mutation, along with N501Y, Q493K/R mutations, has been reported to be the most influential in terms of increasing the strength of interaction between the Omicron spike and ACE2 (Kim S. L., 2023). Disruption of this interaction is favorable, and although we do not observe direct binding of drugs to residue 478, the T478K mutation appears to increase drug binding affinity for both Omicron and Delta variants.

The African and Brazilian variants performed similarly both in terms of mean binding affinity and variability. Both variants consistently displayed lower variance in binding affinity than the wildtype and UK, Delta, and Omicron variants. The number of bonds formed on either side of the interface is not a consistent predictor of binding affinity or variability. The African and Brazilian variants frequently formed a larger total number of bonds, but still exhibited significantly decreased performance compared to the other variants and the wildtype. However, the performance of the selected drugs against the African and Brazilian variants was still within an acceptable binding affinity range and showed lower variance among experimental runs, indicating that these drugs are promising candidates for these variants despite their slightly lower performance.

Conclusion

The COVID-19 global pandemic marks the third outbreak caused by a coronavirus, following the 2002 severe acute respiratory syndrome (SARS-CoV) and 2012 Middle East respiratory syndrome (MERS-CoV) epidemics. Although the previous outbreaks had higher fatality rates, the novel SARS-CoV-2 virus has had a much wider transmission. Furthermore, long-term conditions associated with the contraction of COVID-19 complicate the threat to public health. While quarantining and social distancing protocol was advised in the early stages of the pandemic, normal public life has largely resumed. Widespread vaccination remains the greatest defense against transmission of COVID-19.

Approved therapeutics such as remdesivir, lopinavir, and ritonavir-boosted nirmatrelvir inhibit parts of the viral machinery necessary for replication and assembly. There are currently no FDA-approved inhibitors that inhibit the first stage of infection, which is viral entry. SARS-CoV-2 enters the host cell through the interaction of its spike protein with human cell receptor ACE-2. The spike protein consists of two subunits that facilitate viral entry and subsequent fusion to the host cell, S1 and S2, respectively. The S1 subunit contains the spike receptor binding domain (S-RBD). The receptor binding domain is the region that directly binds the ACE-2 receptor.

There are two possible methods for interfering with the ability of the S-RBD to interact with ACE-2: direct and indirect (allosteric) inhibition. A direct inhibitor competes with the virus to bind the S-RBD. Allosteric inhibitors bind alternative locations on the spike protein to cause a change in the conformation of the S-RBD, preventing its successful function. Specifically, allosteric inhibition aims to prevent the ability of the S-RBD to switch from the ‘down’ state, in

which it cannot interact with ACE-2, to the ‘up’ state. In this research we investigated both direct and allosteric inhibition of the spike protein.

To explore direct inhibition, we tested the ability of FDA-approved druglike compounds from the ZINC database to directly bind the S-RBD and ACE-2 interface. We found that Paritaprevir, Sirolimus, Voxilaprevir, and Cabazitaxel had the highest binding affinity for the wildtype spike receptor binding domain. We further tested these compounds against five variants of SARS-CoV-2: B.1.1.7 (UK), B.1.351 (South Africa), P.1. (Brazil), B.1.617.2 (Delta), and B.1.1.529 (Omicron).

In analyzing the performance of the top compounds against the variants, we noted several patterns in binding affinity. In all cases, the drugs bound all the variant spike proteins with significantly lower binding affinity. After the wildtype, the Delta and UK variants had the next best binding affinities, and tended to perform similarly in terms of average binding affinity and variance, likely due to the structural similarity of these variant spike proteins with the wildtype spike. The Delta and UK variants each contain only one mutation in the RBD region. However, we found that overall structural similarity and number of mutations is not necessarily a predictor of binding affinity.

The Omicron variant consistently outperformed the African and Brazilian variants, even though it has the largest total number of mutations: it contains the Delta and UK mutations (T478K and N501Y), as well as mutated residues at positions 417 and 484 as seen in the African and Brazilian variants. Omicron contains the same K417N mutation as the Brazil variant, but has an E484A mutation, instead of the African and Brazilian E484K mutation. We also found that the

Omicron-bound complexes tend to most closely resemble the Delta-bound complexes. It is possible then that the T478K mutation most influences drug binding to the Omicron spike for this binding region.

The African and Brazilian variants performed similarly both in terms of mean binding affinity and variability. Both variants consistently displayed lower variance in binding affinity than the wildtype and UK, Delta, and Omicron variants. The number of bonds formed on either side of the interface is not a consistent predictor of binding affinity or variability. The African and Brazilian variants frequently formed a larger total number of bonds, but still exhibited significantly decreased performance compared to the other variants and the wildtype. However, the performance of the selected drugs against the African and Brazilian variants was still within an acceptable binding affinity range and showed lower variance among experimental runs, indicating that these drugs are promising candidates for these variants despite their slightly lower performance.

To explore indirect, or allosteric, inhibition of the spike protein, we investigated two allosteric sites previously specified in the literature. The first site, Site A, is located at the junction of the two subunits that comprise the spike protein, SD1 and SD2 (Wang, Qian, et al. 2021). The second site, Site B, is located underneath the spike receptor binding domain and receptor binding motif (RBM) (Li, Bingqian, et al. 2021).

We tested two libraries of compounds against both allosteric sites. Library 1 consisted of 200 compounds downloaded from the ZINC database. Library 2 consisted of compounds CPD1-7, CPD20, and CPD26. The lead compounds in both libraries performed within a similar range of

approximately -7 to -9 kcal/mol. We found a slight preference for Site B for both libraries. This difference may be due to the increased density of potential contact residues at Site B compared to Site A. We found that there were two common clusters of binding for the top 10 lead compounds at Site A that were situated in the pocket between HR1, CTD1, and CTD2. Interactions with residues that make up the CTD and HR proteins have the potential to disrupt movement of these structures during fusion. At Site B, we also found two common clusters of binding that interacted with residues R408 and D405, which are indicated to be crucial for the pre-fusion state of the spike glycoprotein. In addition to analyzing the performance of the top 10 compounds at each site, we also individually analyzed the three lead compounds common to both allosteric sites. These compounds were ZINC06524444, ZINC08101192, and ZINC08101193. The latter two compounds showed a preference for Site B, while ZINC06524444 preferred Site A.

In analyzing the performance of Library 2, we found that compounds CPD1-7, CPD20, and CPD26 bound both Site A and Site B with moderate binding affinity. CPD7 performed the least well at both sites, with some conformations having a binding affinity less than -6 kcal/mol. CPD1, CPD3, and CPD4 consistently had high variation in performance at both sites. In general, the tested compounds had a higher affinity for Site B than for Site A, particularly CPD26, CPD5, and CPD6. The compound that performed the best overall with the least variation between sites was CPD2.

In examining the most common conformations of CPD7, CPD20, and CPD26 against their expected binding poses, none of these compounds bound Site A exactly as expected. Similar binding poses to the expected were observed, in terms of orientation of the ligand and interacting residues, however these conformations were not the most frequent for CPD20 and CPD26.

However, the most frequently observed pose for CPD7 was also the conformation most similar to the predicted pose. In general, we find that both compound libraries had a slight preference for Site B. We identified the top 10 performing compounds at each site, as well as the three common lead compounds. Our results suggest that these compounds have the potential to inhibit spike function.

Future studies may be able to assess the lead compounds identified in this study *in vitro* to test whether these compounds perform as well experimentally as they do in simulation. Experimental evidence is necessary to build a strong case that these compounds can act as inhibitors of the SARS-CoV-2 spike protein. Another direction future research may take is to test the identified allosteric inhibitors against the variant spike proteins. It will be useful to further characterize the exact binding conformations of allosteric inhibitors and how those conformations overlap with mutated residues.

In addition to experimental studies and the inclusion of variant cases, future work may also focus on further describing the stability of compound binding through simulation. Molecular dynamics simulations are a popular type of software used to track the dissolution of a ligand-receptor complex over time. These simulations also provide valuable information about the stability of binding that a binding affinity distribution cannot. Lastly, the lead compounds both allosteric and non-allosteric may be a starting point for *de novo* drug design. If these compounds perform well experimentally, drug optimization techniques may be used to modify the chemical structure to increase binding affinity and stability against the SARS-CoV-2 spike protein.

References

Aarts, E. K. (2005). Simulated annealing. *Search methodologies: introductory tutorials in optimization and decision support techniques*, 187-210.

AIP Publication Board. (1990). *AIP Style Manual, Fourth Edition.* New York, NY: American Institute of Physics.

Aiyegbusi, O. L. (2021). Symptoms, complications and management of long COVID: a review. *Journal of the Royal Society of Medicine*, 114(9), 428-442.

Alkhatib, M. S.-S. (2021). SARS-CoV-2 variants and their relevant mutational profiles: update summer 2021. *Microbiology spectrum*, 9(3), e01096-21.

Amicone, M. B.-Z. (2022). Mutation rate of SARS-CoV-2 and emergence of mutators during experimental evolution. *Evolution, medicine, and public health*, 10(1), 142-155.

Amicone, M. B.-Z. (2022). Mutation rate of SARS-CoV-2 and emergence of mutators during experimental evolution. *Evolution, medicine, and public health*, 10(1), 142-155.

Ban, T. A. (2022). The role of serendipity in drug discovery. *Dialogues in clinical neuroscience*.

Beigel, J. H. (2020). Remdesivir for the treatment of Covid-19—preliminary report. *New England Journal of Medicine*, 383(19), 1813-1836.

Belitzky, E. C.-N. (2022). Determining Binding Affinity (KD) of Radiolabeled Antibodies to Immobilized Antigens. *JoVE (Journal of Visualized Experiments)*, (184), e63927.

Bianchi, M. B. (2020). Sars-CoV-2 envelope and membrane proteins: structural differences linked to virus characteristics? *BioMed Research International*.

Bonanad, C. G.-B.-S.-G. (2020). The effect of age on mortality in patients with COVID-19: a meta-analysis with 611,583 subjects. *Journal of the American Medical Directors Association*, 21(7), 915-918.

Bonvin, A. M. (2006). Flexible protein–protein docking. *Current opinion in structural biology*, 16(2), 194-200.

Bosch, B. J. (2003). The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. *Journal of virology*, 77(16), 8801-8811.

Brouqui, P. A. (2021). Asymptomatic hypoxia in COVID-19 is associated with poor outcome. . *International Journal of Infectious Diseases*, 102, 233-238.

Candido, K. L. (2022). Spike protein of SARS-CoV-2 variants: a brief review and practical implications. *Brazilian Journal of Microbiology*, 53(3), 1133-1157.

Chen, R. E. (2021). Resistance of SARS-CoV-2 variants to neutralization by monoclonal and serum-derived polyclonal antibodies. *Nature medicine*, 27(4), 717-726.

Cheng, Y. H. (2019). Crystal structure of the S1 subunit N-terminal domain from DcCoV UAE-HKU23 spike protein. *Virology*, 535, 74-82. .

Cho, A. E. (2005). Importance of accurate charges in molecular docking: quantum mechanical/molecular mechanical (QM/MM) approach. *Journal of computational chemistry*, 26(9), 915-931.

Day, C. J. (2021). Multidisciplinary approaches identify compounds that bind to human ACE2 or SARS-CoV-2 spike protein as candidates to block SARS-CoV-2–ACE2 receptor interactions. *MBio*, 12(2), 10-1128.

de Oliveira, M. D. (2021). Comparative computational study of SARS-CoV-2 receptors antagonists from already approved drugs.

Dryden-Peterson, S. K. (2023). Nirmatrelvir plus ritonavir for early COVID-19 in a large US health system: a population-based cohort study. *Annals of internal medicine*, 176(1), 77-84.

Ekins, S. M. (2007). In silico pharmacology for drug discovery: methods for virtual ligand screening and profiling. *British journal of pharmacology*, 152(1), 9-20.

Fantini, J. Y. (2021). Structural dynamics of SARS-CoV-2 variants: A health monitoring strategy for anticipating Covid-19 outbreaks. *Journal of Infection*, 83(2), 197-206.

Forni, D. C. (2017). Molecular evolution of human coronavirus genomes. *Trends in microbiology*, 25(1), 35-48.

Fuhrmann, J. R. (2010). A new Lamarckian genetic algorithm for flexible ligand-receptor docking. *Journal of computational chemistry*, 31(9), 1911-1918.

Gao, T. G. (2021). Identification and functional analysis of the SARS-COV-2 nucleocapsid protein. *BMC microbiology*, 21(1), 1-10.

Gohlke, H. H. (2000). Knowledge-based scoring function to predict protein-ligand interactions. *Journal of molecular biology*, 295(2), 337-356.

Gordon, C. J. (2021). Molnupiravir promotes SARS-CoV-2 mutagenesis via the RNA template. *Journal of Biological Chemistry*, 297(1).

Grellet, E. G. (2022). Replication of the coronavirus genome: a paradox among positive-strand RNA viruses. *Journal of Biological Chemistry*, 101923.

Guedes, I. A. (2018). Empirical scoring functions for structure-based virtual screening: applications, critical aspects, and challenges. *Frontiers in pharmacology*, 9, 1089.

Guo, L. L. (2023). Targetable elements in SARS-CoV-2 S2 subunit for the design of pan-coronavirus fusion inhibitors and vaccines. *Signal Transduction and Targeted Therapy*, 8(1), 197.

Gur, M. T. (2020). Conformational transition of SARS-CoV-2 spike glycoprotein between its closed and open states. *The Journal of chemical physics*, 153(7), 075101.

Harrison, S. C. (2005). Mechanism of membrane fusion by viral envelope proteins. *Advances in virus research*, 64, 231-261.

Hartenfeller, M. &. (2011). Enabling future drug discovery by de novo design. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 1(5), 742-759.

Harvey, W. T. (2021). SARS-CoV-2 variants, spike mutations and immune escape. *Nature Reviews Microbiology*, 19(7), 409-424.

Harvey, W. T. (2021). SARS-CoV-2 variants, spike mutations and immune escape. . *Nature Reviews Microbiology*, 19(7), 409-424.

Hill, A. D. (2015). Scoring functions for AutoDock. *Glycoinformatics*, 467-474.

Hodos, R. A. (2016). In silico methods for drug repurposing and pharmacology. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 8(3), 186-21.

Huang, Y. Y. (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.

Huang, Y. Y. (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.

Huang, Y. Y. (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.

Imran, M. K. (2021). Discovery, development, and patent trends on molnupiravir: a prospective oral treatment for COVID-19. *Molecules*, 26(19), 5795.

Irwin, J. J. (2012). ZINC: a free tool to discover chemistry for biology. *Journal of chemical information and modeling*, 52(7), 1757-1768.

Jorgensen, W. L. (1991). Rusting of the lock and key model for protein-ligand binding. *Science*, 254(5034), 954-955.

Jose, S. G.-S. (2022). Potential of phytocompounds from Brassica oleracea targeting S2-domain of SARS-CoV-2 spike glycoproteins: Structural and molecular insights. *Journal of Molecular Structure*, 1254, 132369. .

Jxwtgmys, F. (2016). The Physics of Bubbles. *Physical Review, II*(23), 97-108.

Kalhor, H. S. (2022). Repurposing of the approved small molecule drugs in order to inhibit SARS-CoV-2 S protein and human ACE2 interaction through virtual screening approaches. *Journal of Biomolecular Structure and Dynamics*, 40(3), 1299-1315.

Khan, A. Z. (2021). Higher infectivity of the SARS-CoV-2 new variants is associated with K417N/T, E484K, and N501Y mutants: an insight from structural data. *Journal of cellular physiology*, 236(10), 7045-7057.

Kim, S. L. (2023). Binding of human ACE2 and RBD of omicron enhanced by unique interaction patterns among SARS-CoV-2 variants of concern. *Journal of computational chemistry*, 44(4), 594-601.

Kim, S. N. (2021). SARS-CoV-2 Omicron mutation is faster than the chase: multiple mutations on spike/ACE2 interaction residues. *Immune Network*, 21(6).

Lamb, Y. N. (2022). Nirmatrelvir plus ritonavir: first approval. *Drugs*, 82(5), 585-591.

Lan, J. G. (2020). Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*, 581(7807), 215-220.

Lan, J. G. (2020). Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*, 581(7807), 215-220.

Lei, Z. N. (2020). Chloroquine and hydroxychloroquine in the treatment of malaria and repurposing in treating COVID-19. *Pharmacology & Therapeutics*, 216, 107672.

Lenard, J. (2008). Viral membranes. *Encyclopedia of virology*, 308.

Li, B. W. (2021). Identification of potential binding sites of sialic acids on the RBD domain of SARS-CoV-2 spike protein. *Frontiers in Chemistry*, 9, 659764.

Li, F. (2016). Structure, function, and evolution of coronavirus spike proteins. *Annual review of virology*, 3, 237-261.

Li, Q. C. (2010). PubChem as a public resource for drug discovery. *Drug discovery today*, 15(23-24), 1052-1057.

Liebeschuetz, J. W. (2012). Pose prediction and virtual screening performance of GOLD scoring functions in a standardized test. *Journal of computer-aided molecular design*, 26, 737-748.

Lin, H. X. (2021). Remdesivir in Coronavirus Disease 2019 (COVID-19) treatment: a review of evidence. *Infection*, 49, 401-410.

Liu, X. L. (2022). Rampant C-to-U deamination accounts for the intrinsically high mutation rate in SARS-CoV-2 spike gene. *RNA*, 28(7), 917-926.

Liu, Y. L. (2021). The N501Y spike substitution enhances SARS-CoV-2 transmission. *BioRxiv*.

Lotfi, M. H. (2020). COVID-19: Transmission, prevention, and potential therapeutic opportunities. *Clinica chimica acta*, 508, 254-266.

Mahase, E. (2021). Covid-19: Molnupiravir reduces risk of hospital admission or death by 50% in patients at risk, MSD reports.

Mahase, E. (2022). Covid-19: Has the spread of omicron BA. 2 made antibody treatments redundant? .

Mallapaty, S. (2022). Where did Omicron come from? Three key theories. *Nature*, 26-28.

Mannar, D. S. (2022). SARS-CoV-2 Omicron variant: Antibody evasion and cryo-EM structure of spike protein–ACE2 complex. *Science*, 375(6582), 760-764.

McCreary, E. K. (2020). Efficacy of Remdesivir in COVID-19. *Jama*, 324(11), 1041-1042.

Mesel-Lemoine, M. M. (2012). A human coronavirus responsible for the common cold massively kills dendritic cells but not monocytes. *Journal of virology*, 86(14), 7577-7587.

Morris, G. M. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of computational chemistry*, 30(16), 2785-2791.

Motozono, C. T. (2021). SARS-CoV-2 spike L452R variant evades cellular immunity and increases infectivity. *Cell host & microbe*, 29(7), 1124-1136.

N M O'Boyle, M. B. (2011). Open Babel: An open chemical toolbox. *J. Cheminf*, 3, 33.

Neves, B. J.-F.-F. (2018). QSAR-based virtual screening: advances and applications in drug discovery. *Frontiers in pharmacology*, 9, 1275.

Olotu, F. A. (2020). Leaving no stone unturned: Allosteric targeting of SARS-CoV-2 spike protein at putative druggable sites disrupts human angiotensin-converting enzyme interactions at the receptor binding domain. *Informatics in medicine unlocked*, 21, 100451.

Park, K. (2019). A review of computational drug repurposing. *Translational and clinical pharmacology*, 27(2), 59-63.

Parvathaneni, V. K. (2019). Drug repurposing: a promising tool to accelerate the drug discovery process. *Drug discovery today*, 24(10), 2076-2085.

Pellecchia, M. S. (2002). NMR in drug discovery. *Nature Reviews Drug Discovery*, 1(3), 211-219.

Pham, T. H. (2021). A deep learning framework for high-throughput mechanism-driven phenotype compound screening and its application to COVID-19 drug repurposing. *Nature machine intelligence*, 3(3), 247-257.

Prajapat, M. S. (2020). Virtual screening and molecular dynamics study of approved drugs as inhibitors of spike protein S1 domain and ACE2 interaction in SARS-CoV-2. *Journal of Molecular Graphics and Modelling*, 101, 107716.

Pushpakom, S. I. (2019). Drug repurposing: progress, challenges and recommendations. *Nature reviews Drug discovery*, 18(1), 41-58.

QwcAkpsulka, M. (2017). *My Pancakes Are Burned.* Orange, CA: Cooking with Fire Press.

Rahman, M. S. (2021). Mutational insights into the envelope protein of SARS-CoV-2. *Gene reports*, 22, 100997.

Rao, V. S. (2011). Modern drug discovery process: An in silico approach. *Journal of bioinformatics and sequence analysis*, 2(5), 89-94.

Raveendran, A. V. (2021). Long COVID: an overview. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 15(3), 869-875.

Reaume, A. G. (2011). Drug repurposing through nonhypothesis driven phenotypic screening. *Drug Discovery Today: Therapeutic Strategies*, 8(3-4), 85-88.

Schrodinger, L. (2015). The PyMOL Molecular Graphics System, Version 1.8.

Shiehzadegan, S. A. (2021). Analysis of the delta variant B. 1.617. 2 COVID-19. *Clinics and practice*, 11(4), 778-784.

Shrestha, L. B. (2021). Broadly-neutralizing antibodies against emerging SARS-CoV-2 variants. *Frontiers in immunology*, 4025.

Shukla, R. H. (2021). Signature-based approaches for informed drug repurposing: targeting CNS disorders. *Neuropsychopharmacology*, 46(1), 116-130.

Simonis, A. T. (2021). A comparative analysis of remdesivir and other repurposed antivirals against SARS-CoV-2. *EMBO molecular medicine*, 13(1), e13105.

Singh, A. K. (2021). Molnupiravir in COVID-19: A systematic review of literature. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 15(6), 102329.

Socher, E. C. (2021). Computational decomposition reveals reshaping of the SARS-CoV-2–ACE2 interface among viral variants expressing the N501Y mutation. *Journal of cellular biochemistry*, 122(12), 1863-1872.

Starr, T. N. (2022). ACE2 binding is an ancestral and evolvable trait of sarbecoviruses. *Nature*, 603(7903), 913-918.

Stobart, C. C. (2013). Chimeric exchange of coronavirus nsp5 proteases (3CLpro) identifies common and divergent regulatory determinants of protease activity. *Journal of virology*, 87(23), 12611-12618.

Tian, F. T. (2021). N501Y mutation of spike protein in SARS-CoV-2 strengthens its binding to receptor ACE2. *Elife*, 10, e69091.

Toelzer, C. G. (2020). Free fatty acid binding pocket in the locked structure of SARS-CoV-2 spike protein. *Science*, 370(6517), 725-730.

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

Turabian, K. (1987). *A Manual for Writers of Term Papers, Theses, and Dissertations, Fifth Edition.* Chicago, Ill: University of Chicago Press.

Ugurlucan, M. M. (2012). Aspirin: from a historical perspective. *Recent Patents on Cardiovascular Drug Discovery (Discontinued)*, 7(1), 71-76.

Van Der Hoek, L. P.-O. (2004). Identification of a new human coronavirus. *Nature medicine*, 10(4), 368-373.

Vankadari, N. K. (2022). Structure of human TMPRSS2 in complex with SARS-CoV-2 spike glycoprotein and implications for potential therapeutics. *The Journal of Physical Chemistry Letters*, 13(23), 5324-5333.

Wadood, A. A. (2013). In-silico drug design: An approach which revolutionarised the drug discovery process. *OA Drug Des Deliv*, 1(1), 3.

Walls, A. C. (2020). Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell*, 181(2), 281-292.

Wang, G. Y. (2021). Dalbavancin binds ACE2 to block its interaction with SARS-CoV-2 spike protein and is effective in inhibiting SARS-CoV-2 infection in animal models. *Cell Research*, 31(1), 17-24. .

Wang, Q. W. (2021). Probing the allosteric inhibition mechanism of a spike protein using molecular dynamics simulations and active compound identifications. *Journal of Medicinal Chemistry*, 65(4), 2827-2835.

Warren, G. L. (2006). A critical assessment of docking programs and scoring functions. *Journal of medicinal chemistry*, 49(20), 5912-5931.

Westendorf, K. Ž. (2022). LY-CoV1404 (bebtelovimab) potently neutralizes SARS-CoV-2 variants. *Cell reports*, 39(7).

Williamson, E. J. (2020). Factors associated with COVID-19-related death using OpenSAFELY. *Nature*, 584(7821), 430-436.

Woo, P. C. (2005). Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. *Journal of virology*, 79(2), 884-895.

Wqytellorzort, J. (2015). *A Fine Day for a Walk in the Park.* Orange, CA: Orpfilia Press.

Wrapp, D. W. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*, 367(6483), 1260-1263.

Wu, L. Z. (2022). SARS-CoV-2 Omicron RBD shows weaker binding affinity than the currently dominant Delta variant to human ACE2. *Signal transduction and targeted therapy*, 7(1), 8.

Wu, L. Z. (2022). SARS-CoV-2 Omicron RBD shows weaker binding affinity than the currently dominant Delta variant to human ACE2. *Signal transduction and targeted therapy*, 7(1), 8.

Yang, T. J. (2021). Effect of SARS-CoV-2 B. 1.1. 7 mutations on spike protein structure and function. *Nature structural & molecular biology*, 28(9), 731-739.

Yang, Y. P. (2020). The deadly coronaviruses: The 2003 SARS pandemic and the 2020 novel coronavirus epidemic in China. *Journal of autoimmunity*, 109, 102434.

Yu, R. C. (2020). Computational screening of antagonists against the SARS-CoV-2 (COVID-19) coronavirus by molecular docking. *International Journal of Antimicrobial Agents*, 56(2), 106012.

Zaki, A. M. (2012). Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *New England Journal of Medicine*, 367(19), 1814-1820.

Zech, F. S. (2021). Spike residue 403 affects binding of coronavirus spikes to human ACE2. *Nature Communications*, 12(1), 6855.

Zhai, J. H. (2022). A multiple-step in silico screening protocol to identify allosteric inhibitors of Spike–hACE2 binding. *Physical Chemistry Chemical Physics*, 24(7), 4305-4316.

Zhu, N. Z. (2020). A novel coronavirus from patients with pneumonia in China, 2019. *New England journal of medicine*.

Appendix A.

A.1 Iteration Script

```perl
#!/usr/bin/perl
print"Ligand_file:\t";
$ligfile=<STDIN>;
chomp $ligfile;
open (FH,$ligfile)||die "Cannot open file\n";
@arr_file=<FH>;

for($i=0;$i<@arr_file;$i++)
{
print"@arr_file[$i]\n";
@name=(split(/\./,@arr_file[$i]))[0];
}
for($i=0;$i<@arr_file;$i++)
{
        chomp @arr_file[$i];
        print"@arr_file[$i]\n";
        mkdir @name;
        for($j=1; j<=3; $j++)
        {
                system("vina --config conf.txt --ligand @arr_file[$i] --log @name/@arr_file[$i]_[$j]_log.log");
        }
}
```

A.2 Output Script

```bash
#!/bin/bash

echo 'num, zincid, filename,mode,affinity,dist_rmsd,best_mode_rmsd' | tee -a output.csv

ligfile="$1"
while read -r line; do
        name=$(echo "$line" | cut -f 1 -d '.')
        echo "$name"
        for i in $name/*.log; do
                echo -n $name, | tee -a output.csv
                id=$(grep -A1 "<zinc_id>" $name.sdf | tail -n 1)
                echo -n $id, | tee -a output.csv
                echo -n $i | tee -a output.csv
                result=$(grep -A3 'mode' $i | tail -n 1 | tr -s " " | sed 's/ /, /g')
                if [ -z "$result" ] ; then echo 'NA,NA,NA,NA'| tee -a output.csv; else echo $result | tee -a output.csv; fi
        done
done < "$ligfile"
```

A.3 Cluster Scripts

```bash
#!/bin/bash

ligfile="$1"
while read -r line; do
        name=$(echo "$line" | cut -f 1 -d '.')
        echo "$name"
        cd $name
        for ((i=79; i<101; i++)); do
                echo "$i"
                vina --config ../conf_6m0j_fullgrid.txt --ligand $line --out [$name]_[$i]_out.pdbqt --log [$name]_[$i]_log.log
        done
        cd ..
done < "$ligfile"
```

```
#!/bin/bash

jobid="$1"

cat > job_$1.sbatch << "END"

#SBATCH --time=16-0
#SBATCH --ntasks-per-node=1
#SBATCH --cpus-per-task=1
#SBATCH --mem-per-cpu=40G
#SBATCH --partition=cpu-long.q
#SBATCH --output=slurm_$1.out
#SBATCH --job-name=serialvina_$1
#SBATCH --mail-type=END,FAIL
#SBATCH --mail-user=grgupta@chapman.edu

module load anaconda3/current
source /cm/shared/apps/anaconda3/etc/profile.d/conda.sh
conda activate myprojenv

(time ./vina_iterate.sh $1.txt) &> $1_time.txt

END
```

```
#!/usr/bin/env bash

#SBATCH --time=16-0
#SBATCH --ntasks-per-node=1
#SBATCH --cpus-per-task=1
#SBATCH --mem-per-cpu=40G
#SBATCH --partition=cpu-long.q
#SBATCH --output=slurm_zinc_1.out
#SBATCH --job-name=zincvina_x1
#SBATCH --mail-type=END,FAIL
#SBATCH --mail-user=grgupta@chapman.edu

module load anaconda3/current
source /cm/shared/apps/anaconda3/etc/profile.d/conda.sh
conda activate myprojenv

(time ./vina_iterate.sh zinc_1.txt) &> zinc_1_time.txt
```