Chapman University

Chapman University Digital Commons

Computational and Data Sciences (MS) Theses

Dissertations and Theses

Summer 8-2021

Drug Repurposing for COVID-19 Using Molecular Docking Tools

Deniz Yasar Oztas

Chapman University, oztas@chapman.edu

Follow this and additional works at: https://digitalcommons.chapman.edu/cads_theses

Part of the [Bioinformatics Commons](#), and the [Computational Biology Commons](#)

Recommended Citation

D. Y. Oztas, "Drug repurposing for COVID-19 using molecular docking tools," M. S. thesis, Chapman University, Orange, CA, 2021. <https://doi.org/10.36837/chapman.000286>

This Thesis is brought to you for free and open access by the Dissertations and Theses at Chapman University Digital Commons. It has been accepted for inclusion in Computational and Data Sciences (MS) Theses by an authorized administrator of Chapman University Digital Commons. For more information, please contact laughtin@chapman.edu.

Drug Repurposing for COVID-19 Using Molecular Docking Tools

A Thesis by

Deniz Yasar Oztas

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

August 2021

Committee in charge:

Gennady M. Verkhivker, Ph.D., Chair

Cyril Rakovski, Ph.D.

Rakesh Tiwari, Ph.D.

CHAPMAN UNIVERSITY
SCHMID COLLEGE OF SCIENCE AND TECHNOLOGY

Computational and Data Sciences

The thesis of Deniz Yasar Oztas is approved.

Gennady M. Verkhivker, Ph.D., Chair

Cyril Rakovski, Ph.D.

Rakesh Tiwari, Ph.D.

May 2021

Drug Repurposing for COVID-19 Using Molecular Docking Tools

Copyright © 2021

by Deniz Yasar Oztas

ACKNOWLEDGEMENTS

I would like to thank to my supervisor Dr. Gennady Verkhivker, who has been always there for me to support me and to encourage me to succeed. It was a great opportunity to work with him. I am very grateful for his guidance, feedbacks and tolerance. I would like to thank Dr. Hesham El-Askary, Dr. Cyril Rakovski, and Dr. Rakesh Tiwari for their help and support.

I would like to especially thank Levent Akkok and Melis Gencel for always being supportive, for their help and motivation; I wouldn't have finished this thesis without them.

A special thanks goes to my best friends Ecem Cayiroglu, Muge Gulbay, Selvin Selvi, Engin Pehlivanoglu, Ibrahim Ayberk, and Onat Sagiroglu for their moral support, tolerance and help from the beginning to the end.

I would like to thank Sina Mohammadi, Duc Le, Riley Kendall, Mohammed Alshahrani, Dante Velasquez, Lily Foster, Matthew Nwerem, Luke Burns, and Ryan Richardson for their friendship and making my time here more enjoyable.

Finally, I am very thankful to my father Osman Öztaş, my mother Nihal Öztaş, and my sister Nergis Öztaş for their encouragement and motivation. Without their unwavering support I would have never made it.

ABSTRACT

Drug Repurposing for COVID-19 Using Molecular Docking Tools

by Deniz Yasar Oztas

Since severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is highly contagious and mortal, finding a treatment is time critical. Drug repurposing is probably the quickest and safest approach in our arsenal. However, testing every drug in a brute force manner would require a lot of resources, and a more sophisticated method is required to filter possible candidates. Since several molecules have already been shown to be effective against SARS-CoV-2 in wet-lab experiments, choosing drugs with similar characteristics would increase our chances of success. In this study, we compare the molecular docking results of FDA-approved drugs from the ZINC database against the molecules with positive experimental results. AutoDock Vina was used to dock the molecules against the SARS-CoV-2 spike receptor bound to the ACE2 receptor (6M0J). Results were pre-filtered to 50 candidates according to their binding affinities and the 10 most promising molecules that have similar interactions with the experimental drugs were identified. Then, the 10 molecules were docked against B.1.1.7, B.1.351, and P.1 variants, and their inhibition potentials were discussed. According to the results, we conclude that some molecules that inhibit the wild type also have the potential to inhibit the variants as well. However, further experimental and clinical studies are needed.

TABLE OF CONTENTS

	<u>Page</u>
ACKNOWLEDGEMENTS	IV
ABSTRACT.....	V
LIST OF TABLES	VIII
LIST OF FIGURES	IX
LIST OF ABBREVIATIONS	XIII
LIST OF SYMBOLS	XV
CHAPTER 1 - AN OVERVIEW OF SARS-COV-2.....	1
1.1 Introduction.....	1
1.2 Coronaviruses	1
1.3 Structure of Coronavirus.....	3
1.4 Entering Mechanism of SARS-CoV-2.....	6
1.5 Receptor Binding Domain	7
1.6 Diagnosis and Treatment	8
CHAPTER 2 - COMPUTER AIDED DRUG REPURPOSING AND ITS APPLICATIONS FOR SARS-COV-2	11
2.1 Introduction.....	11
2.2 Molecular docking algorithms	16
2.3 Scoring Functions	19
2.4 Drug Repurposing.....	21
2.5 Molecular Docking and Drug Repurposing against SARS-CoV-2	23
2.6 Method	25
CHAPTER 3 - <i>IN SILICO</i> STUDY OF THE EXPERIMENTALLY PROVED SARS-COV- 2 INHIBITORS AND USE OF LIGAND EFFICIENCY TO VALIDATE THE MOLECULAR DOCKING METHOD	28
3.1 Introduction.....	28
3.2 Method	29
3.3 Results and Discussion	30
3.4 Conclusion	45
CHAPTER 4 - IDENTIFYING DRUG REPURPOSING CANDIDATES FOR SARS-COV- 2 FROM FDA-APPROVED DRUGS USING MOLECULAR DOCKING TOOLS AND INTERACTION ANALYSIS.....	47
4.1 Introduction.....	47

4.2 Method	48
4.3 Results and Discussion	54
4.4 Conclusion	67
CHAPTER 5 - A COMPARATIVE ANALYSIS OF THE SELECTED DRUG REPURPOSING CANDIDATES AGAINST SARS-COV-2 VARIANTS USING MOLECULAR DOCKING TOOLS.....	71
5.1 Introduction.....	71
5.2 Results and Discussion	72
5.3 Conclusion	87
CONCLUSION	89
REFERENCES.....	95
APPENDICES	117

LIST OF TABLES

	<u>Page</u>
Table 1. IC50 Values, and Autodock Vina results of experimental molecules.	32
Table 2. Ligand Efficiency Results.....	34
Table 3. Molecular Structures and Interactions of Experimental Molecules.....	44
Table 4. Top 50 FDA approved drugs and their binding affinities.....	49
Table 5. Top 10 most occurring similar molecules.....	51
Table 6. Molecular Structures and Interactions of top 10 FDA approved drug molecules.	54
Table 7. Amino acid mutations of the variants.	72
Table 8. Molecular Interactions of the top 10 candidates for wild type and the variants.	73

LIST OF FIGURES

	<u>Page</u>
Figure 1. Illustration of the Structure of SARS-CoV-2.	4
Figure 2. Illustration of the entry mechanism of SARS-CoV-2	6
Figure 3. Illustration of basic procedure for molecular docking.	15
Figure 4. Illustration of the summary of the drug discovery.	22
Figure 5. Molecular structures of (A) DRI-C21041, (B) DRI-C23041, (C) DRI-C24041, (D) DRI-C2204745, (E) DRI-C61041, (F) DRI-C71041, (G) DRI-C91005, (H) Methylene Blue, and (I) Erythrosine B.	29
Figure 6. Autodock Vina Binding Affinity results and the Autodock 4.2 rescoring results of 6M0J and spike protein structures.	31
Figure 7. Ligand Efficiencies of Only Spike Structure and their correlation.	34
Figure 8. Ligand Efficiencies of Whole Structure (6M0J) and their correlation.	35
Figure 9. (A) DRI-C21041 on 6M0J (B) Molecular Interactions between DRI-C21041 and 6M0J	36
Figure 10. (A) DRI-C23041 on 6M0J (B) Molecular Interactions between DRI-C23041 and 6M0J	37
Figure 11. (A) DRI-C24041 on 6M0J (B) Molecular Interactions between DRI-C24041 and 6M0J	38
Figure 12. (A) DRI-C61041 on 6M0J (B) Molecular Interactions between DRI-C61041 and 6M0J	39
Figure 13. (A) DRI-C71041 on 6M0J (B) Molecular Interactions between DRI-C71041 and 6M0J	40

Figure 14. (A) DRI- C2204745 on 6M0J (B) Molecular Interactions between DRI- C2204745 and 6M0J.....	41
Figure 15. (A) DRI-C91005 on 6M0J (B) Molecular Interactions between DRI-C91005 and 6M0J	42
Figure 16. (A) EryB on 6M0J (B) Molecular Interactions between EryB and 6M0J.....	43
Figure 17. (A) MeBlu on 6M0J (B) Molecular Interactions between MeBlu and 6M0J	44
Figure 18. Grid box region which is used for molecular docking.	48
Figure 19. Comparison of Autodock Vina and Autodock 4.2 Rescoring Results.	50
Figure 20. (A) 6M0J Structure docked with Vorapaxar (B) Binding affinity distribution of 100 iterations.....	53
Figure 21. (A) 6M0J Structure docked with Voxilaprevir (B) Binding affinity distribution of 100 iterations (C) Molecular interactions of Voxilaprevir.	56
Figure 22. (A) 6M0J Structure docked with Sirolimus (B) Binding affinity distribution of 100 iterations (C) Molecular interactions of Sirolimus.	57
Figure 23. (A) 6M0J Structure docked with Paritaprevir (B) Binding affinity distribution of 100 iterations (C) Molecular interactions of Paritaprevir.	58
Figure 24. (A) 6M0J Structure docked with Cabazitaxel (B) Binding affinity distribution of 100 iterations (C) Molecular interactions of Cabazitaxel.	59
Figure 25. (A) 6M0J Structure docked with Simeprevir (B) Binding affinity distribution of 100 iterations (C) Molecular interactions of Simeprevir.	60
Figure 26. (A) 6M0J Structure docked with Saquinavir (B) Binding affinity distribution of 100 iterations (C) Molecular interactions of Saquinavir.....	62

Figure 27. (A) 6M0J Structure docked with Glecaprevir (B) Binding affinity distribution of 100 iterations (C) Molecular interactions of Glecaprevir.	63
Figure 28. (A) 6M0J Structure docked with Irinotecan (B) Binding affinity distribution of 100 iterations (C) Molecular interactions of Irinotecan.	64
Figure 29. (A) 6M0J Structure docked with Indacaterol (B) Binding affinity distribution of 100 iterations (C) Molecular interactions of Indacaterol.	66
Figure 30. (A) 6M0J Structure docked with Ergotamine (B) Binding affinity distribution of 100 iterations (C) Molecular interactions of Ergotamine.	67
Figure 31. (A) 6M0J Structure docked with Elbasvir (B) Binding affinity distribution of 100 iterations.	69
Figure 32. Binding affinity distributions and overlayed docking poses through 100 iteration of Voxilaprevir for (A) Wild Type, (B) B.1.1.7 Variant (C) B.1.351 Variant, and (D) P.1 Variant.	77
Figure 33. Binding affinity distributions and overlayed docking poses through 100 iteration of Sirolimus for (A) Wild Type, (B) B.1.1.7 Variant (C) B.1.351 Variant, and (D) P.1 Variant	78
Figure 34. Binding affinity distributions and overlayed docking poses through 100 iteration of Paritaprevir for (A) Wild Type, (B) B.1.1.7 Variant (C) B.1.351 Variant, and (D) P.1 Variant.	79
Figure 35. Binding affinity distributions and overlayed docking poses through 100 iteration of Cabazitaxel for (A) Wild Type, (B) B.1.1.7 Variant (C) B.1.351 Variant, and (D) P.1 Variant.	80

Figure 36. Binding affinity distributions and overlayed docking poses through 100 iteration of Simeprevir for (A) Wild Type, (B) B.1.1.7 Variant (C) B.1.351 Variant, and (D) P.1 Variant	81
Figure 37. Binding affinity distributions and overlayed docking poses through 100 iteration of Saquinavir for (A) Wild Type, (B) B.1.1.7 Variant (C) B.1.351 Variant, and (D) P.1 Variant	83
Figure 38. Binding affinity distributions and overlayed docking poses through 100 iteration of Glecaprevir for (A) Wild Type, (B) B.1.1.7 Variant (C) B.1.351 Variant, and (D) P.1 Variant.....	84
Figure 39. Binding affinity distributions and overlayed docking poses through 100 iteration of Irinotecan for (A) Wild Type, (B) B.1.1.7 Variant (C) B.1.351 Variant, and (D) P.1 Variant.	85
Figure 40. Binding affinity distributions and overlayed docking poses through 100 iteration of Indacaterol for (A) Wild Type, (B) B.1.1.7 Variant (C) B.1.351 Variant, and (D) P.1 Variant	86
Figure 41. Binding affinity distributions and overlayed docking poses through 100 iteration of Ergotamine for (A) Wild Type, (B) B.1.1.7 Variant (C) B.1.351 Variant, and (D) P.1 Variant.....	87

LIST OF ABBREVIATIONS

SARS	Severe Acute Respiratory Syndrome
MERS	Middle East Respiratory Syndrome
ACE2	Angiotensin-Converting Enzyme 2
MeBlu	Methylene Blue
EryB	Erythrosine B
PDB	Protein Data Bank
Kb	Kilobase
kDa	Kilodalton
Ala (A)	Alanine
Arg (R)	Arginine
Asn (N)	Asparagine
Asp (D)	Aspartic Acid
Cys (C)	Cysteine
Glu (E)	Glutamic Acid
Gln (Q)	Glutamine
Gly (G)	Glycine
His (H)	Histidine
Ile (I)	Isoleucine
Leu (L)	Leucine
Lys (K)	Lysine
Met (M)	Methionine

Phe (F)	Phenylalanine
Pro (P)	Proline
Ser (S)	Serine
Thr (T)	Threonine
Trp (W)	Tryptophan
Tyr (Y)	Tyrosine
Val (V)	Valine

LIST OF SYMBOLS

Å Angstrom

Chapter 1 - An Overview of SARS-CoV-2

1.1 Introduction

2020 was a year of great financial distress, product shortages, stay-home orders, panic across widespread populations and possibly lasting changes in the way people conduct their lives. The reason for these events was a virus which first appeared in Wuhan, China in late 2019. Everything started as a local outbreak and quickly spread to the whole world. The virus is highly infectious and mainly spreads through contact/droplet transmission. It can cause pneumonia, Acute Respiratory Distress Syndrome (ARDS), and other complications (Guan, Ni et al. 2020). Its origins remain unknown. However, due to its genetic similarities to the Severe acute respiratory syndrome (SARS), the virus was named as SARS-CoV-2 (Petersen, Koopmans et al. 2020). On March 11th, the World Health Organization (WHO) declared the disease a global pandemic and a global effort to battle the virus followed.

In the year following the WHO declaration, several vaccines were developed, got approvals from regulators around the world and started to be administered (Forni and Mantovani 2021). However, as of right now, there is still no official treatment for the infected patients. Since a large portion of the global population currently does not have access to the vaccinations and the mutations the virus is undergoing, reaching herd immunity might be an elusive goal. Therefore, finding a treatment is still an important and time-critical task.

1.2 Coronaviruses

Coronaviruses are a large family of viruses which belong to the Nidovirales. Nidovirales is an order of single-stranded, positive-strand RNA viruses with envelopes. Nidovirales order includes four families which are *Coronaviridae*, *Mesoniviridae*, *Roniviridae*, and *Arteriviridae* (Enjuanes, Gorbalenya et al. 2008). The Coronavirinae is one of two subfamilies of the Coronaviridae family,

and it is divided into four genera which are alpha, beta, gamma, and delta coronaviruses (Payne 2017). According to the genomic characterization, while the alpha and beta coronaviruses probably originated from bats and rodents, delta and gamma coronaviruses originated from avian species.

Coronaviruses can infect various species such as camels, cats, and bats. They can cause various types of diseases such as respiratory, hepatic, and neurological. They can jump from animal species to humans, albeit usually through an intermediate species (Campana, Parisi et al. 2020). So far, seven coronaviruses that can infect humans were discovered. These seven coronaviruses are known to belong to either alpha or beta coronaviruses genera (Ye, Yuan et al. 2020). Of the seven, three have caused lethal outbreaks: SARS-CoV, MERS-CoV, and SARS-CoV-2. While the first two were zoonotic outbreaks, the origin of the third outbreak remains unknown. There is some evidence of a zoonotic link, but the evidence is not conclusive (Campana, Parisi et al. 2020).

SARS-CoV was first seen 2003 and had a 9% mortality rate (Sørensen, Sørensen et al. 2006). It bound to angiotensin-converting enzyme 2 (ACE2) on the cell membrane and caused pneumonia, which led to acute respiratory distress syndrome (ARDS). It spread to multiple countries but was contained quickly. A total of 8098 people got sick, and 774 of these people died (Hui and Zumla 2019).

MERS-CoV, or Middle East respiratory syndrome, was first seen in 2012 and had a 35% mortality rate (Alsolamy and Arabi 2015). Unlike SARS-CoV, MERS-CoV bound to Dipeptidyl peptidase-4 (DPP4) on the cell membrane and caused ARDS as well. It is a rarely seen disease that does not spread quickly. So far, a total of 2519 people got sick, and 866 died (Organization 2019).

SARS-CoV-2 is similar to SARS-CoV. It was first seen in 2019 and has a 3.4% mortality rate worldwide (Roussel, Giraud-Gatineau et al. 2020). It also binds to angiotensin-converting enzyme

2 (ACE2) on the cell membrane and, like the other two viruses, causes pneumonia, which leads to acute respiratory distress syndrome (ARDS). It is much more contagious and, at the time of writing, was not contained. So far, a total of 141 million people got sick, and 3.01 million people died.

1.3 Structure of Coronavirus

Coronaviruses have highly conserved genomic organization, and the genome of these viruses encode structural proteins, non-structural proteins, and accessory proteins (Artika, Dewantari et al. 2020). Coronaviruses have a single-stranded RNA genome which is very large (up to 33.5 kb genomes) (Fehr and Perlman 2015). The genome is structured like an mRNA; it has a 5' cap and a 3' poly (A) tail. Therefore, it can utilize the host cell for translation of its replicase gene (up to 20kb), which encodes two polyproteins that get cleaved into the non-structural proteins (nsps). These nsps play various important roles like blocking interferons, host immune proteins that the cell produces as an immune response (Kamitani, Narayanan et al. 2006, Kamitani, Huang et al. 2009). They also facilitate the viral RNA replication and sub-genomic transcription of RNA, which creates the structural proteins. Structural proteins form the final virions and facilitate the exit from the host cell. They are necessary to protect the genome and to infect the next host cell (Tanaka, Kamitani et al. 2012). The third group of proteins, accessory proteins, are also produced during the infection (Zhao, Jha et al. 2012). These proteins do not play a part in the replication but are required for viral pathogenesis.

Coronaviruses are spherical viruses, and they are approximately 125 nm in diameter. They are composed of four main structural proteins, membrane (M), envelope (E), nucleocapsid (N), and spike (S) proteins (Fehr and Perlman 2015). All of these proteins are encoded within the 3' end of

the viral genome. Because of the spike glycoproteins on its envelope, under an electron microscope, SARS-COV-2 looks like a crown. It is demonstrated in Figure 1.

Figure 1. Illustration of the Structure of SARS-CoV-2.

The Membrane protein is approximately 25–30 kDa and has three transmembrane domains (Armstrong, Niemann et al. 1984). It is the most expressed structural molecule and gives the viruses their shapes. The M protein has an N-terminal glycosylated ectodomain and a C-terminal endodomain. M protein interacts with all the other structural proteins and organizes the assembly process. According to the studies, it is a functionally dimeric protein. It is suggested that M protein can take two conformations. Membrane curvature is regulated to play different roles during different assembly stages. It captures other structural proteins and gets the virion ready for budding (Neuman, Kiss et al. 2011).

The Envelope protein is approximately 8–12 kDa (Schoeman and Fielding 2019). It is expressed in high quantities, but only a small portion is found in the virions. It has an N-terminal ectodomain and a C-terminal endodomain and also has ion channel activity. This ion channel activity is required for pathogenesis (Nieto-Torres, DeDiego et al. 2014). The E protein is a small

transmembrane protein and is responsible for the assembly & release of the virus. E protein is heavily involved in the budding process (Li, Surya et al. 2014).

The Nucleocapsid protein is approximately 45 kDa (Fan, Ooi et al. 2005). It has an N-terminal domain and a C-terminal domain, and both domains can individually bind to RNA through different mechanisms. However, for optimal binding, contributions from both are required. N protein is heavily phosphorylated (Stohlman and Lai 1979). Phosphorylation is suggested to help increase the N proteins' affinity towards the viral RNA. N protein binds to the viral RNA genome and protects it by forming a shell around it (Hurst, Koetzner et al. 2013, Kuo and Masters 2013). During the assembly phase, N protein binds to the C-terminal RNA binding domain, M protein, nsp3, and a non-structural protein from the replicase complex. It ties the viral RNA genome to the replicase–transcriptase complex (RTC), and then helps package the genome into the viral envelope.

The last piece, S protein, is approximately 150 kDa (Beniac, Andonov et al. 2006). It distinctively exists on the surface of the virus. The S protein consists of a large ectodomain, a small endodomain, and a single transmembrane anchor (Wang, Grunewald et al. 2020). The ectodomain consists of two subunits, S1 and S2. During the infection phase, S protein is cleaved by a host furin-like protease, a cell surface protein, into S1 and S2. S1 subunit binds to the receptor, and S2 subunit handles the membrane fusion, which leads to the viral fusion and entry into the host cell (Bosch, Van der Zee et al. 2003).

1.4 Entering Mechanism of SARS-CoV-2

Understanding the entry mechanism of the SARS-CoV-2 into human cells is vital for designing therapies against it. Like all coronaviruses, SARS-CoV-2 enters the cells via the S protein (Shang, Wan et al. 2020). The mechanism of entry of SARS-CoV-2 is demonstrated in Figure 2.

Figure 2. Illustration of the entry mechanism of SARS-CoV-2

The infection starts with the S1 subunit of the trimeric spike protein binding to the ACE2 receptor, which is a type I membrane protein expressed in lungs, heart, kidneys, and intestine cell surfaces (Hamming, Timens et al. 2004). Unlike other SARS viruses, The SARS-CoV-2 spike protein has a proprotein convertase (PPC) cleavage site, which separates S1 and S2 (Walls, Park et al. 2020). The human type I transmembrane protein furin and transmembrane serine protease 2 (TMPRSS2) on the host cell membrane cleave the S protein (Bestle, Heindl et al. 2020). The resulting S1 subunit dissociates, and S2 subunit transitions to a post-fusion conformation which helps fuse the

host cell and viral membranes. The post-fusion S2 domain, which drives this process, has a peptide region that can be activated by receptor recognition; and once activated it can interact with the host cell membrane (Xiu, Dick et al. 2020).

1.5 Receptor Binding Domain

The area on the SARS-CoV-2 that binds to the ACE2 inhibitor is called RBD, receptor-binding domain. This area and its interactions with ACE2 were studied thoroughly using X-ray crystallography and cryo-electron microscopy (cryo-EM) (Song, Gui et al. 2018). There are multiple 3D structures in the Protein Data Bank, which illustrate this region. Some of the commonly used ones are: PDB: 6M17, PDB: 6M0J, PDB: 6VW1, PDB: 6LZG. While the 6M17, 6M0J, 6VW1, and 6LZG structures represent the crystal structure of the spike protein that is bound to ACE2, 6XEY represents the crystal structure of the SARS-CoV-2 spike glycoprotein that is bound to antibodies (Lan, Ge et al. 2020, Liu, Wang et al. 2020, Shang, Ye et al. 2020, Wang, Zhang et al. 2020, Yan, Zhang et al. 2020).

According to the studies, spike protein and ACE2 make a significant number of hydrogen and hydrophobic bonds. It has been shown that Lys417, Gly446, Tyr449, Tyr453, Leu455, Phe456, Gln474, Ala475, Gly476, Phe486, Asn487, Tyr489, Gln493, Gly496, Gln498, Thr500, Asn501, Gly502, and Tyr505 are the binding residues of SARS-CoV-2, which form hydrogen bonds. On the other side, Ser19, Gln24, Thr27, Phe28, Asp30, Lys31, His34, Glu35, Glu37, Asp38, Tyr41, Gln42, Leu45, Leu79, Met82, Tyr83, Asn330, Lys353, Gly354, Asp355, Arg357, and Arg393 are the contact residues for ACE2 (Shang, Ye et al. 2020, Yan, Zhang et al. 2020).

SARS-CoV-2 binds to the ACE2 receptor with a higher affinity than SARS-CoV (Shang, Ye et al. 2020). However, it also has defenses against the host immune system. Spike protein has two conformations: laying down and standing up. Unlike SARS-CoV, it spends most of its time in the

down setting, which helps it evade the immune response. However, certain host proteases cause it to adopt the standing up conformation, and in this form, it can identify the ACE2 receptor and bind to it with high affinity (Campana, Parisi et al. 2020, Walls, Park et al. 2020).

1.6 Diagnosis and Treatment

For most patients, diagnosis and treatment of SARS-CoV-2 are unnecessary. Patients get common cold symptoms like fever, cough, fatigue, and their immune systems handle the infection without any major incident. Unfortunately, some of the other patients, especially older patients and patients with underlying conditions, get severe symptoms like shortage of breath, chest pains, and confusion, which can lead to hospitalizations and even death. Since the virus is highly infectious, it can easily infect large populations and cause the health services to get overwhelmed by the number of patients with severe symptoms. Therefore, early diagnosis of disease and isolation of the patients from the rest of the population is critical.

Since the symptoms of SARS-CoV-2 show high variance and can be mixed up with other diseases such as flu; radiological imaging is required to diagnose the disease with high prediction rates (Hosseiny, Kooraki et al. 2020). Currently, there are multiple classes of tests available for SARS-CoV-2 screening: serological tests (ex: Elisa) that detect antibodies in the bloodstream, molecular tests (ex: RT-PCR) that isolate and amplify certain parts of the virulent RNA, and nanotechnology-based tests that use nanoparticles to bind to the antigens (Emery, Erdman et al. 2004, Beavis, Matushek et al. 2020, Singhal, Gupta et al. 2020). Of these various methods, PCR has become the most used one (Waller, Kaur et al. 2020). However, serologic assays are also important for collecting data for epidemiological studies.

Currently, the best way to combat SARS-CoV-2 is prevention. Unfortunately, isolation of sick patients, social distancing, and better hygiene have not been enough to contain the disease, and especially with the new variants of the virus showing up, herd immunity might be hard to achieve. Therefore, there is a global campaign to vaccinate people. There are currently multiple vaccines in the market, and more are in the pipeline. These vaccines belong to 5 distinct categories: mRNA vaccines (Biontech and Moderna), replication-defective viral vector vaccines (AstraZeneca, JNJ), inactivated pathogen vaccines (Sinovac), protein subunit vaccines (Novavax) and virus like particle vaccines (Medicago) (Kyriakidis, López-Cortés et al. 2021). Each of these families has cons and pros, but none are 100% effective against all variants, and there is not enough supply of them for the entire population.

Since prevention goes only so far, treatments are required to treat patients with severe symptoms. To this date, no official therapeutics that were specifically designed for SARS-CoV-2 have been approved. There are currently multiple drug therapies being used against the virus. However, these act as supplemental aid that relieves the symptoms and shortens the recovery time. Remdesivir, Lopinavir - Ritonavir, Favipiravir, Ribavirin are some of the commonly used drugs. Remdesivir and Favipiravir are RNA polymerase inhibitors (Du and Chen 2020, Jorgensen, Kebriaei et al. 2020). Lopinavir - Ritonavir are protease inhibitors (Horby, Mafham et al. 2020). And Ribavirin is a guanosine analog (Thomas, Ghany et al. 2012). These are mostly anti-viral drugs that are designed to treat other viral diseases. However, since they interfere with different stages of the viral replication, they limit the symptoms and help the recovery time.

Another group of commonly used drugs is immunosuppressants such as interferon β -1b and Interleukin-6 Inhibitors (Russell, Moss et al. 2020). These are used to treat the cytokine storm that is caused by the host's immune response. When a host's innate immune system gets activated, it

causes the production of cytokines, pro-inflammatory molecules. Under normal circumstances, this is the normal response to a viral infection. However, SARS-CoV-2 can cause an overreaction, where the system goes into a positive feedback loop and creates too much cytokines. This exaggerated immune response can cause tissue damage and cause ARDS (Ragab, Salah Eldin et al. 2020). This group of drugs are used to limit the host's immune response and avoid organ failure.

The last option currently being used is treatment with plasma from recovered SARS-CoV-2 patients (Duan, Liu et al. 2020). Once a patient recovers from the disease, his blood carries antibodies that were used to battle the infection. These antibodies can be harvested and given to another patient who has severe symptoms. Although this method is the most effective amongst all the treatments, it is not scalable.

Since no highly effective and scalable treatment exists yet, new drug candidates need to be found and studied. However, a lead drug takes a very long time to develop. Therefore, strategies like repurposing existing drugs can be used. To treat the disease instead of its symptoms, the entry of the virus into the cells can be blocked. This can be managed by interfering with the ACE2-RDB interaction.

Chapter 2 - Computer Aided Drug Repurposing and Its Applications for SARS-CoV-2

2.1 Introduction

Our bodies are extremely complex machines. At any given time, many different molecules interact with each other and help the system keep functioning. When an external factor, such as SARS-CoV-2, interferes with the system, finding a targeted remedy is no small task. The possible drugs space is infinitely large and when a molecule is being considered as a treatment, its interactions with the other molecules in the system have to be safe for the host. There are some guidelines that rely on heuristics, such as Lipinski's rule of five, to make the process manageable (Lipinski, Lombardo et al. 1997). But even with the restricted input space and simplified constraints, in order to discover a successful treatment, a lot of experiments need to be performed. Since wet-lab experiments are both costly and time-consuming, filtering out improbable candidates would increase efficiency and decrease development time. This is why computational methods are being developed and widely used.

As a first step, computational methods can be used in conjunction with bioinformatics to identify disease-causing protein sequences and predict their structures. Then these structures can be computationally analyzed according to their structural and molecular properties for druggable pockets, where an agent can bind with high affinity. After the druggability of the target is verified, target drug molecules can be screened, and promising ones can be filtered. Later these promising drug molecules can be further optimized and filtered using structure-based optimizations and Quantitative Structure-Activity Relationship (QSAR) models (Kwon, Bae et al. 2019). Once the drug molecule is optimized, its toxicity, pharmacokinetics, and polypharmacology can be studied.

There are three major computer-aided drug design strategies: ligand-based, system-based, and structure-based.

A ligand is a molecule which binds to a biomolecule and causes a biochemical change (Mobley and Dill 2009). Ligands that trigger the receptor and cause a biological response are called agonists and ligands which block the receptor are called antagonists (Berg and Clarke 2018). In the ligand-based approach the chemical properties and the structure of a compound is known (Aparoy, Kumar Reddy et al. 2012). This knowledge is then used to find similar compounds that bind to a target macromolecule or to design/identify other molecules that have enhanced properties or to identify target receptors. In the first two approaches, chemical similarity is used, and similar ligands are considered. The similarity is usually measured by using features called fingerprints. Fingerprints are vectors that define a molecule in terms of paths or substructures. Once molecules are quantified, distance metrics, such as the Tanimoto index, can be used to quantify the similarity between them. Using this approach, similar molecules which might be more favorable could be identified (Lo, Gui et al. 2016). On the other hand, in the third approach, biochemical properties of ligands, such as MolMR, MolLogP, TPSA, can be used to predict their affinities against a library of proteins (Boyles, Deane et al. 2020).

In the system-based approach, a more holistic view is taken. Interactions of the proteins, genomes, pathways, and their responses to certain chemical signals are taken into account (Draghici, Khatri et al. 2007). Computational methods are used to find insights on the systems level, and instead of a minimalist approach that concentrates on molecule pairs, a unified approach is taken. Interactions are analyzed as a network of events and drugs are targeted to modulate that network. In this approach certain drugs that have certain functions can be used to treat other diseases where similar functions are still effective.

In the structure-based approach, the structure of the target protein is known (Batool, Ahmad et al. 2019). This structure may be identified in the wet lab using techniques like multi-dimensional nuclear magnetic resonance (NMR) and X-Ray diffraction technique or it may be a product of molecular modeling and simulations. The first step in this approach is finding a potential receptor site. Target sites usually have certain features like having potential hydrogen donors and acceptors. They can be ridges or pockets on the surface, which act as assembly, communication, or active sites. These sites are identified either through wet-lab experiments, by using homology models or by identifying sites that would interact with functional groups that drug-like molecules usually carry (Laurie and Jackson 2005). The next step is looking for hits, and it can be done in two different ways: *de-novo* design and virtual screening.

In the *de novo* method, molecules are built from scratch according to the properties of the binding site (Lin, Li et al. 2020). Atoms or frequently used molecule fragments such as rings can be stitched together to build molecules that interact well with the binding domain (Reutlinger, Rodrigues et al. 2014). One or multiple reference molecules can also be used as a point, and their fragments can be linked to other scaffolds. *De novo* drug design is an area where machine learning methods like VAE, GAN, and genetic algorithms can be used (Mouchlis, Afantitis et al. 2021).

In the second method, virtual screening, a library of molecules are tested against the structure of interest (Lyne 2002). Ligands and the structure interact due to the intermolecular forces between them and can form hydrogen, ionic, hydrophobic or Van Der Waals bonds. The measure of the strength of the interactions between the ligand and the structure is called binding affinity, and it is used to compare and rank the ligands (Huang, Bonnichon et al. 2017). A lower binding affinity means a stronger attraction and a higher binding affinity means a weaker attraction. Binding affinity is usually expressed by the equilibrium dissociation constant (KD).

KD is calculated by

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

where [L] is the molar concentration of the ligand, [P] is the molar concentration of the protein, and [LP] is the molar concentration of the complex (Kastritis and Bonvin 2013). In order to have a sense of the binding affinity between the molecule and the structure, how they are positioned against each other needs to be known. Same structure and ligand can have very different affinities in different geometries. This is where molecular docking is used.

Molecular docking is a computational method that predicts the most likely pose and location a ligand would take against a structure. Molecular docking tools create thousands of different ligand conformations at the target protein's active site, rank them with a scoring function, and try to find the pose with the lowest binding energy. A low and negative binding energy means that when the ligand binds to the structure, energy is released. This would decrease the total energy of the system and make the system more stable, hence more likely. A simple illustration of molecular docking can be found in Figure 3.

Figure 3. Illustration of basic procedure for molecular docking.

Using large and diverse molecule libraries for screening is a common practice in drug studies (Kodadek 2011). In virtual screening, compounds from databases, such as ZINC, are docked methodically against the target structure and promising ones are identified (Kodadek 2011). These databases are specially curated for virtual screening and contain a significant number of commercial and experimental molecules. In this sense, virtual screening can be considered as the *in silico* version of high-throughput screening (HTS) (Stahura and Bajorath 2004). However, it is a faster and more cost-effective method.

When promising ligands are identified, and their binding modes are known, more insights can be derived about the molecule and how it interacts with ligands. This knowledge can be used to optimize the ligands by modifying them, or even to design new ones. A larger database would enable more diverse candidates and greater insight.

However, the docking method is not without challenges. Since there are too many degrees of freedom in the system, creating a very realistic simulation would be computationally demanding

(Adcock and McCammon 2006). Therefore, simple approximations, such as treating the protein and ligand as separate rigid objects, are used to make the simulations more manageable. This, of course, comes at an accuracy cost (Perez and Tvaroška 2014). There are many molecular docking programs, and they all use different algorithms, assumptions, and approximations. So, they all have their advantages and disadvantages.

2.2 Molecular docking algorithms

In 1894 Fisher proposed that ligands interacted with receptors that had geometric matching structures (Fischer 1894). He used an analogy to describe his thesis: lock and key (Koshland Jr 1958). This model was too simplistic, and in 1958 Koshland proposed an alternative: Induced Fit Theory (Ohue, Shimoda et al. 2014). In this model, both ligands and proteins were flexible and changed shapes after their interaction for a better fit.

There are two types of molecular docking that follow these two models: rigid-body docking and flexible docking. Rigid-body docking is usually done for pre-screening and to identify binding sites. It does not consider the flexibility of the agents. Therefore, it is fast but not accurate. The old versions of DOCK and FLOG softwares use this methodology.

On the other hand, in flexible docking not only translational and rotational axes are considered, but also possible conformations of all agents are included in the simulation as well. This causes a significantly larger search space. Therefore, flexible docking is more computationally intense, but more accurate. There are many variants of these two approaches as well. Most docking tools assume that the ligand is flexible, and the receptor is rigid (Pagadala, Syed et al. 2017). Molecular Dynamics (MD) simulations are better suited for modeling completely flexible ligands and receptors. However, since molecular dynamics simulations are computationally expensive, some semi-flexible methods were attempted (Salmaso and Moro 2018). For example, the Lennard-Jones

8-4 potential in GOLD and smooth potential in AutoDock 3.0 belong to this class (Meng, Zhang et al. 2011). These methods don't have perfect flexibility, but they are computationally more efficient than MD simulations.

Docking tools try to find stable conformations of the ligand-structure complex. To do that they use search algorithms that go through the possible conformations space and try to converge on the minimum energy setting. These can be categorized into two categories: systemic algorithms (exhaustive search algorithms, fragment-based search algorithms) and stochastic algorithms (genetic algorithms and Monte Carlo algorithms) (Sliwoski, Kothiwale et al. 2014).

Exhaustive search algorithms span the whole conformations space with set intervals and consider every possible degree of freedom. However, if the possible conformation space is too large, certain shortcuts can be taken to decrease the required computational power. For example, initially, a lower resolution could be used to identify an area of interest, where a higher resolution search can be continued. Grid-based searching is another option. After voxelization ligands and the protein grids can be transformed with 3D FFT and scored using convolutions, which are computationally more efficient (Ohue, Shimoda et al. 2014). Another strategy that is used to shrink down the search space is using heuristics. Depending on the structure characteristics, certain areas could be picked as low energy candidates. This strategy is employed by GLIDE (Friesner, Banks et al. 2004).

Fragment-based algorithms divide the ligand into smaller building blocks and form the final result by using the results of the fragments. There are a couple of different fragment-based algorithms: incremental construction, distance geometry, and shape matching.

Incremental construction (IC) separates the ligand into fragments and then starts to dock into the target's active site (Rarey, Kramer et al. 1996). The first docked fragment is usually the largest one

or has the most potential against the target. The first docked fragment can be considered as an anchor, and the remaining fragments are added to this anchor incrementally. Therefore, this method is also called the anchor and grow method. Since different poses are generated by construction, this method can handle the flexibility of the ligands as well. IC is used in DOCK, FLOG, and SLIDE (Kuntz, Blaney et al. 1982, Miller, Kearsley et al. 1994, Schnecke and Kuhn 2000).

Matching Algorithm is an approach that is based on molecular shapes (Fischer, Norel et al. 1993). This algorithm tries to maximize the overlap between areas that are derived from the molecules using their molecular shape (& complements) and electrostatics properties. Although Matching algorithms are fast, their accuracies are not the highest (Moitessier, Englebienne et al. 2008). They are often used in conjunction with other methods such as incremental construction and FFT too. For example, DOCK algorithm matches spheres, which are derived from ligand fragments, with the complement of the receptor and builds back the ligand to find the optimal pose (Kuntz, Blaney et al. 1982). Matching Algorithms are also used in LibDock, and FLOG (Miller, Kearsley et al. 1994, Diller and Merz Jr 2001)

Stochastic search algorithms use probability distribution functions to make random changes and move in the direction of changes with positive results. Due to its stochastic nature, this strategy has a better chance to avoid local minima. However, it also requires more computing power for a wide-area search. Monte Carlo and genetic algorithms are the most commonly used stochastic search algorithms.

Monte Carlo (MC) is a stochastic method that randomly modifies the ligand conformation (Hart and Read 1992). MC method starts with a random initial pose of the ligand and scores this pose according to a specific energy-based function. Then using a probabilistic approach, small changes

are made to the pose, and the new pose is rescored. Poses with better scores are kept and poses with worse scores are discarded with respect to the output of another probability function. This operation continues till the target number of conformations are generated and the minimum energy pose is found. Monte Carlo was used in earlier versions of AutoDock (Goodsell and Olson 1990).

Genetic algorithms (GA) is another stochastic method that tries to find the conformation with global minimum energy (Jones, Willett et al. 1997). The idea comes from Darwin's evolution theory. Like the theory, there are mutations, mating, and crossovers in this method as well. The current poses of the ligands and the receptors can be considered as chromosomes. In the same manner, rotation, conformation, and translation parameters of the ligand can be considered as genes. Chromosomes are stochastically modified, paired with other chromosomes, and go through the crossover to produce offsprings. Then these offsprings are evaluated with a scoring function and offsprings with better scores are retained for the next round. Genetic algorithms are used in AutoDock and GOLD (Morris, Goodsell et al. 1998, Verdonk, Cole et al. 2003).

2.3 Scoring Functions

Docking is not only used to find the lowest energy poses of a ligand. It is also used to compare the performances of different ligands during virtual screening. In order to be able to compare the binding affinities of two different ligands a metric called scoring function is used. There are several scoring functions that are used in different docking tools. Scoring functions are mathematical approximation methods for estimating binding affinity. They usually depend on free energy calculations. They need to be consistent and accurate for all types of ligands. Three basic types of scoring functions are used for protein–ligand docking: force-field, empirical, and knowledge-based functions.

Force field-based scoring functions use van der Waals interactions, electrostatic interactions, and torsional forces to calculate the affinities (Huang, Kalyanaraman et al. 2006). Some of the terms used in the function can be tuned with experimental data as well. DOCK force field score is estimated by using the equation,

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

where i and j are atom indexes for the ligand and the protein, r stands for the distance between atoms, A/B stands for van der Waals parameters, and q stands for atomic charges (Weiner, Kollman et al. 1984). Force field-based scoring functions can utilize modern force fields like CHARMM and Amber as well (Debroise, Shakhnovich et al. 2017). Force field-based scoring functions do not consider solvation, and lose precision due to the lack of focus on the non-bond interactions between far away atoms. Force field-based scoring is used by DOCK, Autodock and GOLD (Meng, Shoichet et al. 1992, Jones, Willett et al. 1997, Morris, Goodsell et al. 1998).

Empirical based functions use different intermolecular interaction terms that are estimated by multiple linear regression of binding energies from lab experiments (Liu and Wang 2015). A typical empirical scoring function would look like this:

$$\Delta G = \sum_i W_i \cdot \Delta G_i$$

where ΔG_i represents different energy terms such as VDW energy, electrostatics, hydrogen bond, entropy, hydrophobicity. The corresponding weight coefficients are determined by empirical methods. Binding affinities of a protein–ligand training set is fitted with three-dimensional representations of known structures (Sottriffer, Sanschagrin et al. 2008). The empirical-based

functions are generally faster than force field-based scoring functions (Jain 1996). Empirical scoring is used by Glide, ChemScore, and X-Score (Wang, Lai et al. 2002, Verdonk, Cole et al. 2003, Halgren, Murphy et al. 2004).

Knowledge-based scoring functions are also statistics-based scoring functions. They use statistical observations of intermolecular interactions which are derived from the experimental structures (Tanaka and Scheraga 1976). The score is calculated as the sum of all pairwise contact scores between the ligand and protein atoms. These energies are estimated by analyzing large datasets like the Protein Data Bank (PDB). The pairwise scores are calculated according to the assumptions that if an interaction has a lower energy, it will happen more often (Thomas and Dill 1996). So frequent pairs get higher scores, and rare pairs get penalized. For protein–ligand studies, the potentials are calculated by

$$w(r) = -k_B T \ln[g(r)]$$

$$g(r) = \rho(r) / \rho^*(r)$$

where k_B is the Boltzmann constant, T is the absolute temperature of the system, $\rho(r)$ is the number density of the protein–ligand atom pair at distance r , and $\rho^*(r)$ is the pair density in a reference state where the interatomic interactions are zero. Knowledge-based scoring is used in PMF, ITScore, and DrugScore (Gohlke, Hendlich et al. 2000, Huang and Zou 2006, Huang, Grinter et al. 2010).

2.4 Drug Repurposing

Drug discovery on average takes 10 to 15 years. Most of the tested molecules cannot progress to the advanced stages (DiMasi, Hansen et al. 2003). Since 1995, only 10 percent of molecules in

phase 1 studies actually made it to the clinical phase. That is why computational methods like molecular docking are being adopted at every stage of the drug discovery pipeline and being used to identify and eliminate targets. But even with the computer aid, drug discovery is a long process.

Figure 4. Illustration of the summary of the drug discovery.

One of the strategies to accelerate this process is drug repurposing. Drug repurposing (also known as drug repositioning or reprofiling) is a technique for discovering new uses for existing drugs that go beyond the original medical indication (Rudrapal, Khairnar et al. 2020). Since most of the pipeline can be skipped, this approach ensures a fast time to market, an existing supply chain, lower cost, and in some cases significant amounts of clinical data. Clinical data is especially important as it would be very useful for expediting the pre-clinical and clinical test stages. Also, since 45% of the drug lead failures are due to toxicity and safety issues, this approach would increase the chances of success (Rudrapal, Khairnar et al. 2020).

There are many places where drug repurposing has been used and has been successful. For example, Viagra, one of the most well-known drugs, was originally developed as an antihypertensive drug, but Pfizer has repurposed and marketed this drug for erectile dysfunction (Ghofrani, Osterloh et al. 2006). Another example is repurposing of the cancer drug Zidovudine as an anti-HIV drug. Thalidomide is another famous example. It was developed as a sedative in

the 1950s; but then repurposed as a leprosy and multiple myeloma therapies (Rahmat, Sklavenitis Pistofidis et al. 2019).

However, drug repurposing has its own challenges too. Drugs are usually approved against certain diseases at certain doses. When a drug is repurposed, it needs to go through clinical trials again, and the appropriate doses need to be identified. Another issue is the delivery systems. A repurposed molecule that is targeted to reach a certain tissue might require a different targeting mechanism which would require some of the steps in the pipeline. Also, when a drug is being repurposed, its interactions with other drugs that are used for the new indication need to be studied (Oprea, Bauman et al. 2011). Regardless of these challenges a successful repurposing would fast track a drug and save lives.

The best way to increase possible drug repurposing opportunities is to use computation methods. All three computational methods would be applicable for this strategy. Virtual screening can be done to identify approved drug molecules that interact with a specific target disease causing protein. Or a specific drug molecule can be tested against all the crystal structures in the databases to see if it interacts with any. Molecule libraries can be searched to find druggable, more effective or free analogues of drug targets. Using networks, similarities between diseases and biological processes can be identified, and drugs that work for one case can be tried on the other as well. Molecular docking is used for virtually all these scenarios.

2.5 Molecular Docking and Drug Repurposing against SARS-CoV-2

Despite all the safety measures and vaccinations, at the time of writing SARS-CoV-2 was still spreading quickly and causing a high number of deaths. Therefore, there is a huge worldwide effort to find viable treatments against it. In the past year, a significant number of studies that used

computational methods against SARS-CoV-2 were published. Due to the urgency of the matter, molecular docking and drug repurposing was a common theme across these studies. Some of the suggested repurposing targets have gone through clinical trials and are currently approved in different countries.

Remdesivir, which is used for hepatitis C, is the only FDA-approved drug against SARS-CoV-2. Remdesivir was approved by FDA in October, but before that, it was predicted as a promising candidate by some *in silico* studies (Elfiky 2020, Naik, Munikumar et al. 2020). Later docking studies that were carried out for this drug, also affirmed it as a promising candidate (Deshpande, Tiwari et al. 2020). Finally, in a double-blind, randomized, placebo-controlled clinical study, Remdesivir has been proven to shorten the healing process in adults who were hospitalized due to Covid-19 (Beigel, Tomashek et al. 2020).

Ribavirin is another promising candidate against SARS-CoV-2. It is an antiviral medication which is used to treat hepatitis C and respiratory syncytial virus (RSV) infection. Ribavirin was suggested as a potential drug against SARS-CoV-2 *in silico* studies (Elfiky 2020, Peng, Shen et al. 2021). Later, in another study, *in silico* results were strengthened with *in vitro* and it was observed that Ribavirin showed anti-viral activity (Unal, Bitirim et al. 2020). In another clinical study, it was observed that a treatment including Ribavirin shortened the duration of viral shedding (Hung, Lung et al. 2020).

Arbidol, which is also known as Umifenovir, is another antiviral (anti-influenza) drug. It was suggested that Arbidol may have efficacy against SARS-CoV-2 in an *in silico* study (Vankadari 2020). Then, in an *in vitro* study, it was shown to inhibit SARS-CoV-2 by partially blocking its entry to the cells (Wang, Cao et al. 2020). In addition to this, in a clinical study conducted, it was

shown to positively contribute to oxygen saturation, hospitalization duration, chest CTs. As a result it was recommended for a study with a larger sample size (Nojomi, Yassin et al. 2020).

These 3 molecules were not the only suggested drug repurposing targets. Many other studies were published, and different approaches were ventured. Different datasets such as, ZINC15, SuperDRUG2, PubChem, ChEMBL, MTiOpenScreen, Thiol-reacting FDA-approved drugs, medicinal plants, were docked against different targets such as PL protease (PDB: 6W9C), Main protease (PDB: 6LU7), nsp15 protein (PDB: 6W01) and nsp-16 (PDB: 6WH4) (Arun, Sharanya et al. 2020, da Silva Hage-Melim, Federico et al. 2020, Elmezayen, Al-Obaidi et al. 2020, Gyebi, Ogunro et al. 2020, Lobo-Galo, Terrazas-López et al. 2020). Some of these studies did MD and ADMED simulations as well. Main protease (Mpro), RdRp, spike protein and TMPRSS2 were common docking targets. Saquinavir, Darunavir, lopinavir, Remdesivir, Indinavir, Galidesivir, Saikosaponin, Raltegravir and Simeprevir were the most frequently suggested candidates. Most of the studies did not have experimental validation data (Dotolo, Marabotti et al. 2021).

As can be seen from the examples above, some of the molecular docking results are actually promising and can be verified by *in vitro* or clinic studies. Therefore, we investigated the efficacy and the interactions of FDA-approved drugs against SARS-CoV-2 in this study.

2.6 Method

In this study, Autodock 4.2 and Autdock Vina were used as the main docking tools (Morris, Huey et al. 2009, Trott and Olson 2010). Autodock 4.2 is an established and effective docking tool used for protein-ligand docking. It uses a genetic algorithm called Lamarckian algorithm for confirmation search and a semi-empirical force field function for scoring. It is being maintained by Scripps Research and is succeeded by Autodock Vina. Autodock Vina is about two orders of

magnitude faster than Autodock 4.2 and also has better accuracy at binding mode predictions. Vina is multithreaded and can achieve parallelism on multi-core devices. It uses a stochastic search method called Iterated Local Search global optimizer, which resembles a genetic algorithm with a metropolis criterion. It also uses the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method for choosing the optimization direction. AutoDock Vina has a semi-empirical and semi-knowledge based scoring function which is considered more of a machine learning algorithm (Trott and Olson 2010).

Since Experimental DRI-C molecules are novel drugs and are only synthesized in the lab, they are not found in any of the databases. Therefore, the structures of the experimental DRI-C molecules were manually drawn using the Avogadro software for molecular docking (Hanwell, Curtis et al. 2012). Methylene Blue and Erythrosine B structures were taken from the PubChem (Kim, Thiessen et al. 2016). Three-dimensional crystal structures of SARS-CoV-2 spike receptor-binding domain that is bound to ACE2 (PDB ID: 6M0J) at a resolution of 2.45 Å was downloaded from the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PDB) (Lan, Ge et al. 2020). Another PDB file was also created from 6M0J to isolate the spike protein.

Ligand Molecules were downloaded from the ZINC database, which is a collection of compounds specifically designed for virtual screening. The downloaded molecules were in SDF format, which is a text format that defines structure data through atoms, bonds, and coordinates. Then SDF files were converted to PDB format, which is another text format that describes 3D molecules. This conversion was done by OpenBabel, a toolkit designed to be the translator between different data formats and disciplines (O'Boyle, Banck et al. 2011). PDB files were converted once again to PDBQT format, which is very similar to PDB, but carries extra info about partial charges and atom types. This conversion was done by MGTOOLS, which is a utility box that comes with Autodock

4.2. At this point pdbqt files of the target protein (6M0J) and ligand were given to Autodock Vina along with a configuration file which defines the area of interest. Once Vina is done, it prints out the binding affinities for the top 9 poses. These were parsed using a BASH script and appended to a CSV file. The best binding pose from Vina was rescored using Autodock 4.2. If the rescoring values were similar PDBQT output files from Vina, they were overlayed on top of the original structure using PyMol, which is a python-based 3D model visualization tool (Schrodinger 2015). The resulting structure was then fed into PLIP, which identifies the non-covalent bonds between ligands and structures (Salentin, Schreiber et al. 2015).

Chapter 3 - *in silico* Study of The Experimentally Proved SARS-Cov-2 Inhibitors and Use of Ligand Efficiency to Validate the Molecular Docking Method

3.1 Introduction

In this study, we would like to filter and find promising drug molecules against SARS-CoV-2 using virtual screening/molecular docking tools. Before moving on to finding promising drug molecules, we would like to verify our method. For this purpose, a number of experimental molecules from other studies were used to evaluate how reliable the molecular docking method is (Bojadzic, Alcazar et al. 2020, Bojadzic, Alcazar et al. 2020). These experimental molecules were shown to inhibit SARS-CoV-2 in wet-lab experiments. Therefore, they were good reference points for our study. FDA molecules with common features and interactions would be more likely to be successful at inhibiting the virus. Also, since real world data for these molecules were available, simulation results could be cross checked, and our method could be verified.

The experimental molecules that were used are novel drug molecules that were synthesized by the researchers in their laboratories. Some of these molecules were used to target the CD40–CD154 Costimulatory Protein-Protein Interaction and acted as inhibitors (Bojadzic, Chen et al. 2018). Therefore, researchers wanted to try their novel molecules against SARS-CoV-2. In their study, they have tested DRI-C21041, DRI-C23041, DRI-C24041, DRI-C61041, DRI-C71041, DRI-C91005, DRI-C2204745, Methylene Blue and Erythrosine B molecules against SARS-CoV-2, and they showed that all of these molecules inhibited the interaction of the spike proteins of SARS-CoV-2 with ACE2. Methylene Blue and Erythrosine B are actually dye molecules and are already FDA approved. MeBlu is used for the treatment of methemoglobinemia, and has several other therapeutic applications (Jack Clifton and Leikin 2003). EryB is a food colorant, and in another

study it was shown to be a potential protein–protein interaction inhibitor (Ganesan, Margolles-Clark et al. 2011). The chemical structures of all the experimental molecules are shown in Figure 5.

Figure 5. Molecular structures of (A) DRI-C21041, (B) DRI-C23041, (C) DRI-C24041, (D) DRI-C2204745, (E) DRI-C61041, (F) DRI-C71041, (G) DRI-C91005, (H) Methylene Blue, and (I) Erythrosine B.

3.2 Method

Since Experimental DRI-C molecules are novel drugs and are only synthesized in the lab, they are not found in any of the databases. Therefore, the structures of the experimental DRI-C molecules were manually drawn using the Avogadro software for molecular docking (Hanwell, Curtis et al. 2012). Methylene Blue and Erythrosine B structures were taken from the PubChem database (Kim, Thiessen et al. 2016). Three-dimensional crystal structure of SARS-CoV-2 spike receptor-binding domain that is bound to ACE2 (PDB ID: 6M0J) at a resolution of 2.45 Å was downloaded from the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PDB) (Lan, Ge et al. 2020). Another PDB file was also created from 6M0J to isolate the spike protein.

DRI-C molecules, Methylene Blue and Erythrosine B were docked against both 6M0J and the spike protein using Autodock Vina (Trott and Olson 2010). In order to account for the stochastic behavior of docking, this exercise was repeated 100 times. Then, the molecules were rescored using Autodock 4.2 to validate the method and results (Morris, Huey et al. 2009). The receptor binding region with a sizable margin was used as the grid box. The grid box mapping parameters for 6M0J along x, y and z directions respectively are as follows: (Å) x= 38 y= 78 z= 40 and Center (Å) x= -36.139 center_y= 33.528 center_z= 3.444. And for the spike only structure parameters are: (Å) x= 48 y= 60 z= 40 and Center (Å) x= -43.083 center_y= 38.667 center_z= 0.111. The best binding pose of the molecule was overlayed to the target structure using PyMol to create a combined new PDB file. Then PLIP was used to analyze and visualize the interactions of the combined structure (Salentin, Schreiber et al. 2015, Schrodinger 2015).

3.3 Results and Discussion

The first step in validating our method was making sure that the binding affinity results were sane and sensible. To verify this, structures were docked by Autodock Vina and rescored by Autodock 4.2. As can be seen in Figure 6, there is a clear trend between the Autodock Vina Binding Affinity results and the Autodock 4.2 rescoring results. Based on this, we can say that the docking protocol is validated.

Figure 6. Autodock Vina Binding Affinity results and the Autodock 4.2 rescoring results of 6M0J and spike protein structures.

After the method validation, the second step is to see whether there was a correlation between the experimental results and the molecular docking results. Finding a significant correlation would mean that when we apply our molecular docking method to the FDA approved drug molecules and pick some targets, they would be likely to inhibit SARS-CoV-2 in the lab experiments as the experimental molecules do.

Table 1 shows the binding affinity and rescoring results as well as the experimental IC₅₀ values. IC₅₀ is a metric that indicates how much of a particular inhibitor is required *in vitro* experiments to decrease a biochemical activity by half. (Aykul and Martinez-Hackert 2016).

Table 1. IC50 Values, and Autodock Vina results of experimental molecules.

	IC50 Values (nM)	Whole Structure Vina Results (kcal/mol)	Spike Structure Vina Results (kcal/mol)
DRI-C91005	0.16	-10.9	-8.9
EryB	0.43	-8.1	-6.4
DRI-C23041	0.52	-9.7	-8.4
DRI-C21041	1.82	-10.3	-8
MeBlu	3.03	-6.3	-5.6
DRI-C24041	3.35	-9.5	-7.6
DRI-C2204745	4.04	-9.4	-8.5
DRI-C61041	7.67	-10.4	-7.6
DRI-C71041	9.95	-10	-8

According to the IC50 values, DRI-C23041, DRI-C91005, and EryB can inhibit SARS-CoV-2 at low molarities. However, there is no clear relationship between the docking/rescoring results and experimental IC50 values of these molecules. Although DRI-C91005 ranks first in all categories, EryB with the second lowest IC50 value would be expected to have the second-best docking results. However, this is not the case, and since the IC50 values and binding affinity scores are completely different concepts that measure different things, this is completely expected (Mikra, Rossos et al. 2017). As can be seen in Figure 5, the molecules' structures and their sizes are significantly different, which affects the docking results and possible interactions.

Typically, larger compounds make more interactions with the target and therefore display higher levels of binding energy. However, higher levels of binding energy may be misleading and do not necessarily mean a more efficient binding. A smaller molecule with comparable binding energy would probably make less interactions but bind more strongly. Since drug candidates often get bulkier during the drug optimization stage, a smaller but more efficiently binding molecule is more desirable. Ligand efficiency metric can be used to find such candidates during screening. Ligand efficiency is calculated by dividing the binding energy of a compound by the number of its

nonhydrogen atoms (Hopkins, Groom et al. 2004). This would mean that smaller molecules would be picked over larger molecules exhibiting similar binding energies.

EryB is a significantly smaller molecule than DRI-C91005 and DRI-C23041. Therefore, it would be expected to bind to the target molecule with less binding energy than the other two. However, when the binding energy for all molecules were normalized with respect to their molecular weights, comparison of the results would be more meaningful. Also, since binding efficiency can be calculated both from IC50 and binding energy values, it can be used as a way of verifying the relationship between the measured IC50 values in the lab and the docking results from simulations.

First ligand efficiencies for the experimental molecules were calculated using the equation 1, where R is the gas constant ($1.987 \times 10^{-3} \text{ kcal} \cdot \text{K}^{-1} \cdot \text{mol}^{-1}$) and T is temperature (K). Then, the predicted ligand efficiencies were calculated using equation 2, where the Autodock Vina binding affinity results were used.

$$\frac{-RT \ln \text{IC50}}{\text{Number of nonhydrogen atoms}} \quad (1)$$

$$\frac{\text{Binding free energy}}{\text{Number of nonhydrogen atoms}} \quad (2)$$

The results were shown in Table 2.

Table 2. Ligand Efficiency Results

	Experimental (IC50) Ligand Efficiencies	6M0J Autodock Vina Results' Ligand Efficiencies	Only Spike Autodock Vina Results' Ligand Efficiencies
DRI-C91005	0.1997	0.2319	0.1893
EryB	0.2837	0.2612	0.2064
DRI-C23041	0.2117	0.2365	0.2048
DRI-C21041	0.1934	0.2512	0.1951
MeBlu	0.3630	0.3	0.2666
DRI-C24041	0.1844	0.2317	0.1853
DRI-C2204745	0.1774	0.2238	0.2023
DRI-C61041	0.1723	0.2536	0.1853
DRI-C71041	0.1685	0.2439	0.1951

According to the paper, the experimental molecules were only bound to the spike protein (Bojadzic, Alcazar et al. 2020). Figure 7 shows the comparison of actual and predicted ligand efficiency results for the spike protein structure. There is a clear trend between the two sets of results and the correlation is 0.90.

Figure 7. Ligand Efficiencies of Only Spike Structure and their correlation.

However, when we are screening the FDA approved drugs, we decided to consider drugs that would bind to the interface as well. Therefore, we repeated the procedure for the whole (PDB ID: 6M0J) structure as well. As Figure 8 shows, when the actual and predicted ligand efficiency results for the whole structure are compared, a similar trend can be seen and the correlation is 0.86, which is still quite high. This would mean that using the whole structure for docking would still give us meaningful results and that the results that we get from docking closely follow the results from the laboratory experiments.

Figure 8. Ligand Efficiencies of Whole Structure (6M0J) and their correlation.

As stated earlier, a high binding affinity may not always lead to an efficient binding or low IC50 value. To make better predictions, properties of the known to be inhibiting molecules can also be used. One of the most important properties of the experimental drugs is where they bind to. Molecules that make similar bonds with the experimental molecules would be more likely to

inhibit SARS-CoV-2. Also finding hotspots that experimental molecules bind to can give us clues about the inhibition potential of certain interactions.

With respect to the binding affinity results, DRI-C21041 is one of the top candidates (-10.3 kcal/mol). Its binding location to the 6M0J structure is shown in Figure 9(A), and its interactions are shown in Figure 9(B). As seen in Figure 9(A), DRI-C21041 molecule binds to the interface of ACE2 and spike protein of SARS-CoV-2. DRI-C21041 forms Hydrogen bonds with Asn33, Thr92, Gln96 residues on ACE2, and with Arg403 on the spike protein. It only had hydrophobic interactions with the ACE2 structure: Asn33, Glu37, Thr92, and Pro389.

Figure 9. (A) DRI-C21041 on 6M0J (B) Molecular Interactions between DRI-C21041 and 6M0J

DRI-C23041 molecule is slightly different from DRI-C21041. Their molecular sizes are the same; but the SO_3H group is bound to another carbon in DRI-C23041. As can be seen in Figure 10(A), they bind to similar places on the 6M0J structure, and they share common interactions, which are

visualized in Figure 10(B). DRI-C23041 makes hydrogen bonds with Glu3, Asn90 on the ACE2 and with Arg403 on the spike protein. It forms the exact same hydrophobic bonds that DRI-C21041 forms: Asn33, Glu37, Thr92, Pro389 DRI-C21041. DRI-C23041 also forms a salt bridge, a non-covalent interaction formed between two ionized molecules, with Lys26 on the ACE2 (Kurczab, Śliwa et al. 2018).

Figure 10. (A) DRI-C23041 on 6M0J (B) Molecular Interactions between DRI-C23041 and 6M0J

DRI-C24041 has the same molecular weight and atoms as DRI-C21041 and DRI-C23041. Just like DRI-C23041, it has SO_3H in a different location. Despite the similar structure it exhibits a lower affinity and a higher IC_{50} value. However, as Figure 11(A) shows, it makes about the same interactions as the two other variants. The only minor difference is the pi-stacking it makes. pi-stacking is non-covalent interactions of aromatic molecules (Martinez and Iverson 2012).

Figure 11. (A) DRI-C24041 on 6M0J (B) Molecular Interactions between DRI-C24041 and 6M0J

The next molecule is DRI-C61041. It has one of the higher binding affinities. It is similar to DRI-C21041; but it has a Nitrogen atom bound to one of its aromatic rings. It exhibits most of the same interactions as the previous molecules. It forms hydrogen bonds with Asn33, Gln96 on the ACE2 and with Arg403 on the spike protein. It forms hydrophobic bonds with Asn33, His34, Glu37, Pro389, Tyr505. Its interaction with Tyr505 is interesting. Tyr505 is mentioned in the literature as a hotspot (Veeramachaneni, Thunuguntla et al. 2020). Although due to its interactions and binding score, DRI-C61041 seems promising, it actually has a higher IC₅₀ value than the previous molecules. It is one of the few molecules that stops the correlation from being 1.0 and implies that the relationship between interactions and IC₅₀ values is a complex one.

Figure 12. (A) DRI-C61041 on 6M0J (B) Molecular Interactions between DRI-C61041 and 6M0J

DRI-C71041 is another molecule which has a high binding affinity. It has a similar structure as DRI-C61041. However, one of its aromatic rings is five membered and has a Sulphur atom as the fifth member. It exhibits similar hydrogen bonds as the other molecules. It makes hydrogen bonds with Asn33, Gln96 on ACE2, and with Arg403 on the spike protein. It also makes two pi bonds, one with His34 on ACE2 and one with Arg403 on the spike protein. Just like DRI-C61041, DRI-C71041 seems promising but actually has one of the highest IC₅₀ values.

Figure 13. (A) DRI-C71041 on 6M0J (B) Molecular Interactions between DRI-C71041 and 6M0J

Compared to the previous molecules, DRI-C2204745 is a larger molecule. Therefore, it has more interactions with the target structure. As Figure 14(A) shows, it binds to a different place than the previous molecules. This molecule makes hydrogen bonds with Asp30, Asn33, Met323, Gln325, Gln388 on ACE2, and with Gly504, Tyr505 on the spike protein. It also makes hydrophobic bonds with Asn33, Thr324, Ala387, Pro389 on ACE2, and with Val503, Tyr505 on the spike protein. It has some salt bridge interactions as well. It forms salt bridges with His34 and Arg403. Although it shares some interactions with the previous molecules, it also exhibits unique interactions with the RBD as well. Also, it has a decent binding affinity score, -9.4 kcal/mol.

Figure 14. (A) DRI- C2204745 on 6M0J (B) Molecular Interactions between DRI- C2204745 and 6M0J

DRI-C91005 is the top molecule in the group. It ranks highest both in IC₅₀ and binding affinity results. As Figure 15(A) shows, this molecule binds to a different place than the other molecules and does not have any common interactions. However, according to the literature, its binding location is still within the RBD. The reason for a different binding location might be DRI-C91005 being a larger molecule than the others. As molecules get larger their binding areas grow as well, and this leads to higher affinity results (Kuntz, Chen et al. 1999). DRI-C91005 makes hydrogen bonds with Asn49, Asn58, Gln76 on ACE2, and with Gly485, Asn487 on the spike protein. Gly485 is an important interaction. According to literature Gly485 is another hotspot on the spike protein (Shang, Ye et al. 2020). DRI-C91005 also makes hydrophobic bonds with Lys31, Leu39, Gln42, Ala46, Asn61, Lys68.

Figure 15. (A) DRI-C91005 on 6M0J (B) Molecular Interactions between DRI-C91005 and 6M0J

Unlike the DRI-C molecules, EryB is a dye molecule. As Figure 5 shows, it is a significantly smaller molecule than the others. This explains why it has lower binding affinity than the others (Kuntz, Chen et al. 1999). Although it has a lower binding affinity, according to the IC₅₀ values, it can inhibit SARS-CoV-2 with one of the lowest molarities within the group. Just like the others it binds to the RBD. It makes hydrogen bonds with Arg393 on ACE2, and with Arg408, Tyr505 on the spike protein. All three of these interactions are mentioned as hotspots in the literature (Ali and Vijayan 2020, Shang, Ye et al. 2020, Veeramachaneni, Thunuguntla et al. 2020). It also makes hydrophobic bonds with Thr324, Phe356, Met383, Ala386, Tyr505; all of which are in the RBD. Another interesting property of Ery B is it was suggested as a potential PPI in a study (Bojadzic, Chen et al. 2018). Due to all of these properties it makes a good reference point for our screening.

Figure 16. (A) EryB on 6M0J (B) Molecular Interactions between EryB and 6M0J

Our last molecule is MeBlu, an FDA approved dye. Although it is a dye molecule it is also used to treat methemoglobinemia, a condition where blood can't carry oxygen to the body (Jack Clifton and Leikin 2003). MeBlu was also tested in a clinical study against SARS-CoV-2. It was administered as a part of a cocktail and showed promising results. (Alamdari, Moghaddam et al. 2020). As Figure 17(A) shows it is the only molecule in the group that only binds to ACE2. It has a smaller structure than the others and does not make any hydrogen bonds. However, it makes hydrophobic bonds with Phe40, Leu73, Phe390, Leu391. Due to its small size and lack of hydrogen bonds, it has a lower binding affinity than the others (Kuntz, Chen et al. 1999). Since it was suggested as a potential SARS-CoV-2 drug in other studies as well, it is another important reference point for our screening (Ghahestani, Shahab et al. 2020, Scigliano and Scigliano 2021).

Figure 17. (A) MeBlu on 6M0J (B) Molecular Interactions between MeBlu and 6M0J

All of the experimental molecules' interactions are listed in Table 3 for better visualization. Interactions with ACE2 and S protein were annotated with (A) and (E) suffix, respectively.

Table 3. Molecular Structures and Interactions of Experimental Molecules.

Molecule	Binding Affinity (kcal/mol)	Hydrogen / pi Interactions / Salt Bridge	Hydrophobic Interactions
DRI-C21041	-10.3	Asn33(A), His34(A), Thr92(A), Gln96(A), Arg403(E)	Asn33(A), His34(A), Glu37(A), Thr92(A), Pro389(A)
DRI-C23041	-9.7	His34(A), Glu37(A), Asn90(A), Arg403(E)	Asn33(A), Glu37(A), Thr92(A), Pro389(A), Tyr495(E)
DRI-C24041	-9.5	Glu37(A), Asn90(A), Thr92(A), Arg403(E), pi: Arg403(E)	Asn33(A), Glu37(A), Thr92(A), Pro389(A), Tyr495(E)
DRI-C61041	-10.4	Asn33(A), Gln96(A), Arg403(E)	Asn33(A), His34(A), Glu37(A), Pro389(A), Tyr505(E)
DRI-C71041	-10	Asn33(A), Thr92(A), Gln96(A), Arg403(E), pi: His34(A), Arg403(E)	Lys26(A), Leu29(A), Glu37(A), Val93(A), Pro389(A), Tyr505(E)
DRI-C91005	-10.9	Asn49(A), Asn58(A), Gln76(A), Gly485(E), Asn487(E)	Lys31(A), Leu39(A), Gln42(A), Ala46(A), Asn61(A), Lys68(A)

DRI-C220474 5	-9.4	Asp30(A), Asn33(A), Met323(A), Gln325(A), Gln388(A), Gly504(E), Tyr505(E), Salt Bridge: His34(A), Arg403(E)	Asn33(A), Thr324(A), Ala387(A), Pro389(A), Val503(E), Tyr505(E)
MeBlu	-6.3	-	Phe40(A), Leu73(A), Phe390(A), Leu391(A)
Ery B	-8.1	Arg393(A), Arg408(E), Tyr505(E)	Thr324(A), Phe356(A), Met383(A), Ala386(A), Tyr505(E)

All of the experimental drugs inhibit SARS-CoV-2 in a laboratory setting, albeit with different IC50 values. As Table 3 shows, with the exception of MeBlu, all of them have good binding affinity scores, which were verified by two different scoring tools. The validity of the docking method was also tested by calculating the ligand efficiency both from simulation results and the real world IC50 data. Since the results were consistent, we have a high degree of confidence that simulation results for FDA approved drugs would also be fairly reliable.

3.4 Conclusion

Simulations are approximations to real life. Their results are not always accurate or even reliable. Since this study relies on simulations heavily, first our method had to be verified and its predictions had to be compared against the real-life experiment results. These seven novel experimental molecules and two FDA-approved dye molecules that were shown to inhibit SARS-CoV-2 in a laboratory setting were analyzed. These molecules were docked against the spike protein and ACE2 complex. The resulting poses were scored using two different functions and showed similar trends, which verified our scoring function. IC50 values of these molecules from the wet-lab experiments were available and binding affinity scores had to be compared against them. However, IC50 and binding affinity do not measure the same things and cannot be compared against each other directly. Therefore, a metric, which can be calculated from both values was used as the common ground. This metric is called ligand efficiency, and it measures binding energy per atom. When it was calculated from real-life IC50 values and binding affinity score, resulting datasets

showed similar trends and a correlation of 0.86. This showed that there was a strong relationship between the lab results and simulation results. Therefore, we have a high degree of confidence that simulation results for FDA-approved drugs would also be fairly reliable.

However, docking results could be supported by extra predictors. Although the experimental molecules have some unique interactions, they also have a decent amount of common interactions, which is significant. They also bind to some residues that are mentioned in the literature as hotspots. Any FDA approved drug having a decent binding affinity and binding to the residues that the experimental drugs bind would likely have similar properties as the experimental drugs and would be a potential target for repurposing.

Of the nine experimental molecules, 8 of them exhibited high binding affinities and made interactions with both ACE2 and spike protein. The remaining one, MeBlu, only interacted with ACE2 and had a lower binding affinity. This suggests that drugs that do not interact with the spike protein at all, could still be effective against SARS-CoV-2.

Amongst the experimental molecules, hydrophobic bonds with Pro389, Glu37, Asn33, Thr92 and Tyr505, as well as hydrogen bonds with Asn33, His34, Thr92, Gln96, Tyr505 and Arg403 were common. Of these residues Glu37, His34 and Tyr505 were mentioned as hotspots in the literature. However, the molecule with the best performance, DRI-C91005, did not have any common interactions with the rest of the molecules. On the other hand, it interacted with Asn487, which is suggested as a hotspot on the spike protein. Since all of these molecules inhibit SARS-CoV-2, any FDA approved drug that shows similarities will be considered as promising.

Chapter 4 - Identifying Drug Repurposing Candidates for SARS-CoV-2 from FDA-Approved Drugs Using Molecular Docking Tools and Interaction Analysis

4.1 Introduction

Every drug prescribed in the United States has to be tested for safety and efficacy first. The government body that performs these tests and regulates the drug market is called The United States Food and Drug Administration, FDA. FDA tests are stringent and time-consuming. Of the 5000 compounds that reach the preclinical study stage, only one gets approved. Of the 10 compounds that reach the clinical studies, only one makes it to the market. The whole process takes 10 to 15 years. Therefore, a *de novo* drug getting approval in a short amount of time is unlikely. This makes the repurposing of previously approved drugs the most viable strategy for quick results.

The high death toll and the time sensitive nature of the pandemic has led to a lot of *in silico* and *in vivo* repurposing studies. Some of these were covered in the second chapter. In these studies, different stages of the virus activity and different parts of the virus were targeted. While some studies concentrated on the main protease, others focused on the spike protein and RBD (Abo-Zeid, Ismail et al. 2020, Absalan, Doroud et al. 2020, De Vita, Chini et al. 2020).

The goal of this study is to find FDA-approved drug candidates that could bind to RBD and inhibit the virus activity. Molecular docking method, which was verified on the experimental drugs in chapter 3, will be used to analyze how FDA approved molecules interact with the spike protein. Results from experimental drugs and previous repurposing studies will be used as reference points to identify the best candidates.

4.2 Method

During the virtual screening 1750 FDA approved molecules from the ZINC database were used. Each molecule was docked against the 6M0J structure using Autodock Vina. In order to decrease the bias, docking boxes from other studies were expanded. However, if the box is chosen too large, then docking takes a very long time and the probability of missing the binding domain increases. On the other hand, if a very small box is chosen, then the molecules are forced to bind at a setting which might not be the optimal energy solution. Therefore, a midsize box that contains the RBD was used. It is shown in Figure 18.

Figure 18. Grid box region which is used for molecular docking.

Molecules were ranked according to their binding affinities and the top 50 molecules were picked for further analysis. These 50 drugs and their affinities can be seen in Table 4.

Table 4. Top 50 FDA approved drugs and their binding affinities.

Ligand	Binding Affinity (kcal/mol)	Ligand	Binding Affinity (kcal/mol)	Ligand	Binding Affinity (kcal/mol)
Voxilaprevir	-10.9	Altabax	-9.6	Orap	-9
Eltrombopag	-10.5	Erismodegib	-9.5	Itc	-9
Sirolimus	-10.4	Retapamulin	-9.5	Atacand	-9
Venetoclax	-10.4	Montelukast	-9.5	Everolimus	-9
Lumacaftor	-10.1	Dihydroergotamine	-9.5	Axitinib	-8.9
Avapro	-10.1	Imatinib	-9.5	Glimepiride	-8.8
Paritaprevir	-10.1	Glecaprevir	-9.5	Suvorexant	-8.8
Emend	-10	Abemaciclib	-9.4	Eribulin	-8.8
Rolapitant	-10	Irinotecan	-9.4	Stendra	-8.7
Cabazitaxel	-10	Ibrutinib	-9.4	Lomitapide	-8.7
Rifabutin	-9.9	Ledipasvir	-9.4	Stivarga	-8.7
Nilotinib	-9.8	Elbasvir	-9.3	Vorapaxar	-8.7
Simeprevir	-9.8	Midostaurin	-9.2	Flibanserin	-8.6
Saquinavir	-9.8	Conivaptan	-9.2	Ergotamine	-8.6
Avodart	-9.8	Indacaterol	-9.1		
Lurasidone	-9.8	Tpv	-9		
Naldemedine	-9.7	Amaryl	-9		
Olaparib	-9.6	Telmisartan	-9		

As mentioned in chapter 2, in order to make computations manageable, docking tools use approximations and stochastic methods. Therefore, their results should be verified. Autodock 4.2 was used to rescore the top 50 candidates. Figure 19 shows the comparison of Autodock Vina and Autodock 4.2 scores. Results follow similar trends and despite a couple of outliers, the root mean square distance between the two sets is 0.85.

To make sure that our remaining 50 molecules were not random, an extra step of verification was done. If there were any drugs that are similar to the experimental drugs, it would be likely that at least some of them would be in the remaining 50. However, measuring the similarities of 1750 molecules is not an easy task. To be able to do this programmatically these molecules need to be quantized and a measurement metric needs to be set.

The former can be done using a technique called molecular fingerprinting. There are two major fingerprinting strategies, structure based, and hash based. In the former, bits usually show the presence of certain substructures and their order. In the later, all linear substructures in the molecule are enumerated and passed through a hash function (Haider 2010).

Once the molecules are quantized into vectors of bits, the next step is comparing them. The most commonly used metric for this task is Tanimoto Similarity coefficient (Bajusz, Rácz et al. 2015). Tanimoto coefficient is calculated by dividing the number of common bits between the two vectors

to the number of bits in their union. A value of 0 would mean that there is no similarity, and 1 would mean that the two vectors are identical (Tanimoto 1958). Since different fingerprinting algorithms use different bit schemes, a universal similarity threshold is not possible to be set. However, when using the Daylight fingerprinting algorithm a Tanimoto Similarity coefficient larger than 0.85 means that two molecules are similar (Cereto-Massagué, Ojeda et al. 2015).

There are many different fingerprinting algorithms that use different approaches. In order to decrease the variance, three different algorithms, Daylight, Morgan, and Torsion, were used. Experiments were done on Python using the RDKit library (Landrum 2016). Each of the 9 experimental drugs were compared against each of the 1750 FDA-approved drugs using the three different fingerprinting algorithms and Tanimoto Similarity. For each experiment, 10 FDA approved drugs with the highest similarity scores were picked. As a result, 3 collections (one for each fingerprinting algorithm) of 10 results were gathered for each of the 9 experimental molecules; in other words, a total of 270 data points were collected. In order to remove the outliers, only the most commonly recurring molecules were considered. From these results variants of the same molecules were dropped and the top 10 most occurring molecules were picked. Table 5 shows the results.

Table 5. Top 10 most occurring similar molecules.

Molecule	Binding Affinity (kcal/mol)	Number of Occurrence in Similarity Experiments	Exist in TOP 50
Dihydroergotamine	-9.5	26	✓
Ergotamine	-8.3	22	✓
Saquinavir	-9.7	18	✓
Elbasvir	-9.3	12	✓
Bromocriptine	-8.0	12	✗
Velpatasvir	-8.1	10	✗
Paritaprevir	-10.1	9	✓

Naldemedine	-9.7	9	✓
Rifabutin	-9	9	✓
Ledipasvir	-9.1	9	✓

From the top 10 most occurring molecules, 8 of them were in the remaining 50, which means that our filter worked as expected. However, the Tanimoto similarity scores for the Daylight fingerprint experiments ranged between 0.7-0.73, which means that none of the FDA approved drugs passed the 0.85 threshold.

Once there were only 50 molecules left, the next step was studying the binding locations and interactions. However due to the stochastic nature of the docking tools relying on a single set of results would not be enough. Therefore, docking experiments were repeated 100 times to decrease the variance in the results. The BASH scripts which were used for our method were included in Appendix A. Results from the iterations were overlaid using PyMol, and binding affinity distributions were plotted using Matplotlib library of Python (Hunter 2007).

Some of the molecules showed significant variance across the iterations. Their binding affinities, interactions, and even binding locations changed significantly. Some did not even consistently bind to RBD. These molecules were filtered out. Figure 20 shows one of the eliminated molecules, Vorapaxar. It binds to multiple different places on the structure and exhibits binding affinities ranging from -8.4 to -9.7. Since it cannot consistently bind to the binding domain, it would be unlikely to inhibit the spike protein activity.

Figure 20. (A) 6M0J Structure docked with Vorapaxar (B) Binding affinity distribution of 100 iterations

The remaining molecules were evaluated with respect to their interactions. Each molecule's most frequently seen binding affinity was determined and 5 poses with that affinity were randomly picked. These samples were analyzed with the PLIP software, and their interactions were recorded. Molecules which show incoherent interactions were filtered out.

According to the literature Glu35, Tyr83, Asp38, Lys31, Glu37, His 34, Lys353, and Arg393 are hotspots on ACE2. On the other hand, Gln493, Gln 498, Asn487, Tyr505, and Lys417 are hotspots on the spike protein (Ali and Vijayan 2020, Veeramachaneni, Thunuguntla et al. 2020). Drug molecules that interact with these residues were favored over ones that don't. Another metric used was similarity to the experimental drugs. Some of the interactions were common amongst the experimental molecules. Asn33(5/9), His34(5/9), Arg403(5/9), Glu37(5/9), Tyr505(4/9),

Thr92(3/9), and Gln96(3/9) interactions were frequently seen between experimental molecules and the RBD. Drug molecules that made these interactions were preferred over the others.

4.3 Results and Discussion

After rounds of eliminations 10 molecules were left. They are shown in Table 6.

Table 6. Molecular Structures and Interactions of top 10 FDA approved drug molecules.

Molecule	Structure	Binding Affinity kcal/mol	Hydrogen / pi / Halogen Interactions	Hydrophobic Interactions
Voxilaprevir		-10.9	Asn33(A), His34(A), Asp38(A), Gln96(A), Lys353(A), Arg403(E), pi: His34(A), Arg403(E)	Lys26(A), Leu29(A), Asn33(A), Glu37(A), Phe390(A), Lys417(E), Tyr453(E)
Sirolimus		-10.4	Thr27(A), His34(A), Lys417(E)	Asn33(A), Pro389(A), Asp405(E)
Paritaprevir		-10.1	Ser19(A), Thr27(A), Lys417(E), Tyr421(E), Tyr473(E), pi: Lys26(A)	Glu23(A), Lys26(A), Leu29(A), Val93(A), Tyr473(E)
Cabazitaxel		-10	Asn33(A), Gln409(E), Lys417(E)	Lys26(A), Asn33(A), Val93(A), Lys417(E)
Simeprevir		-9.8	Asp30(A), Asn33(A), Arg393(A), Lys417(E), Tyr505(E), pi: His34(A)	Lys26(A), Lys417(E)
Saquinavir		-9.7	Lys26(A), Asp30(A), His34(A), Arg403(E), Glu406(E), Lys417(E), pi: His34(A), Salt Bridge: Asp30(A)	Lys26(A), His34(A), Glu37(A), Val93(A), Pro389(A), Tyr495(E)
Glecaprevir		-9.5	Arg393(A), Arg403(E), Arg408(E), Gln409(E), Lys417(E), Tyr505(E), Halogen: Asn33(A)	Leu29(A), Asp30(A), Ala387(A), Pro389(A), Glu406(E)
Irinotecan		-9.4	Thr27(A), Asn33(A), Arg393(A), Lys417(E), Tyr473(E)	Glu23(A), Lys26(A)
Indacaterol		-9.1	Ser47(A), Asp350(A), pi: Trp349(A)	Phe40(A), Asp350(A), Phe390(A), Arg393(A)

Ergotamine		-8.6	Asp30(A), Arg403(E), Tyr505(E)	Leu29(A), Asn33(A), Val93(A), Pro389(A)
-------------------	---	------	-----------------------------------	--

Voxilaprevir is an anti-viral drug used to treat Hepatitis C. It reversibly binds to and inhibits the nonstructural (NS) protein 3/4A protease (Heo and Deeks 2018). When used with other Hepatitis C drugs, daclatasvir and sofosbuvir, it has been shown to inhibit SARS-CoV-2 in multiple studies (Jockusch, Tao et al. 2020, Sacramento, Fintelman-Rodrigues et al. 2020). In our study, the Voxilaprevir molecule has the highest binding affinity (-10.9 kcal/mol) amongst the selected molecules. As the Figure 21(A) demonstrates, Voxilaprevir binds directly to the RBD region of the whole structure and shows very little variance in binding location and affinity during the iterations (Figure 21(B)). As can be seen in Figure 21(C), Voxilaprevir forms 6 Hydrogen, 2 pi, and 7 Hydrophobic interactions with the interface. It binds to Asn33, His34, and Arg403 which are common with DRI-C21041, DRI-C23041, DRI-C61041, DRI-C71041, DRI-C2204745.

Its Arg403 and Lys353 interactions are worth a mention. Arg403 interaction was seen in other studies where proposed molecules were claimed to inhibit SARS-CoV-2 (Gurung, Ali et al. 2021). In the same manner, Lys353 interaction was seen in a study where Dalbavancin molecule was claimed to be effective against SARS-CoV-2 in animal trials (Wang, Yang et al. 2021). In another study Lys353 was shown to be bonding with the spike protein, despite the inhibition of the ACE2 (Gangadevi, Badavath et al. 2021). Since Voxilaprevir makes both of the interactions, it is a promising candidate. Also, since it is a protease inhibitor, it can be effective against the SARS-CoV-2 replication.

Figure 21. (A) 6M0J Structure docked with Voxilaprevir (B) Binding affinity distribution of 100 iterations (C) Molecular interactions of Voxilaprevir.

The next drug in our list is a mTOR inhibitor, which is used as an immunosuppressant agent. mTOR inhibitors are used to reduce the sensitivity of T and B immune cells to interleukin-2 (IL-2), which inhibits these cell's activation (Nguyen, Vautier et al. 2019). There is a study that claims that mTor inhibitors can be used against SARS-CoV-2 (Husain and Byraredddy 2020). In our study, Sirolimus exhibits a low binding affinity (-10.4 kcal/mol) and stable results throughout the 100 iterations (Figure 22(B)). It forms 3 hydrogen bonds, one of which is with the His34 residue, a common interaction with the experimental molecules and also a hotspot.

It also makes a hydrogen bond with Lys417, which is a hotspot as well. Lys417 was studied in an *in vitro* study as well and was deemed as an important interaction (Gangadevi, Badavath et al. 2021). Although this drug does not make too many hydrogen bonds, distances for the ones it makes are below 3 Å, which suggests a strong binding.

Sirolimus also makes a hydrophobic bond with Asp405, which is not a common interaction among the experimental or FDA-approved drugs. However in some studies it was suggested to be a promising interaction (Mohammadi, Heidarizadeh et al. 2020, Tariq, Mateen et al. 2020). Sirolimus makes its other hydrophobic bonds with Asn33 and Pro389 of ACE2, which are both common hydrophobic bonds that the experimental molecules make.

Figure 22. (A) 6M0J Structure docked with Sirolimus (B) Binding affinity distribution of 100 iterations (C) Molecular interactions of Sirolimus.

Paritaprevir is an antiviral that is used to treat Hepatitis C and it targets the NS3-4A serine protease. It shows very little variance in the iterations and has a binding affinity of -10.1 kcal/mol. Paritaprevir makes 4 hydrogen bonds. One of its hydrogen bonds is with Lys417, which is a hotspot. It also makes a hydrogen bond with Tyr473, which is similar to Glecaprevir. Tyr473 is considered to have modestly important residue. Its other hydrogen bond interactions, Ser19 and Thr27 are not seen in the experimental molecules. However, they are both located in the beginning

of the helix region of ACE2, which is considered as one of the most important regions (Guo, Chen et al. 2020). Paritaprevir also makes a pi stacking interaction with Lys26 and 5 hydrophobic bonds. Although these residues are not seen in experimental molecules, they are seen in other Hepatitis C drugs in our list. It is also suggested as a RNA polymerase inhibitor for SARS-CoV-2, which makes it more promising (Gurung, Ali et al. 2021). And it is suggested as a repurposing target for SARS-CoV-2 in another study as well (Mevada, Dudhagara et al.).

Figure 23. (A) 6M0J Structure docked with Paritaprevir (B) Binding affinity distribution of 100 iterations (C) Molecular interactions of Paritaprevir.

One of the molecules with the highest binding affinity scores in the list is Cabazitaxel. It was fairly consistent during the iterations and is used to treat prostate cancer. It makes 3 hydrogen and 4 hydrophobic bonds. Its hydrogen and hydrophobic bonds with Asn33 are common with the DRI-C experimental molecules. Its other interaction is with Lys417, which is a hotspot. It makes a hydrogen and a hydrophobic bond with this residue. And its last hydrogen bond is with Gln409,

which was seen in other studies but is not the most promising interaction. Although Cabazitaxel is a strong candidate due to its high affinity and important interactions, it also has a possible downside. Cabazitaxel is a chemotherapeutic agent and it can cause Neutropenia, a low count of a certain white blood cell type (Assi, Ibrahim et al. 2020). Due to this risk Cabazitaxel needs to be carefully studied.

Figure 24. (A) 6M0J Structure docked with Cabazitaxel (B) Binding affinity distribution of 100 iterations (C) Molecular interactions of Cabazitaxel.

Simeprevir is another hepatitis C drug that inhibits the nonstructural (NS) protein 3/4A protease (Ouwerkerk-Mahadevan, Snoeys et al. 2016). Simeprevir has already been suggested to be effective against other viruses (Li, Yao et al. 2019). It also has been proposed to be effective against SARS-CoV-2 in another *in silico*, molecular dynamics study (Trezza, Iovinelli et al. 2020). In our study, Simeprevir has exhibited a high binding affinity (-9.8 kcal/mol) and results were homogeneous (Figure 25B). As can be seen in Figure 25C, Simeprevir forms 5 Hydrogen, 1 pi,

and 2 Hydrophobic interactions with the interface. It makes a hydrogen bond and hydrophobic bond with Lys417, which is a hotspot (Gangadevi, Badavath et al. 2021). It shows similarities to DRI-C23041, DRI-C61041, DRI-C71041, DRI-C2204745 as well: it makes hydrogen bonds with Asn33, Tyr505 and a pi bond with His34, another hotspot.

Simeprevir has also 2 common interactions with Ery B: Arg393 and Tyr505, both of which are hotspots. They are commonly seen *in silico* studies (Benítez-Cardoza and Vique-Sánchez 2020, Sethi, Sanam et al. 2020). In an *in vivo* study Arg393 was also suggested to be a key residue for inhibition (Wang, Yang et al. 2021). Like Voxilaprevir, Simeprevir is a protease inhibitor and can also be effective against virus replication as well.

Figure 25. (A) 6M0J Structure docked with Simeprevir (B) Binding affinity distribution of 100 iterations (C) Molecular interactions of Simeprevir.

Saquinavir is an antiretroviral drug used to treat HIV. It is a protease inhibitor that is used to reduce the HIV virus load in the body, which slows down the disease and treats the symptoms. It has been

suggested as a SARS-CoV-2 remedy in multiple studies (Bello, Martínez-Muñoz et al. 2020, Montenegro, Al-Abed et al. 2020). Saquinavir is one the drugs with the lowest binding affinities (-9.7 kcal/mol) and the most uniform interactions in our study. It forms 6 Hydrogen, 6 Hydrophobic, and 1 pi bonds. Saquinavir interacts with His34 on ACE2 and with Arg403 in S-protein, which are common in DRI-C molecules. It also makes hydrogen bonds with Asp30 and Lys417. Additionally, it makes a salt bond with Asp30 as well. Lys417 is a hotspot on the RBD and in the absence of an inhibitor it usually interacts with Asp30 on ACE2 (Ali and Vijayan 2020). Therefore, its bonds with Asp30 and Lys417 are important.

Saquinavir also makes a hydrogen bond with Glu406 on the spike protein. The distance for this bond is below 3 Å. In a study published in 2015, a compound found in soybeans, Nicotianamine, was shown to inhibit ACE2. This compound was tested against SARS-CoV-2 *in silico* and was found to bind to Glu406 (Takahashi, Yoshiya et al. 2015, Chen and Du 2020). Saquinavir's other interaction with the spike protein is the hydrogen bond it makes with Gln409. This interaction was not seen in experimental molecules. However, according to the literature, it is one of the strong bonds that Naltrexone makes in molecular dynamics studies. According to *in vitro* studies, Naltrexone is a promising molecule for SARS-CoV-2 inhibition (Choubey, Dehury et al. 2020). Additionally, Saquinavir makes a hydrophobic bond with Tyr495. Amongst the experimental molecules, this interaction was only seen in DRI-C23041. However, it was mentioned as a promising interaction in literature (Sethi, Sanam et al. 2020, Farouk, Baig et al. 2021).

Saquinavir is one of the molecules with the most number of interactions in our list. It interacts with multiple hotspots and residues covered in the literature. It is also a protease inhibitor as well. So, like other drugs in our list, it has potential to disrupt SARS-CoV-2 at different stages.

Figure 26. (A) 6M0J Structure docked with Saquinavir (B) Binding affinity distribution of 100 iterations (C) Molecular interactions of Saquinavir.

Glecaprevir is the third hepatitis C antiviral drug in our list. Like Voxilaprevir and Simeprevir, it inhibits the nonstructural (NS) protein 3/4A protease, which prevents the cleavage of the HCV polyprotein (Ng, Tripathi et al. 2018). In our study, Glecaprevir has exhibited a high binding affinity (-9.5 kcal/mol), and consistently made the same bonds throughout the iteration (Figure 27B).

Just like Saquinavir, it has a high number of interactions. As can be seen in Figure 27C, Glecaprevir forms 6 Hydrogen, 1 Halogen, and 5 Hydrophobic interactions within the interface. Its hydrogen bonds with Arg403 and Tyr505 are common with the DRI-C molecules. It also makes a halogen bond with Asn33, which is a residue, experimental molecules frequently interact with. Its interactions with Arg393, Arg408, and Tyr505 are exactly the same as Ery B. Furthermore, Glecaprevir exhibits the Lys417 interaction which is common with Simeprevir and is a hotspot.

More importantly, it was suggested as an inhibitor for the SARS-CoV-2 main protease as well (Shamsi, Mohammad et al. 2020).

With the exception of Gln409, all of Glecaprevir's interactions are either with hotspots or frequently seen amongst the experimental molecules or covered in the literature. Actually Gln409 was also seen in some *in silico* studies (Tariq, Mateen et al. 2020, Jena, Kanungo et al. 2021). However, in our study, its bond distance was 3.96 Å. Although distances smaller than 4 Å are considered as hydrogen bonds in literature, this bond is unlikely to be as strong as the others. Despite this both bonds with ACE2 and bonds with the spike protein are close to each other. This suggests a strong inhibition potential.

Figure 27. (A) 6M0J Structure docked with Glecaprevir (B) Binding affinity distribution of 100 iterations (C) Molecular interactions of Glecaprevir.

Another molecule that has a low binding affinity (-9.4 kcal/mol) and exhibits uniform interactions throughout the 100 iterations is Irinotecan (Figure 28(B)). This drug inhibits topoisomerase, an

enzyme which controls the structure of DNA and plays a role in replication. It is used to treat various kinds of cancer. Irinotecan forms 5 Hydrogen bonds: with Thr27, Asn33, and Arg393 from ACE2, and with Lys417 and Tyr473 from the RBD. The Asn33 residue is common with most of the DRI-C molecules and Arg393 is common with EryB. Also, Lys417 is an important hotspot. Its other hydrogen bond with the spike protein is noteworthy as well. Tyr473 is mentioned in multiple studies as forming modestly strong bonds with Asp30 of ACE2 (Alexander, Schoeder et al. 2020, Ali and Vijayan 2020). Although it is not a hotspot, interacting with it would increase the chances of inhibition. Besides its interactions and binding affinity, there is another factor which makes Irinotecan a candidate. In one *in vivo* study, Topoisomerase 1 inhibition therapy was shown to be effective against inflammation caused by SARS-CoV-2 (Ho, Mok et al. 2020).

Figure 28. (A) 6M0J Structure docked with Irinotecan (B) Binding affinity distribution of 100 iterations (C) Molecular interactions of Irinotecan.

Our next molecule is Indacaterol, which is used to treat bronchitis (Cazzola, Calzetta et al. 2014). Unlike our other molecules, it only binds to ACE2. Although this looks like an issue at the first glance, there are studies suggesting that it is possible to inhibit SARS-CoV-2 by only binding to ACE2. Actually, one of the experimental molecules does exactly that. MeBlu only binds to ACE2 and was shown to inhibit SARS-CoV-2 *in vitro* experiments. Just like MeBlu, Indacaterol is a small molecule, and it exhibits a decent binding affinity (-9.1 kcal/mol). However, its interactions are very different from the other molecules. It makes a double hydrogen bond with Ser47, a hydrogen bond with Asp350, a triple hydrophobic bond with Phe40, a double hydrophobic bond with Phe390, and hydrophobic bonds with Asp350 and arg393. It also makes a pi stacking interaction with Trp349. All these interactions happen inside of the ACE2 binding domain. These exact interactions were seen in another *in silico* study, where Catechin, a plant metabolite, was tested against SARS-CoV-2 (Jena, Kanungo et al. 2021). Since Catechin was shown to be effective SARS-CoV-2 *in vivo* experiments, Indacaterol is promising as well.

Figure 29. (A) 6M0J Structure docked with Indacaterol (B) Binding affinity distribution of 100 iterations (C) Molecular interactions of Indacaterol.

The last drug in our list with a lower end binding affinity score is Ergotamine. It is used to treat migraines and has a similar structure to EryB. It is a smaller molecule, and it only makes hydrogen bonds with the RBD of spike protein. Ergotamine makes hydrogen bonds with Arg403 and Tyr505, which are both common interactions amongst the experimental molecules. The latter, Tyr505, is also a hotspot. Ergotamine also makes a hydrogen bond with Asp30. In normal conditions Asp30 interacts with Lys417 of the spike protein, another hotspot. So Asp30 interaction has inhibiting potential. Ergotamine makes 4 hydrophobic bonds, two of which, Asn33 and Pro389, are common in experimental molecules. Apart from its promising interactions, there is *in silico* research claiming that it can be used as a main protease enzyme inhibitor against SARS-CoV-2 as well (Gurung, Ali et al. 2020).

Figure 30. (A) 6M0J Structure docked with Ergotamine (B) Binding affinity distribution of 100 iterations (C) Molecular interactions of Ergotamine.

4.4 Conclusion

Due to the challenging nature of drug design and stringent FDA approval process, finding a new treatment for SARS-CoV-2 in a short period of time is unlikely. Therefore, in this chapter, drugs that were already approved by the FDA were investigated and molecules that have potential for inhibiting the spike protein were identified.

First, molecules that exhibited the highest binding affinities towards the spike protein and ACE2 complex were identified. A midsize box containing the RBD was selected, and 1750 FDA-approved drugs were docked with Autodock Vina. The resulting poses were rescored with another docking tool called Autodock 4.2. Scores from both tools were very similar, which verified our scoring function. Then, 50 molecules with the highest binding affinity scores were picked, and the rest of the molecules were discarded.

The more similar a molecule is to the experimental drugs from Chapter 3, the more likely it is to inhibit SARS-CoV-2. To verify that our filter created sane results, FDA approved drugs that were most similar to the experiment drugs were identified. This was done using two computational methods, molecular fingerprinting and Tanimoto Similarity. Of the top 10 FDA approved drugs that were most similar to the experimental molecules, 8 of them were in the final 50 list. This verifies that the initial filter does not act randomly.

Once the list is down to a manageable size, a more detailed analysis was done on the remaining drug candidates. In order to remove the variance that is caused due to the stochastic nature of the docking tools, each candidate was docked 100 times. Molecules that target specific places on the spike protein and ACE2 complex were preferred over the ones that were not consistently binding to the same place. Molecules which showed high variance in binding location and binding affinity were eliminated from the list. Interactions of the remaining molecules were analyzed, and molecules that make bonds with hotspots or with residues that were commonly seen in the analysis of experimental drugs were picked.

Of the top 10 candidates, 5 of them, Paritaprevir, Glecaprevir, Saquinavir, Simeprevir and Voxilaprevir, are antivirals. One is an immunosuppressant; one is used to treat bronchitis; one is used for migraine and two are used for cancer treatment. Most of these drugs are targeted for relevant diseases. They are either for diseases that affect areas of the body where ACE2 are expressed, or they are meant to treat virus infections. This is encouraging for *in vitro* studies. Another encouraging finding is, one of the molecules from the similarity list (Table 2), Elbasvir, has been shown to inhibit SARS-CoV-2 *in vitro* experiments (Milani, Donalisio et al. 2021). Elbasvir did not show consistent behavior during the iterations and was not considered for the final list (Figure 31). However, three other molecules from that list, Ergotamine, Saquinavir and

Paritaprevir showed consistent binding and affinity results and made it to the final list, which we believe is very promising.

Figure 31. (A) 6M0J Structure docked with Elbasvir (B) Binding affinity distribution of 100 iterations

Binding affinities of the final 10 drugs are similar with the binding affinities of the experimental drugs, which again is very good. The most promising candidate Voxilaprevir exhibited a binding affinity of -10.9, which was on par with the top experimental drug, DRI-C91005. Ergotamine had the lowest affinity of the group with -8.1, which was the same as EryB from the experimental drugs. The average affinity for the experimental drugs (with the exception of MeBlu) was -9.78 kcal/mol, while the average for the final 10 FDA approved candidates were -9.7 kcal/mol.

The molecules in the final list interacted most commonly with His34, Glu37, Lys417, and Tyr505 hotspot residues. Some of the interactions that were commonly made by the experimental molecules, Asn33, Arg403 and Pro389, were also frequently exhibited by the final 10 as well.

Voxilaprevir interacted with 5 hotspots and 2 residues that were common in the experimental drug interactions. Sirolimus interacted with 2 hotspots and 2 common residues. Cabazitaxel interacted with one hotspot and one common residue. Simeprevir interacted with 4 hotspots and one common residue. Saquinavir interacted with 4 hotspots and 2 common residues. Glecaprevir interacted with 3 hotspots and 3 common residues. Irinotecan interacted with 2 hotspots and 1 common residue. Ergotamine interacted with 1 hotspot and 3 common residues. Paritaprevir interacted with only 1 hotspot. Indacaterol only interacted with 1 hotspot but had 2 common interactions with MeBlu.

Overall, the high number of interactions with both hotspots and residues that were commonly seen during the analysis of experimental drugs, consistent binding locations, low binding affinities, other studies that suggest some of the top 10 molecules as treatment candidates, and the computationally calculated chemical similarity of some of the top 10 molecules to the experimental drugs suggest that these FDA approved drugs have high potential against SARS-CoV-2.

Chapter 5 - A Comparative Analysis of the Selected Drug Repurposing Candidates Against SARS-CoV-2 Variants Using Molecular Docking Tools

5.1 Introduction

When the viral genome is being replicated, sometimes there are copying errors, and a slightly different version of the virus is produced. When these changes are favorable for the virus that particular error, or mutation, makes it to the new generations and the virus genome changes. This new genome is called a strain. When the mutations change the phenotype of the virus, new strains are called new variants. The original, standard strain is called the wild type.

Since SARS-CoV-2 is a RNA virus, it has a higher mutation rate than DNA viruses (Lauring and Andino 2010). When the residues on the surface protein change significantly, the way the virus functions and avoids the immune system changes too. For example, the A82V mutation of the Ebola virus has increased its infectiousness and mortality (Diehl, Lin et al. 2016). This makes SARS-CoV-2 even more of a serious threat. As time passes and SARS-CoV-2 infects more and more people, new variants are starting to show up too. In this study we will only cover three major ones, B.1.1.7, B.1.351 and P.1.

The B.1.1.7 variant has spread in the UK first (Volz, Mishra et al. 2021). Eight of the variant mutations are in the spike protein. The most noteworthy mutations are N501Y in the RBD, and P681H in the cleavage site. These mutations alter the way the virus binds to ACE2 and replicate. Especially the N501Y is thought to be increasing the affinity of the spike protein to the ACE2 and making the virus to be more infectious (Liu, Liu et al. 2021). Another variant with the N501Y mutation is the B.1.351 variant which was first seen in South Africa. B.1.351 has two other key mutations in the spike protein, E484K and K417N (Wibmer, Ayres et al. 2021). The third variant,

P.1, originated in Brazil and has the N501Y mutation as well. It also has the E484K, K417T mutations in the RBD (Faria, Claro et al. 2021). The RBD mutations that these variants have may affect how effectively the virus binds to ACE2 and its replication in humans. They may also affect how effective the experimental drugs can inhibit SARS-CoV-2.

There are currently no approved treatments for the wild type, and this study aims to find drugs that could be repurposed to treat SARS-CoV-2. However, the mutations in the variants could change the way SARS-CoV-2 binds to ACE2 and could render possible candidates useless. Therefore, we have used computational methods to analyze how the mutations might affect the promising FDA-approved drugs from chapter 4.

5.2 Results and Discussion

Variants are created by modifying 6M0J using the PyMol mutagenesis tool. The mutations are shown in Table 7. The top 10 candidates from the chapter 4 were docked against the variants 100 times each and most common conformations were picked for interaction analysis. Molecular Interactions of the top 10 candidates for wild type and the variants are shown in Table 8.

Table 7. Amino acid mutations of the variants.

Lineage	Mutations
B.1.1.7	<ul style="list-style-type: none"> • Asn501 → Tyr (N501Y)
B.1.351	<ul style="list-style-type: none"> • Lys417 → Thr (K417T) • Glu484 → Lys (E484K) • Asn501 → Tyr (N501Y)
P.1	<ul style="list-style-type: none"> • Lys417 → Asn (K417N) • Glu484 → Lys (E484K) • Asn501 → Tyr (N501Y)

Table 8. Molecular Interactions of the top 10 candidates for wild type and the variants.

Molecule	Type	Binding Affinity (kcal/mol)	Hydrogen / pi stacking / Halogen / Salt Bridge Interactions	Hydrophobic Interactions
Voxilaprevir	Wild	-10.9	Asn33(A), His34(A), Asp38(A), Gln96(A), Lys353(A), Arg403(E), pi: His34(A), Arg403(E)	Lys26(A), Leu29(A), Asn33(A), Glu37(A), Phe390(A), Lys417(E), Tyr453(E)
	B.1.1.7	-9.8	Asn33(A), Asp38(A), Gln96(A), Lys353(A), Arg403(E), pi: His34(A), Arg403(E)	Lys26(A), Asn33(A), Glu37(A), Phe390(A), Lys417(E), Tyr453(E)
	B.1.35 1	-8.4	Thr92(A), Gln96(A), Arg403(E), Asn417(E), Tyr421(E), Tyr505(E) Halogen: Gln409(E)	Lys26(A), Asp405(E)
	P.1	-9.6	Asp30(A), Asn33(A), Asp38(A), Lys353(A), Arg403(E), pi: His34(A), Arg403(E)	Leu29(A), Asn33(A), Glu37(A), Val93(A), Phe390(A), Thr417(E), Tyr453(E)
Sirolimus	Wild	-10.4	Thr27(A), His34(A), Lys417(E)	Asn33(A), Pro389(A), Asp405(E)
	B.1.1.7	-9.6	Glu23(A), Lys417(E)	Lys26(A), Asn33(A), Pro389(A), Asp405(E)
	B.1.35 1	-9.3	Glu23(A), Tyr421(E)	Thr92(A), Leu392(A)
	P.1	-9.3	Asp30(A), Tyr421(E)	Thr92(A), Leu392(A)
Paritaprevir	Wild	-10.1	Ser19(A), Thr27(A), Lys417(E), Tyr421(E), Tyr473(E), pi: Lys26(A)	Glu23(A), Lys26(A), Leu29(A), Val93(A), Tyr473(E)
	B.1.1.7	-9	Asp30(A), Asn33(A), Lys417(E), Tyr421(E), Asn460(E)	Glu23(A), Lys26(A), Thr27(A), Phe456(E)
	B.1.35 1	-9.1	Lys31(A), Glu35(A), Lys484(E), Ser494(E)	Ala71(A), Glu75(A), Leu79(A), Tyr449(E), Tyr489(E), Phe490(E)
	P.1	-9.1	Lys31(A), Glu35(A), Lys484(E), Ser494(E)	Ala71(A), Glu75(A), Leu79(A), Tyr449(E), Tyr489(E), Phe490(E)
Cabazitaxel	Wild	-10	Asn33(A), Gln409(E), Lys417(E)	Lys26(A), Asn33(A), Val93(A), Lys417(E)
	B.1.1.7	-9.7	Asn33(A), Gln96(A), Gln409(E), Lys417(E), Salt Bridge: Arg403(E)	Lys26(A), Asn33(A), Val93(A), Lys417(E)
	B.1.35 1	-9.5	Thr92(A), Gln96(A), Arg403(E), Asn417(E), Salt Bridge: His34(A)	Leu29(A), Asn33(A), His34(A), Glu37(A), Pro389(A), Phe390(A), Tyr505(E)
	P.1	-9.4	Asn33(A), Gln96(A), Thr417(E)	Lys26(A), Asn33(A), Val93(A), Thr417(E)

Simeprevir	Wild	-9.8	Asp30(A), Asn33(A), Arg393(A), Lys417(E), Tyr505(E), pi: His34(A)	Lys26(A), Lys417(E)
	B.1.1.7	-8.6	Lys31(A), Glu35(A), Gln76(A), Glu484(E), Ser494(E)	Glu35(A), Leu39(A), Phe72(A), Leu79(A), Leu452(E), Tyr489(E), Phe490(E)
	B.1.35 1	-8.6	Gln76(A), Lys484(E), Ser494(E), pi: Phe72(A)	Lys31(A), Leu39(A), Glu75(A), Tyr449(E), Lys484(E), Tyr489(E), Phe490(E)
	P.1	-8.6	Gln76(A), Lys484(E), Ser494(E), pi: Phe72(A)	Lys31(A), Glu35(A), Leu39(A), Glu75(A), Tyr449(E), Lys484(E), Tyr489(E), Phe490(E)
Saquinavir	Wild	-9.7	Lys26(A), Asp30(A), His34(A), Arg403(E), Glu406(E), Lys417(E), pi: His34(A), Salt Bridge: Asp30(A)	Lys26(A), His34(A), Glu37(A), Val93(A), Pro389(A), Tyr495(E)
	B.1.1.7	-9.6	Asp30(A), His34(A), Arg403(E), Glu406(E), Gln409(E), Lys417(E), Tyr453(E), pi: His34(A)	Lys26(A), Asp30(A), His34(A), Glu37(A), Val93(A), Pro389(A), Tyr453(E), Tyr495(E)
	B.1.35 1	-9.4	Asp30(A), His34(A), Arg403(E), Glu406(E), Asn417(E), Tyr453(E), pi: His34(A)	Lys26(A), Asp30(A), His34(A), Glu37(A), Val93(A), Pro389(A), Tyr453(E), Tyr495(E)
	P.1	-9.5	Asp30(A), His34(A), Arg403(E), Glu406(E), Gln409(E), Thr417(E), Tyr453(E)	Lys26(A), His34(A), Glu37(A), Val93(A), Pro389(A), Tyr453(E), Tyr495(E)
Glecaprevir	Wild	-9.5	Arg393(A), Arg403(E), Arg408(E), Gln409(E), Lys417(E), Tyr505(E), Halogen: Asn33(A)	Leu29(A), Asp30(A), Ala387(A), Pro389(A), Glu406(E)
	B.1.1.7	-8.4	Arg393(A), Gln409(E), Lys417(E), Tyr505(E), Halogen: Asn33(A), Tyr505(E)	Leu29(A), Asp30(A), Asn33(A), Pro389(A), Glu406(E), Tyr505(E)
	B.1.35 1	-8.7	Glu35(A)	Leu39(A), Lys68(A), Phe72(A), Tyr449(E), Leu452(E)
	P.1	-8.2	Arg393(A), Gln409(E), Thr417(E), Tyr505(E), Halogen: Asn33(A), Tyr505(E)	Leu29(A), Asn33(A), Glu406(E), Tyr505(E)
Irinotecan	Wild	-9.4	Thr27(A), Asn33(A), Arg393(A), Lys417(E), Tyr473(E)	Glu23(A), Lys26(A)
	B.1.1.7	-9.1	Arg393(A), Gly504(E), Tyr505(E), pi: Tyr505(E), Arg403(E), Salt Bridge: His34(A), Arg403(E)	Asn33(A), Pro321(A), Met383(A), Ala386(A), Ala387(A), Phe555(A)

	B.1.35 1	-9.1	Arg393(A), Gly504(E), Tyr505(E), pi: Tyr505(E), Arg403(E), Salt Bridge: His34(A), Arg403(E)	Asn33(A), Pro321(A), Met383(A), Ala386(A), Ala387(A), Phe555(A)
	P.1	-9.1	Arg393(A), Gly504(E), Tyr505(E), pi: Tyr505(E), Arg403(E), Salt Bridge: His34(A), Arg403(E)	Asn33(A), Pro321(A), Met383(A), Ala386(A), Ala387(A), Phe555(A)
Indacaterol	Wild	-9.1	Ser47(A), Asp350(A), pi: Trp349(A)	Phe40(A), Asp350(A), Phe390(A), Arg393(A)
	B.1.1.7	-9	Ser47(A), Asp350(A), pi: Trp349(A)	Phe40(A), Asp350(A), Phe390(A), Arg393(A)
	B.1.35 1	-9	Asp350(A), Phe390(A), Arg393(A)	Trp69(A), Leu73(A), Leu100(A), Phe390(A), Leu391(A), Arg393(A)
	P.1	-9	Asp350(A), Arg393(A), Asn394(A)	Phe40(A), Trp69(A), Leu73(A), Leu100(A), Phe390(A), Leu391(A), Arg393(A)
Ergotamine	Wild	-8.1	Asp30(A), Arg403(E), Tyr505(E)	Leu29(A), Asn33(A), Val93(A), Pro389(A)
	B.1.1.7	-7.7	Arg403(E), Tyr505(E)	Leu29(A), Val93(A), Pro389(A)
	B.1.35 1	-7.8	Arg403(E), Tyr505(E)	Leu29(A), Val93(A), Pro389(A)
	P.1	-7.7	Lys26(A), Arg403(E), Tyr505(E)	Leu29(A), Asn33(A), Val93(A), Pro389(A)

Voxilaprevir is an anti-viral drug used to treat Hepatitis C. It has the lowest binding affinity to the wild type (-10.9 kcal/mol). As Figure 32(A) shows, when interacting with the wild type Voxilaprevir always binds to the same place with the same binding affinity. As Figure 32(B), 32(C), 32(D) show, this changes slightly for the variants. Most of the time Voxilaprevir binds to the RBD, but in a small number of iterations it binds only to ACE2. While the distance of bonds with the wild type range between 3-4 Å, they are slightly larger with the variants at 3.5-4 Å. The interactions that Voxilaprevir makes with B.1.1.7 and P1 are similar, while its interactions with B.1.351 are different from both the wild type and the two other variants. Voxilaprevir makes hydrogen bonds with His34, Asp38, Lys353 of the wild type, which are hotspots, as well as with Asn33 and Arg403, which are common bonds that the experimental molecules make. When

docked against B.1.1.7 and P.1, it no longer makes a hydrogen bond with the hotspot His34 and its binding affinity drops to -9.6 kcal/mol range, which is still very good. It keeps the hydrogen bonds and its pi bonds with His34 and Arg403.

However, Voxilaprevir does not bind with B.1.351 as well as it does with the other variants. Its binding affinity drops to -8.4 kcal/mol. It makes hydrogen bonds with Arg403 and Tyr505, which are important residues; but it does not interact with His34 or Lys353. It interacts with Asn417 instead, the mutation that is unique to B.1.351. Lys417 is a hotspot on the wild type, but this mutation decreases the importance of this residue, which might explain the affinity loss (Pavlova, Zhang et al. 2021). It might also explain the interaction with Tyr421. Some *in silico* studies also suggested that Tyr421 was an important residue (Allam, Ghrifi et al. 2020, Idress and Kumar 2021). However, there is not *in vitro* or *in vivo* data to support this claim.

Although Voxilaprevir binds to the RBD with more variance, results suggest that it could still be effective against the B.1.1.7 and P.1. However, it might not be as effective against B.1.351.

Figure 32. Binding affinity distributions and overlaid docking poses through 100 iterations of Voxilaprevir for (A) Wild Type, (B) B.1.1.7 Variant (C) B.1.351 Variant, and (D) P.1 Variant.

Sirolimus is an immunosuppressant agent which consistently binds to the same area of the wild type with a low affinity (-10.4 kcal/mol). Although its binding affinity increases against the variants, it keeps making stable bindings. Sirolimus makes a hydrogen bond with His34 and Lys417 of the wild type, both of which are hotspots. However, the His34 interaction is not seen with any of the variants, and Lys417 interaction is only seen with the B.1.17 variant. The binding affinities and bond distances increase against the variants as well. While bond distances are below 3 Å for the wild type, this number goes up to 4 Å range for the variants and binding affinities increase about 1 kcal/mol. When Sirolimus is docked against P.1 and B.1.351, it makes a hydrogen bond with Tyr421. Bonds with this residue are seen when other molecules are docked against these two variants. Since Sirolimus achieves a decent binding affinity with only two hydrogen bonds, it is possible that the K417N mutation causes this residue to become more

important. However, since there is not enough evidence to support this claim, the effectiveness of Sirolimus against the variants with the K417N mutation is optimistically doubtful.

Figure 33. Binding affinity distributions and overlaid docking poses through 100 iterations of Sirolimus for (A) Wild Type, (B) B.1.1.7 Variant (C) B.1.351 Variant, and (D) P.1 Variant.

Paritaprevir is an antiviral that is used to treat Hepatitis C. Like the other molecules in our list, it binds very consistently to the wild type and has a very low binding affinity. When docked against the variants, Paritaprevir still binds consistently to the same place. However, its binding affinity increases 1 kcal/mol. The average bond distance with the wild type is around 3 Å. However, this number goes up to 3.5 Å with B.1.1.7 and to 4 Å with the other variants. Paritaprevir makes a hydrogen bond with Lys417 of the wild type, which is a hotspot. This interaction exists with B.1.1.7, but not with B.1.351 and P.1. Instead, when docked against these two variants, it interacts with Lys484 and Ser494. This is a pattern that we have seen in other drugs as well. When the Lys417 is mutated to Asn417, the hotspot disappears, and these two residues take over the bonds.

As covered before Ser494 might have some potential. It is the opposite residue to a hotspot on ACE2, Arg393. Therefore, Paritaprevir might still be effective against P.1 and B.1.351. However, it is likely to be effective against B.1.1.7.

Figure 34. Binding affinity distributions and overlaid docking poses through 100 iterations of Paritaprevir for (A) Wild Type, (B) B.1.1.7 Variant (C) B.1.351 Variant, and (D) P.1 Variant.

Cabazitaxel is a chemotherapeutic agent (Assi, Ibrahim et al. 2020). It binds to the RBD of the wild type fairly consistently and it has one of the best binding affinities amongst the group (-10 kcal/mol). When docked against the variants, its binding affinity changes slightly, which is positive. However, while it is somewhat stable against B.1.1.7, its binding location shows more variance against the P.1 and B.1.351. Like some other drugs in the list, Cabazitaxel makes a hydrogen bond with the Lys417 hotspot of the wild type. This interaction is seen across the variants, even after the Lys417 mutations. This is a positive sign for the effectiveness of the

molecule against the variants. Also, Cabazitaxel makes almost identical hydrophobic bonds with all the variants and some of these bonds are with hotspots. This is also another positive sign.

Figure 35. Binding affinity distributions and overlaid docking poses through 100 iterations of Cabazitaxel for (A) Wild Type, (B) B.1.1.7 Variant (C) B.1.351 Variant, and (D) P.1 Variant.

Simeprevir is the other hepatitis C drug that has been suggested as a SARS-CoV-2 treatment (Li, Yao et al. 2019, Trezza, Iovinelli et al. 2020). It has one of the lowest binding affinities against the wild type in this study (-9.8 kcal/mol) and as Figure 36 shows it consistently binds to the same place. However, when docked against the variants, it binds to 3 or 4 different locations. The most common conformations are still binding to the receptor binding domain; but the exact location is different from the location it binds to on the wild type and its binding affinity is significantly higher. Bond distances increase when docked against the variants as well. While they are in 2.5-3 Å range for the wild type, those numbers increase to 3.5-4 Å range for the variants. Simeprevir makes hydrogen bonds with Arg393, Lys417, Tyr505, and a pi bond with His34 of the wild type. These are all hotspots. However, when Simeprevir is docked against the variants it does not make

these bonds. The interactions for P.1 and B.1.135 are exactly the same; but they are slightly different from the interactions B.1.1.7 make. Ser494 is a common interaction amongst all variants. P.1 and B.1.351 make hydrogen bonds with Lys484, while B.1.1.7 makes a hydrogen bond with Glu484, which is the original residue before the mutation to Lys484. Since Simeprevir binds to the same area of all three variants, this area might be an active hotspot for the variants. Also, Ser494 from spike protein RBD normally interacts with Arg393 from the ACE2 and Arg393 is a hotspot. So, hydrogen bonds with Ser494 can disrupt the ACE2 spike protein bonding and inhibit the virus. (Choudhary, Malik et al. 2020). These predictions need to be studied further. Although there is reason for optimism, the variations in the binding locations on the variants need to be considered as well.

Figure 36. Binding affinity distributions and overlaid docking poses through 100 iterations of Simeprevir for (A) Wild Type, (B) B.1.1.7 Variant (C) B.1.351 Variant, and (D) P.1 Variant.

Saquinavir is an antiretroviral drug used to treat HIV. It has been suggested as a SARS-CoV-2 remedy in multiple studies (Bello, Martínez-Muñoz et al. 2020, Montenegro, Al-Abed et al. 2020). It has the most uniform interactions with the wild type and one of the lowest binding affinities (-9.7 kcal/mol). It makes hydrogen bonds with important hotspots of the wild type, His34, Arg403, and Lys417. It behaves more or less the same against the variants as well. Its binding affinity does not change much. It keeps making hydrogen bonds with His34, Arg403, and the mutated versions of Lys417. Although it does not make the Asp30 salt bridge with the variants, it makes pi stacking interactions with B.1.1.7 and B.1.351. It also makes more hydrophobic bonds with the variants. Also, the distances for its hydrogen bonds barely change and stay around 3 Å. Therefore, it is safe to say that Saquinavir does not get affected from the mutations.

Figure 37. Binding affinity distributions and overlaid docking poses through 100 iterations of Saquinavir for (A) Wild Type, (B) B.1.1.7 Variant (C) B.1.351 Variant, and (D) P.1 Variant.

Glecaprevir is another hepatitis C antiviral drug that binds to the wild type with a low binding affinity (-9.5 kcal/mol) and it consistently made the same bonds throughout the iterations. Glecaprevir makes important interactions with the wild type. It makes hydrogen bonds with Lys417, Tyr505, Arg393, and Arg403, which are all important residues. The average distance for these bonds is 2.5 Å, which makes this molecule a very promising candidate for the wild type. However, when docked against the variants, it exhibits a high variance in binding location and binding affinity. The affinity decreases significantly as well. When the most common conformations are selected and analyzed, Glecaprevir is still found to be interacting with the hotspots of B.1.1.7 and P.1, but not B.1.351. The hydrogen bonds it makes with the hotspots Arg393, Lys417, and Tyr505 of the wild type is carried over to the former variants. Additionally, these variants make halogen bonds with Asn33 and Tyr505 as well, which might suggest that it

might still be effective against them. However, Glecaprevir only makes a hydrogen bond with Glu35 of the B.1.351, which is not a good sign.

Figure 38. Binding affinity distributions and overlaid docking poses through 100 iterations of Glecaprevir for (A) Wild Type, (B) B.1.1.7 Variant (C) B.1.351 Variant, and (D) P.1 Variant.

Irinotecan is a cancer drug that binds to the wild type with -9.4 kcal/mol affinity. It exhibits uniform interactions across the iterations. However, this behavior changes when it is docked against the variants. Although most of the time it still binds to the RBD, sometimes it does not. The binding affinities against the variants are higher as well; but stable across the variants (-9.1 kcal/mol). The interactions with the variants are exactly the same, which makes this molecule interesting. Although Irinotecan no longer makes a hydrogen bond with the hotspot Lys417 of the variants, it makes a hydrogen bond with Tyr505 instead, which is another hotspot. It also keeps its hydrogen bond with Arg393, another hotspot, across the mutations. Aside from its hydrogen bonds, it makes pi bonds with Tyr505 and Arg403 of the variants; and salt bridges with His34 and

Arg403; all of which are hotspots. Although the average bond distance increases by 0.5 Å, Irinotecan makes more hydrogen, hydrophobic and pi bonds with the variants. When docked against the variants, it still interacts with hotspots, which suggests that it might be effective against them as well.

Figure 39. Binding affinity distributions and overlaid docking poses through 100 iterations of Irinotecan for (A) Wild Type, (B) B.1.1.7 Variant (C) B.1.351 Variant, and (D) P.1 Variant.

Indacaterol is used to treat bronchitis (Cazzola, Calzetta et al. 2014) and unlike the other molecules in the list, it only binds to ACE2. Since there are other drug candidates that can inhibit SARS-Cov-2 by only interacting with ACE2, Indacaterol was added to the final list. In this aspect, it is also similar to MeBlu from the experimental drugs. As expected Indacaterol does not get affected much from the mutations. In fact, in some of the 100 iterations it binds to the RBD of the P.1 and B.1.351 variants, which results in lower binding affinities. However, since these are not the common case, this study will concentrate on the ACE2 only binding poses. While the interactions of Indacaterol

are exactly the same for the wild type and B.1.1.7, they are slightly different for the other two variants. When docked against B.1.351 and P.1, Indacaterol makes a hydrogen bond with Arg393, which is a hotspot. Although it does not make the pi stacking interaction that it does with the wild type, it makes one extra hydrogen bond in both cases. Also, it makes more hydrophobic bonds as well. Given all these, Indacaterol is expected to as well or better against the variants.

Figure 40. Binding affinity distributions and overlaid docking poses through 100 iterations of Indacaterol for (A) Wild Type, (B) B.1.1.7 Variant (C) B.1.351 Variant, and (D) P.1 Variant.

Ergotamine is a smaller molecule used to treat migraines. When docked against the wild type it binds consistently to the same place but has a higher binding affinity than the other molecules (-8.1 kcal/mol). However, since it only binds to the spike protein and is a smaller molecule, this is not a big issue. When it is docked against the variants, although it sometimes binds to different places, majority of the time it still binds to the RDB of the spike protein. Its binding affinity changes only slightly against the variants as well. Ergotamine makes hydrogen bonds with Arg403

and Tyr505 of the wild type, both of which are hotspots. These interactions exist with all the variants, which is very positive. Although the hydrogen bond with Asp30 only exists with the wild type, Ergotamine still makes most of the hydrophobic bonds with the variants. Therefore, it is expected to be effective against the variants as well.

Figure 41. Binding affinity distributions and overlaid docking poses through 100 iterations of Ergotamine for (A) Wild Type, (B) B.1.1.7 Variant (C) B.1.351 Variant, and (D) P.1 Variant.

5.3 Conclusion

Some of the variants we have covered here, or others that will evolve from their lineages might become the dominant strain at some point. Therefore, when choosing molecules for repurposing, taking the mutations and variants into account is critical. Of the 10 molecules that were picked in chapter 4, most lose binding affinity against the new variants, especially B.1.351. Since the K417T mutation replaces a charged side chain with a polar uncharged one, the affinity of the spike protein decreases, and this makes it harder for both drug targets and antibodies to bind to the RBD. This

shows how dangerous the new mutations and variants are. It also casts doubt on the effectiveness of the current vaccination efforts and shows how important finding a treatment still is.

From the top 10 drugs, Saquinavir performs the best. It consistently binds to RBD, keeps its interactions with the hotspots across the variants and does not lose much affinity against the variants. Also, it is an anti-viral which is also a bonus.

Ergotamine and Indacaterol are in the second tier. They still bind to their original binding locations fairly consistently and do not lose much affinity. Ergotamine interacts mostly with the spike protein and keeps its hotspots across the variants. On the other hand, Indacaterol only binds to ACE2, and it does not get affected much from the changes on the spike protein. These two drugs are also used to treat relevant diseases, which is promising. Paritaprevir is in the next tier. It binds very consistently and strongly to the wild type. It keeps consistently binding to the same locations on the variants. However, it loses its hotspot interactions, which is why it is ranked lower. Voxilaprevir, which performs the best against the wild type and 2 of the variants but does poorly against B.1.351. Therefore, it needs to be studied further. Another drug that needs further research is Sirolimus. It binds fairly consistently to the variants. However, its interactions are not common or studied at all, which makes its performance questionable.

Although Irinotecan and Cabazitaxel binds well to the variants with consistently low affinities, since they are chemotherapy drugs, they rank lower in the list. If their side effects can be kept under control, they might be candidates. On the hand Simeprevir and Glecaprevir are not the best candidates. They lose a high percentage of the affinity against the variants and do not bind consistently to the same locations on the variants.

Conclusion

SARS-CoV-2 is a highly contagious and deadly viral disease which has caused a global pandemic. Despite increasing vaccination rates, it is spreading quickly and affecting our lives deeply. There are no approved novel treatments yet. Since drug development is a long and time-consuming process, repurposing an already approved drug would cut down the costs and research time considerably. Some drugs, like Remdesivir, have already been repurposed to treat the disease. These drugs mainly try to disrupt the viral replication. They treat the symptoms and decrease the recovery time but do not decrease the mortality rate or stop the infection.

SARS-CoV-2 virus is made up of 5 major components. The component that is responsible for infecting the cells is the spike protein. It attaches to ACE2, a protein on the cell membrane, to get entry to the cell. There are already some experimental molecules that were shown to inhibit the ACE2 - Spike protein interaction in vitro studies. However, they are in very early stages of the development. This study aims to identify FDA-approved drugs that behave similar to these molecules and stop the infection.

1750 FDA drugs were considered as candidates and analyzed with the molecular docking method. Docking predicts the optimal orientation and position that a molecule would take while making a bond with another molecule. It also quantifies the strength of the interaction and scores it. A negative score means that energy would be released when the two molecules make a bond. The smaller the score is, the more energy gets released, and the bond gets stronger.

The first step in the study was to verify that the docking methodology works as expected. Experimental molecules were docked against the ACE2-spike protein structure (PDB ID: 6M0J) using a docking tool called AutoDock Vina. Ligand efficiencies of all the experimental molecules

were calculated both from simulation results and wet-lab IC50 values. There was an 0.86 correlation between the two sets of the outputs, which showed a strong relationship between the simulations and wet-lab observations. Next, all of the FDA-approved molecules were docked against 6M0J using AutoDock Vina. Then results were rescored with another docking tool, Autodock 4.2, and similar scores were seen. Once scoring was also verified, 50 molecules with the lowest scores were picked for further analysis.

Since docking is a stochastic process, in order to reduce randomness in the results, the remaining molecules were docked 100 times. Five criteria were looked at to filter the remaining molecules: a low binding affinity to the 6M0J, having similar interactions as the experimental molecules, interacting with hotspot residues on 6M0J, having uniform docking results, and is considered as a potential treatment in other studies.

Some of the molecules like Elbasvir and Eltrombopag exhibited excellent binding affinities. However, they also exhibited significant variance in their binding domains during the iterations and did not always bind to the interface. These were filtered out.

Since the experimental DRI-C, MeBlu, and Ery B molecules were shown to inhibit SARS-CoV-2 *in vitro* studies, finding drugs that bind to similar places on the combined protein would increase the chances of success in clinical studies. Molecules like Rifabutin exhibited very small variance in their binding domain and decent affinity numbers, but shared little to no interactions with the experimental molecules. These were filtered out as well.

From the remaining molecules, which had decent affinity numbers and uniform binding domain distributions, 10 that has the most interactions with the hotspot residues (Asp30, Asp38, Glu35, Glu37, Thr27, Arg393, Lys 31, Lys353, His34 on ACE2 and Arg403, Lys417, Tyr505 on S-

protein) were picked. These top 10 ligands form various hydrogen, hydrophobic, halogen and pi bonds with the hotspot residues and can potentially inhibit the ACE2 receptor.

From the 10 top drugs, 5 are antiviral/antiretrovirals, one is an immunosuppressant drug; one is used to treat bronchitis; one is used for migraine, and two are used for cancer treatment. Ergotamine, Indacaterol, Paritaprevir, Sirolimus, Irinotecan, Saquinavir, and Simiprevir exhibited almost the same interactions in every docking iteration. The binding domains for Cabazitaxel, Voxilaprevir, and Glecaprevir showed a little bit of variance, but not much. All 10 molecules showed similarities with the experimental drugs and interacted with hotspots. Among the 10, Voxilaprevir exhibited the lowest binding affinity and highest number of interactions with hotspots. Indacaterol was the only molecule which only interacted with ACE2.

Once the final candidates were determined, their performances against different SARS-CoV-2 variants were studied. Binding affinities of the molecules consistently got worse against the variants. Some of the molecules no longer made bonds with known hotspots on the S-protein, and their binding domain variance increased. Mutations to the Lys417(E) were especially critical in the binding performance. For example, Voxilaprevir, which performed the best against the wild type, still performed well against the other two variants. But it lost about 1.5 kcal/mol in binding affinity against the B.1.351 variant, showed greater binding location variance, and could no longer bind to Asp38(A), His34(A), and Lys353(A) hotspots.

After the performance analysis against the variants, two other molecules, Simeprevir and Glecaprevir, were filtered out. Two of the remaining molecules, Irinotecan and Cabazitaxel, are chemotherapy drugs with known side effects. Although they could be used as a final option, their wide-spread use is not ideal. This brings the list down to 6.

Saquinavir behaved the most consistently across the variants. It had a decent binding affinity against all variants and showed little binding location variance against the variants. It interacted with both ACE2 and spike protein. It made bonds with His34(A), Arg403(E), and Lys417(E) hotspots. It was also suggested as a SARS-CoV-2 treatment in other publications. Therefore, it takes the top spot in the list.

Despite its reduced performance against B.1.351, Voxilaprevir, which is an antiviral drug, takes the second spot. It exhibited the lowest binding affinity scores against three of the four variants. It made bonds with His34(A), Asp38(A), Lys353(A), and Arg403(E) hotspots. It exhibited fairly low binding location variance. And it has 3 common interactions with the experimental molecules.

The next molecule in the list is Ergotamine. Although it exhibited a lower binding affinity than the other molecules in the list, it is also a smaller molecule, like EryB. It has a fairly low binding location variance across all variants. It made interactions with two hotspots on the spike protein, Arg403, and Tyr505, and this was consistent for all variants. It was also suggested as a SARS-CoV-2 treatment in other in silico studies as well.

The next molecule in the list is Paritaprevir, which is another antiviral drug. Paritaprevir exhibited low binding affinity scores and a very low binding location variance against all of the variants. It is also suggested as an RNA polymerase inhibitor against SARS-CoV-2, which is very promising. It made a high number of interactions with the wild type, one of which, Lys417(E), is a hotspot. Although it kept a high number of interactions, this hotspot interaction was not seen against the variants, which is why it is lower on the list.

The next molecule in the list is Indacaterol which is used to treat bronchitis. It exhibited a fairly high binding affinity score against all the variants and was not affected from the mutations. Unlike

the other molecules in our list, this molecule binds only to ACE2, which is similar to the experimental molecule, MeBlu. It made interactions with the hotspot Arg393(A) of all the variants. Also, another molecule that makes very similar interactions, Catechin, was shown to be effective against SARS-CoV-2 *in vivo* experiments.

The last molecule in the list is Sirolimus. Sirolimus binds very consistently to all of the variants and exhibited decent binding affinity scores across all the variants. However, it did not make as many interactions as the other molecules in the list. It made bonds with two hotspots, His34(A) and Lys417(E), on the wild type, but these bonds were not seen with the variants. More importantly, there is not much research on the interactions it made with the variants. Therefore, it remains a candidate but requires further research.

This study intends to pave the way for further *in vitro* and clinical studies of the molecules we have found to be promising against the SARS-CoV-2 and its variants. Our results suggest that the binding properties of these molecules make them potential inhibitors for SARS-CoV-2. Of our final six molecules, three are antivirals, and one is used to treat bronchitis, which is very promising. Three molecules from the top six scored high chemical similarity scores against the experimental molecules, which is another positive sign. However, all these molecules need to be researched further.

Going forward these molecules can be studied further by molecular dynamics. Molecular dynamics tools simulate the physical movements of molecules over time and can predict whether these molecules can bind to the spike protein in a stable manner. In parallel, these drugs can be tried against SARS-CoV-2 and its variants *in vitro* studies. Cells that express ACE2, active virions, and

different concentrations of the drugs from the list can be combined, and infection rates can be recorded. If these tests prove to be successful, further testing can be done in vitro studies.

Even if these molecules fail to inhibit SARS-CoV-2, they are still great reference points for further *de novo* drug discovery research. These molecules can be perturbed using generative machine learning networks such as GANs and resulting molecules can be put through fitness tests till more effective molecules can be found. In the same manner, an opposite approach can also be taken. Chemical structures of these molecules can be analyzed, and their sub-structures that interact with the spike protein can be isolated for bottom to top *de novo* drug design.

REFERENCES

- Abo-Zeid, Y., et al. (2020). "A molecular docking study repurposes FDA approved iron oxide nanoparticles to treat and control COVID-19 infection." European Journal of Pharmaceutical Sciences **153**: 105465.
- Absalan, A., et al. (2020). "Computation screening and molecular docking of FDA approved viral protease inhibitors as a potential drug against COVID-19." Gastroenterology and hepatology from bed to bench **13**(4): 355.
- Adcock, S. A. and J. A. McCammon (2006). "Molecular dynamics: survey of methods for simulating the activity of proteins." Chemical reviews **106**(5): 1589-1615.
- Alamdari, D. H., et al. (2020). "Application of methylene blue-vitamin C–N-acetyl cysteine for treatment of critically ill COVID-19 patients, report of a phase-I clinical trial." European journal of pharmacology **885**: 173494.
- Alexander, M. R., et al. (2020). "Predicting susceptibility to SARS-CoV-2 infection based on structural differences in ACE2 across species." The FASEB Journal **34**(12): 15946-15960.
- Ali, A. and R. Vijayan (2020). "Dynamics of the ACE2–SARS-CoV-2/SARS-CoV spike protein interface reveal unique mechanisms." Scientific reports **10**(1): 1-12.
- Allam, L., et al. (2020). "Targeting the GRP78-Dependant SARS-CoV-2 Cell Entry by Peptides and Small Molecules." Bioinformatics and biology insights **14**: 1177932220965505.
- Alsolamy, S. and Y. M. Arabi (2015). "Infection with Middle East respiratory syndrome coronavirus." Canadian journal of respiratory therapy: CJRT= Revue canadienne de la therapie respiratoire: RCTR **51**(4): 102.
- Aparoy, P., et al. (2012). "Structure and ligand based drug design strategies in the development of novel 5-LOX inhibitors." Current medicinal chemistry **19**(22): 3763-3778.

Armstrong, J., et al. (1984). "Sequence and topology of a model intracellular membrane protein, E1 glycoprotein, from a coronavirus." Nature **308**(5961): 751-752.

Artika, I. M., et al. (2020). "Molecular biology of coronaviruses: current knowledge." Heliyon: e04743.

Arun, K., et al. (2020). "Drug repurposing against SARS-CoV-2 using E-pharmacophore based virtual screening, molecular docking and molecular dynamics with main protease as the target." Journal of Biomolecular Structure and Dynamics: 1-12.

Assi, T., et al. (2020). "The management of patients with metastatic prostate cancer during the COVID-19 pandemic." Future Oncology **16**(20): 1455-1461.

Aykul, S. and E. Martinez-Hackert (2016). "Determination of half-maximal inhibitory concentration using biosensor-based protein interaction analysis." Analytical biochemistry **508**: 97-103.

Bajusz, D., et al. (2015). "Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations?" Journal of cheminformatics **7**(1): 1-13.

Batool, M., et al. (2019). "A structure-based drug discovery paradigm." International Journal of Molecular Sciences **20**(11): 2783.

Beavis, K. G., et al. (2020). "Evaluation of the EUROIMMUN Anti-SARS-CoV-2 ELISA Assay for detection of IgA and IgG antibodies." Journal of Clinical Virology **129**: 104468.

Beigel, J. H., et al. (2020). "Remdesivir for the treatment of Covid-19—preliminary report." The New England journal of medicine.

Bello, M., et al. (2020). "Identification of saquinavir as a potent inhibitor of dimeric SARS-CoV2 main protease through MM/GBSA." Journal of molecular modeling **26**(12): 1-11.

Beniac, D. R., et al. (2006). "Architecture of the SARS coronavirus prefusion spike." Nature structural & molecular biology **13**(8): 751-752.

Benítez-Cardoza, C. G. and J. L. Vique-Sánchez (2020). "Potential inhibitors of the interaction between ACE2 and SARS-CoV-2 (RBD), to develop a drug." Life sciences **256**: 117970.

Berg, K. A. and W. P. Clarke (2018). "Making sense of pharmacology: inverse agonism and functional selectivity." International Journal of Neuropsychopharmacology **21**(10): 962-977.

Bestle, D., et al. (2020). "TMPRSS2 and furin are both essential for proteolytic activation of SARS-CoV-2 in human airway cells." Life science alliance **3**(9).

Bojadzic, D., et al. (2020). "Methylene Blue Inhibits In Vitro the SARS-CoV-2 Spike–ACE2 Protein-Protein Interaction—A Mechanism That Can Contribute to Its Antiviral Activity Against COVID-19." bioRxiv.

Bojadzic, D., et al. (2020). "Small-Molecule In Vitro Inhibitors of the Coronavirus Spike-ACE2 Protein-Protein Interaction as Blockers of Viral Attachment and Entry for SARS-CoV-2." bioRxiv.

Bojadzic, D., et al. (2018). "Design, synthesis, and evaluation of novel immunomodulatory small molecules targeting the CD40–CD154 costimulatory protein-protein interaction." Molecules **23**(5): 1153.

Bosch, B. J., et al. (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.

Boyles, F., et al. (2020). "Learning from the ligand: using ligand-based features to improve binding affinity prediction." Bioinformatics **36**(3): 758-764.

Campana, P., et al. (2020). "Dendritic Cells and SARS-CoV-2 Infection: Still an Unclassified Connection." Cells **9**(9): 2046.

Cazzola, M., et al. (2014). "Use of indacaterol for the treatment of COPD: a pharmacokinetic evaluation." Expert opinion on drug metabolism & toxicology **10**(1): 129-137.

Cereto-Massagué, A., et al. (2015). "Molecular fingerprint similarity search in virtual screening." Methods **71**: 58-63.

Chen, H. and Q. Du (2020). "Potential natural compounds for preventing SARS-CoV-2 (2019-nCoV) infection."

Choubey, A., et al. (2020). "Naltrexone a potential therapeutic candidate for COVID-19." Journal of Biomolecular Structure and Dynamics: 1-8.

Choudhary, S., et al. (2020). "Identification of SARS-CoV-2 cell entry inhibitors by drug repurposing using in silico structure-based virtual screening approach." Frontiers in immunology **11**: 1664.

da Silva Hage-Melim, L. I., et al. (2020). "Virtual screening, ADME/Tox predictions and the drug repurposing concept for future use of old drugs against the COVID-19." Life sciences **256**: 117963.

De Vita, S., et al. (2020). "Accelerating the repurposing of FDA-approved drugs against coronavirus disease-19 (COVID-19)." RSC Advances **10**(67): 40867-40875.

- Debroise, T., et al. (2017). "A hybrid knowledge-based and empirical scoring function for protein–ligand interaction: SMOG2016." Journal of chemical information and modeling **57**(3): 584-593.
- Deshpande, R. R., et al. (2020). "In silico molecular docking analysis for repurposing therapeutics against multiple proteins from SARS-CoV-2." European journal of pharmacology **886**: 173430.
- Diehl, W. E., et al. (2016). "Ebola virus glycoprotein with increased infectivity dominated the 2013–2016 epidemic." Cell **167**(4): 1088-1098. e1086.
- Diller, D. J. and K. M. Merz Jr (2001). "High throughput docking for library design and library prioritization." Proteins: Structure, Function, and Bioinformatics **43**(2): 113-124.
- DiMasi, J. A., et al. (2003). "The price of innovation: new estimates of drug development costs." Journal of health economics **22**(2): 151-185.
- Dotolo, S., et al. (2021). "A review on drug repurposing applicable to COVID-19." Briefings in Bioinformatics **22**(2): 726-741.
- Draghici, S., et al. (2007). "A systems biology approach for pathway level analysis." Genome research **17**(10): 1537-1545.
- Du, Y. X. and X. P. Chen (2020). "Favipiravir: pharmacokinetics and concerns about clinical trials for 2019-nCoV infection." Clinical Pharmacology & Therapeutics **108**(2): 242-247.
- Duan, K., et al. (2020). "Effectiveness of convalescent plasma therapy in severe COVID-19 patients." Proceedings of the National Academy of Sciences **117**(17): 9490-9496.

Elfiky, A. A. (2020). "Ribavirin, Remdesivir, Sofosbuvir, Galidesivir, and Tenofovir against SARS-CoV-2 RNA dependent RNA polymerase (RdRp): A molecular docking study." Life sciences **253**: 117592.

Elmezayen, A. D., et al. (2020). "Drug repurposing for coronavirus (COVID-19): in silico screening of known drugs against coronavirus 3CL hydrolase and protease enzymes." Journal of Biomolecular Structure and Dynamics: 1-13.

Emery, S. L., et al. (2004). "Real-time reverse transcription–polymerase chain reaction assay for SARS-associated coronavirus." Emerging infectious diseases **10**(2): 311.

Enjuanes, L., et al. (2008). "Nidovirales." Encyclopedia of virology: 419.

Fan, H., et al. (2005). "The nucleocapsid protein of coronavirus infectious bronchitis virus: crystal structure of its N-terminal domain and multimerization properties." Structure **13**(12): 1859-1868.

Faria, N. R., et al. (2021). "Genomic characterisation of an emergent SARS-CoV-2 lineage in Manaus: preliminary findings." Virological.

Farouk, A.-E., et al. (2021). "Screening of inhibitors against SARS-CoV-2 spike protein and their capability to block the viral entry mechanism: A viroinformatics study." Saudi Journal of Biological Sciences.

Fehr, A. R. and S. Perlman (2015). "Coronaviruses: an overview of their replication and pathogenesis." Coronaviruses: 1-23.

Fischer, D., et al. (1993). "Surface motifs by a computer vision technique: searches, detection, and implications for protein–ligand recognition." Proteins: Structure, Function, and Bioinformatics **16**(3): 278-292.

Fischer, E. (1894). "Einfluss der Configuration auf die Wirkung der Enzyme. II." Berichte der deutschen chemischen Gesellschaft **27**(3): 3479-3483.

Forni, G. and A. Mantovani (2021). "COVID-19 vaccines: where we stand and challenges ahead." Cell Death & Differentiation **28**(2): 626-639.

Friesner, R. A., et al. (2004). "Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy." Journal of medicinal chemistry **47**(7): 1739-1749.

Ganesan, L., et al. (2011). "The food colorant erythrosine is a promiscuous protein–protein interaction inhibitor." Biochemical pharmacology **81**(6): 810-818.

Gangadevi, S., et al. (2021). "Kobophenol A Inhibits Binding of Host ACE2 Receptor with Spike RBD Domain of SARS-CoV-2, a Lead Compound for Blocking COVID-19." The journal of physical chemistry letters **12**(7): 1793-1802.

Ghahestani, S. M., et al. (2020). "Methylene blue may have a role in the treatment of COVID-19." Medical Hypotheses **144**: 110163.

Ghofrani, H. A., et al. (2006). "Sildenafil: from angina to erectile dysfunction to pulmonary hypertension and beyond." Nature reviews Drug discovery **5**(8): 689-702.

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

Goodsell, D. S. and A. J. Olson (1990). "Automated docking of substrates to proteins by simulated annealing." Proteins: Structure, Function, and Bioinformatics **8**(3): 195-202.

Guan, W.-j., et al. (2020). "Clinical characteristics of coronavirus disease 2019 in China." New England journal of medicine **382**(18): 1708-1720.

Guo, X., et al. (2020). "Investigation of the genetic variation in ACE2 on the structural recognition by the novel coronavirus (SARS-CoV-2)." Journal of Translational Medicine **18**(1): 1-9.

Gurung, A. B., et al. (2020). "In silico screening of FDA approved drugs reveals ergotamine and dihydroergotamine as potential coronavirus main protease enzyme inhibitors." Saudi Journal of Biological Sciences **27**(10): 2674-2682.

Gurung, A. B., et al. (2021). "Identification of potential SARS-CoV-2 entry inhibitors by targeting the interface region between the spike RBD and human ACE2." Journal of infection and public health **14**(2): 227-237.

Gurung, A. B., et al. (2021). "The potential of Paritaprevir and Emetine as inhibitors of SARS-CoV-2 RdRp." Saudi Journal of Biological Sciences **28**(2): 1426-1432.

Gyebi, G. A., et al. (2020). "Potential inhibitors of coronavirus 3-chymotrypsin-like protease (3CLpro): An in silico screening of alkaloids and terpenoids from African medicinal plants." Journal of Biomolecular Structure and Dynamics: 1-13.

Haider, N. (2010). "Functionality pattern matching as an efficient complementary structure/reaction search tool: an open-source approach." Molecules **15**(8): 5079-5092.

Halgren, T. A., et al. (2004). "Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening." Journal of medicinal chemistry **47**(7): 1750-1759.

Hamming, I., et al. (2004). "Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis." The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland **203**(2): 631-637.

Hanwell, M. D., et al. (2012). "Avogadro: an advanced semantic chemical editor, visualization, and analysis platform." Journal of cheminformatics **4**(1): 1-17.

Hart, T. N. and R. J. Read (1992). "A multiple-start Monte Carlo docking method." Proteins: Structure, Function, and Bioinformatics **13**(3): 206-222.

Heo, Y.-A. and E. D. Deeks (2018). "Sofosbuvir/velpatasvir/voxilaprevir: a review in chronic hepatitis C." Drugs **78**(5): 577-587.

Ho, J. S. Y., et al. (2020). "Topoisomerase 1 inhibition therapy protects against SARS-CoV-2-induced inflammation and death in animal models." bioRxiv.

Hopkins, A. L., et al. (2004). "Ligand efficiency: a useful metric for lead selection." Drug discovery today **9**(10): 430-431.

Horby, P. W., et al. (2020). "Lopinavir–ritonavir in patients admitted to hospital with COVID-19 (RECOVERY): a randomised, controlled, open-label, platform trial." The Lancet **396**(10259): 1345-1352.

Hosseiny, M., et al. (2020). "Radiology perspective of coronavirus disease 2019 (COVID-19): lessons from severe acute respiratory syndrome and Middle East respiratory syndrome." American Journal of Roentgenology **214**(5): 1078-1082.

Huang, N., et al. (2006). "Molecular mechanics methods for predicting protein–ligand binding." Physical Chemistry Chemical Physics **8**(44): 5166-5177.

Huang, R., et al. (2017). "Protein-ligand binding affinity determination by the waterLOGSY method: an optimised approach considering ligand rebinding." Scientific reports **7**(1): 1-6.

Huang, S.-Y., et al. (2010). "Scoring functions and their evaluation methods for protein–ligand docking: recent advances and future directions." Physical Chemistry Chemical Physics **12**(40): 12899-12908.

Huang, S. Y. and X. Zou (2006). "An iterative knowledge-based scoring function to predict protein–ligand interactions: I. Derivation of interaction potentials." Journal of computational chemistry **27**(15): 1866-1875.

Hui, D. S. and A. Zumla (2019). "Severe acute respiratory syndrome: historical, epidemiologic, and clinical features." Infectious Disease Clinics **33**(4): 869-889.

Hung, I. F.-N., et al. (2020). "Triple combination of interferon beta-1b, lopinavir–ritonavir, and ribavirin in the treatment of patients admitted to hospital with COVID-19: an open-label, randomised, phase 2 trial." The Lancet **395**(10238): 1695-1704.

Hunter, J. D. (2007). "Matplotlib: A 2D graphics environment." IEEE Annals of the History of Computing **9**(03): 90-95.

Hurst, K. R., et al. (2013). "Characterization of a critical interaction between the coronavirus nucleocapsid protein and nonstructural protein 3 of the viral replicase-transcriptase complex." Journal of virology **87**(16): 9159-9172.

Husain, A. and S. N. Byrareddy (2020). "Rapamycin as a potential repurpose drug candidate for the treatment of COVID-19." Chemico-biological interactions: 109282.

Idress, D. and V. Kumar (2021). "SARS-CoV-2 spike protein interactions with amyloidogenic proteins: Potential clues to neurodegeneration." Biochemical and Biophysical Research Communications.

Jack Clifton, I. and J. B. Leikin (2003). "Methylene blue." American journal of therapeutics **10**(4): 289-291.

Jain, A. N. (1996). "Scoring noncovalent protein-ligand interactions: a continuous differentiable function tuned to compute binding affinities." Journal of computer-aided molecular design **10**(5): 427-440.

Jena, A. B., et al. (2021). "Catechin and curcumin interact with S protein of SARS-CoV2 and ACE2 of human cell membrane: insights from computational studies." Scientific reports **11**(1): 1-14.

Jockusch, S., et al. (2020). "Sofosbuvir terminated RNA is more resistant to SARS-CoV-2 proofreader than RNA terminated by Remdesivir." Scientific reports **10**(1): 1-9.

Jones, G., et al. (1997). "Development and validation of a genetic algorithm for flexible docking." Journal of molecular biology **267**(3): 727-748.

Jorgensen, S. C., et al. (2020). "Remdesivir: review of pharmacology, pre-clinical data, and emerging clinical experience for COVID-19." Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy **40**(7): 659-671.

Kamitani, W., et al. (2009). "A two-pronged strategy to suppress host protein synthesis by SARS coronavirus Nsp1 protein." Nature structural & molecular biology **16**(11): 1134.

Kamitani, W., et al. (2006). "Severe acute respiratory syndrome coronavirus nsp1 protein suppresses host gene expression by promoting host mRNA degradation." Proceedings of the National Academy of Sciences **103**(34): 12885-12890.

Kastritis, P. L. and A. M. Bonvin (2013). "On the binding affinity of macromolecular interactions: daring to ask why proteins interact." Journal of The Royal Society Interface **10**(79): 20120835.

Kim, S., et al. (2016). "PubChem substance and compound databases." Nucleic acids research **44**(D1): D1202-D1213.

Kodadek, T. (2011). "The rise, fall and reinvention of combinatorial chemistry." Chemical communications **47**(35): 9757-9763.

Koshland Jr, D. (1958). "Application of a theory of enzyme specificity to protein synthesis." Proceedings of the National Academy of Sciences of the United States of America **44**(2): 98.

Kuntz, I., et al. (1999). "The maximal affinity of ligands." Proceedings of the National Academy of Sciences **96**(18): 9997-10002.

Kuntz, I. D., et al. (1982). "A geometric approach to macromolecule-ligand interactions." Journal of molecular biology **161**(2): 269-288.

Kuo, L. and P. S. Masters (2013). "Functional analysis of the murine coronavirus genomic RNA packaging signal." Journal of virology **87**(9): 5182-5192.

Kurczab, R., et al. (2018). "Salt bridge in ligand–protein complexes—systematic theoretical and statistical investigations." Journal of chemical information and modeling **58**(11): 2224-2238.

Kwon, S., et al. (2019). "Comprehensive ensemble in QSAR prediction for drug discovery." BMC bioinformatics **20**(1): 1-12.

Kyriakidis, N. C., et al. (2021). "SARS-CoV-2 vaccines strategies: a comprehensive review of phase 3 candidates." npj Vaccines **6**(1): 1-17.

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

Landrum, G. (2016). "RDKit: open-source cheminformatics <http://www.rdkit.org>." Google Scholar There is no corresponding record for this reference.

Laurie, A. T. and R. M. Jackson (2005). "Q-SiteFinder: an energy-based method for the prediction of protein–ligand binding sites." Bioinformatics **21**(9): 1908-1916.

Lauring, A. S. and R. Andino (2010). "Quasispecies theory and the behavior of RNA viruses." PLoS Pathog **6**(7): e1001005.

Li, Y., et al. (2014). "Structure of a conserved Golgi complex-targeting signal in coronavirus envelope proteins." Journal of Biological Chemistry **289**(18): 12535-12549.

Li, Z., et al. (2019). "Antiviral effects of simeprevir on multiple viruses." Antiviral research **172**: 104607.

Lin, X., et al. (2020). "A review on applications of computational methods in drug screening and design." Molecules **25**(6): 1375.

Lipinski, C. A., et al. (1997). "Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings." Advanced drug delivery reviews **23**(1-3): 3-25.

Liu, J. and R. Wang (2015). "Classification of current scoring functions." Journal of chemical information and modeling **55**(3): 475-482.

Liu, L., et al. (2020). "Potent neutralizing antibodies against multiple epitopes on SARS-CoV-2 spike." Nature **584**(7821): 450-456.

Liu, Y., et al. (2021). "The N501Y spike substitution enhances SARS-CoV-2 transmission." bioRxiv: the preprint server for biology.

Lo, Y.-C., et al. (2016). Quantitative Methods in System-Based Drug Discovery. Complex Systems, Sustainability and Innovation, IntechOpen.

Lobo-Galo, N., et al. (2020). "FDA-approved thiol-reacting drugs that potentially bind into the SARS-CoV-2 main protease, essential for viral replication." Journal of Biomolecular Structure and Dynamics: 1-9.

Lyne, P. D. (2002). "Structure-based virtual screening: an overview." Drug discovery today **7**(20): 1047-1055.

Martinez, C. R. and B. L. Iverson (2012). "Rethinking the term “pi-stacking”." Chemical Science **3**(7): 2191-2201.

Meng, E. C., et al. (1992). "Automated docking with grid-based energy evaluation." Journal of computational chemistry **13**(4): 505-524.

Meng, X.-Y., et al. (2011). "Molecular docking: a powerful approach for structure-based drug discovery." Current computer-aided drug design **7**(2): 146-157.

Mevada, V., et al. "Drug repurposing of approved drugs Elbasvir, Ledipasvir, Paritaprevir, Velpatasvir, Antrafenine and Ergotamine for combating COVID19."

Mikra, C., et al. (2017). "Molecular Docking and Structure Activity Relationship Studies of NSAIDs. What do they Reveal about IC50?" Letters in Drug Design & Discovery **14**(8): 949-958.

Milani, M., et al. (2021). "Combined in silico and in vitro approaches identified the antipsychotic drug lurasidone and the antiviral drug elbasvir as SARS-CoV2 and HCoV-OC43 inhibitors." Antiviral research **189**: 105055.

Miller, M. D., et al. (1994). "FLOG: a system to select ‘quasi-flexible’ligands complementary to a receptor of known three-dimensional structure." Journal of computer-aided molecular design **8**(2): 153-174.

Mobley, D. L. and K. A. Dill (2009). "Binding of small-molecule ligands to proteins:“what you see” is not always “what you get”." Structure **17**(4): 489-498.

Mohammadi, S., et al. (2020). "In silico Investigation on the Inhibiting Role of Nicotine/Caffeine by Blocking the S Protein of SARS-CoV-2 Versus ACE2 Receptor." Microorganisms **8**(10): 1600.

Moitessier, N., et al. (2008). "Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go." British journal of pharmacology **153**(S1): S7-S26.

Montenegro, M. F., et al. (2020). "HIV protease inhibitors saquinavir and nelfinavir are potent inhibitors of cathepsin L activity: a potential treatment for COVID-19 patients."

Morris, G. M., et al. (1998). "Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function." Journal of computational chemistry **19**(14): 1639-1662.

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

Mouchlis, V. D., et al. (2021). "Advances in De Novo Drug Design: From Conventional to Machine Learning Methods." International Journal of Molecular Sciences **22**(4): 1676.

Naik, V. R., et al. (2020). "Remdesivir (GS-5734) as a therapeutic option of 2019-nCoV main protease–in silico approach." Journal of Biomolecular Structure and Dynamics: 1-14.

Neuman, B. W., et al. (2011). "A structural analysis of M protein in coronavirus assembly and morphology." Journal of structural biology **174**(1): 11-22.

Ng, T. I., et al. (2018). "In vitro antiviral activity and resistance profile of the next-generation hepatitis C virus NS3/4A protease inhibitor glecaprevir." Antimicrobial agents and chemotherapy **62**(1).

Nguyen, L. S., et al. (2019). "Sirolimus and mTOR inhibitors: a review of side effects and specific management in solid organ transplantation." Drug safety **42**(7): 813-825.

Nieto-Torres, J. L., et al. (2014). "Severe acute respiratory syndrome coronavirus envelope protein ion channel activity promotes virus fitness and pathogenesis." PLoS Pathog **10**(5): e1004077.

Nojomi, M., et al. (2020). "Effect of Arbidol (Umifenovir) on COVID-19: a randomized controlled trial." BMC infectious diseases **20**(1): 1-10.

O'Boyle, N. M., et al. (2011). "Open Babel: An open chemical toolbox." Journal of cheminformatics **3**(1): 1-14.

Ohue, M., et al. (2014). "MEGADOCK 4.0: an ultra-high-performance protein-protein docking software for heterogeneous supercomputers." Bioinformatics **30**(22): 3281-3283.

Oprea, T. I., et al. (2011). "Drug repurposing from an academic perspective." Drug Discovery Today: Therapeutic Strategies **8**(3-4): 61-69.

Organization, W. H. (2019). Middle East respiratory syndrome coronavirus (MERS-CoV).

Ouwerkerk-Mahadevan, S., et al. (2016). "Drug-drug interactions with the NS3/4A protease inhibitor simeprevir." Clinical pharmacokinetics **55**(2): 197-208.

Pagadala, N. S., et al. (2017). "Software for molecular docking: a review." Biophysical reviews **9**(2): 91-102.

Pavlova, A., et al. (2021). "Critical interactions for SARS-CoV-2 spike protein binding to ACE2 identified by machine learning." bioRxiv.

Payne, S. (2017). "Family Coronaviridae." Viruses: 149.

Peng, L., et al. (2021). "Prioritizing antiviral drugs against SARS-CoV-2 by integrating viral complete genome sequences and drug chemical structures." Scientific reports **11**(1): 1-11.

Perez, S. and I. Tvaroška (2014). "Carbohydrate–protein interactions: Molecular modeling insights." Advances in carbohydrate chemistry and biochemistry **71**: 9-136.

Petersen, E., et al. (2020). "Comparing SARS-CoV-2 with SARS-CoV and influenza pandemics." The Lancet infectious diseases.

Ragab, D., et al. (2020). "The COVID-19 cytokine storm; what we know so far." Frontiers in immunology **11**: 1446.

Rahmat, M., et al. (2019). "Repositioning the Repurposed Drug, a Structural Study of Thalidomide Analogs." The Hematologist **16**(3).

Rarey, M., et al. (1996). "A fast flexible docking method using an incremental construction algorithm." Journal of molecular biology **261**(3): 470-489.

Reutlinger, M., et al. (2014). "Multi-objective molecular de novo design by adaptive fragment prioritization." Angewandte Chemie International Edition **53**(16): 4244-4248.

Roussel, Y., et al. (2020). "SARS-CoV-2: fear versus data." International journal of antimicrobial agents **55**(5): 105947.

Rudrapal, M., et al. (2020). Drug Repurposing (DR): An Emerging Approach in Drug Discovery. Drug Repurposing-Hypothesis, Molecular Aspects and Therapeutic Applications, IntechOpen.

Russell, B., et al. (2020). "Associations between immune-suppressive and stimulating drugs and novel COVID-19—a systematic review of current evidence." ecancermedicalscience **14**.

Sacramento, C. Q., et al. (2020). "The in vitro antiviral activity of the anti-hepatitis C virus (HCV) drugs daclatasvir and sofosbuvir against SARS-CoV-2." bioRxiv.

Salentin, S., et al. (2015). "PLIP: fully automated protein–ligand interaction profiler." Nucleic acids research **43**(W1): W443-W447.

Salmaso, V. and S. Moro (2018). "Bridging molecular docking to molecular dynamics in exploring ligand-protein recognition process: An overview." Frontiers in pharmacology **9**: 923.

Schnecke, V. and L. A. Kuhn (2000). "Virtual screening with solvation and ligand-induced complementarity." Virtual Screening: An Alternative or Complement to High Throughput Screening?: 171-190.

Schoeman, D. and B. C. Fielding (2019). "Coronavirus envelope protein: current knowledge." Virology journal **16**(1): 1-22.

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

Scigliano, G. and G. A. Scigliano (2021). "Methylene blue in covid-19." Medical Hypotheses **146**: 110455.

Sethi, A., et al. (2020). "Understanding the role of galectin inhibitors as potential candidates for SARS-CoV-2 spike protein: in silico studies." RSC Advances **10**(50): 29873-29884.

Shamsi, A., et al. (2020). "Glecaprevir and Maraviroc are high-affinity inhibitors of SARS-CoV-2 main protease: possible implication in COVID-19 therapy." Bioscience reports **40**(6).

Shang, J., et al. (2020). "Cell entry mechanisms of SARS-CoV-2." Proceedings of the National Academy of Sciences **117**(21): 11727-11734.

Shang, J., et al. (2020). "Structural basis of receptor recognition by SARS-CoV-2." Nature **581**(7807): 221-224.

Singhal, N. K., et al. (2020). "Nanotechnology-based approaches for the detection of SARS-CoV-2." Frontiers in Nanotechnology **2**: 6.

Sliwoski, G., et al. (2014). "Computational methods in drug discovery." Pharmacological reviews **66**(1): 334-395.

Song, W., et al. (2018). "Cryo-EM structure of the SARS coronavirus spike glycoprotein in complex with its host cell receptor ACE2." PLoS pathogens **14**(8): e1007236.

Sørensen, M. D., et al. (2006). "Severe acute respiratory syndrome (SARS): development of diagnostics and antivirals." Annals of the New York Academy of Sciences **1067**(1): 500.

Sottriffer, C. A., et al. (2008). "SFCscore: scoring functions for affinity prediction of protein–ligand complexes." Proteins: Structure, Function, and Bioinformatics **73**(2): 395-419.

Stahura, F. L. and J. Bajorath (2004). "Virtual screening methods that complement HTS." Combinatorial chemistry & high throughput screening **7**(4): 259-269.

Stohlman, S. A. and M. Lai (1979). "Phosphoproteins of murine hepatitis viruses." Journal of virology **32**(2): 672-675.

Takahashi, S., et al. (2015). "Nicotianamine is a novel angiotensin-converting enzyme 2 inhibitor in soybean." Biomedical Research **36**(3): 219-224.

Tanaka, S. and H. A. Scheraga (1976). "Medium-and long-range interaction parameters between amino acids for predicting three-dimensional structures of proteins." Macromolecules **9**(6): 945-950.

Tanaka, T., et al. (2012). "Severe acute respiratory syndrome coronavirus nsp1 facilitates efficient propagation in cells through a specific translational shutoff of host mRNA." Journal of virology **86**(20): 11128-11137.

Tanimoto, T. T. (1958). "Elementary mathematical theory of classification and prediction."

Tariq, A., et al. (2020). "Paromomycin: A potential dual targeted drug effectively inhibits both spike (S1) and main protease of COVID-19." International Journal of Infectious Diseases **98**: 166-175.

Thomas, E., et al. (2012). "The application and mechanism of action of ribavirin in therapy of hepatitis C." Antiviral Chemistry and Chemotherapy **23**(1): 1-12.

Thomas, P. D. and K. A. Dill (1996). "An iterative method for extracting energy-like quantities from protein structures." Proceedings of the National Academy of Sciences **93**(21): 11628-11633.

Trezza, A., et al. (2020). "An integrated drug repurposing strategy for the rapid identification of potential SARS-CoV-2 viral inhibitors." Scientific reports **10**(1): 1-8.

Trott, O. and A. J. Olson (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.

Unal, M. A., et al. (2020). "Ribavirin shows antiviral activity against SARS-CoV-2 and downregulates the activity of TMPRSS2 and the expression of ACE2 In Vitro." bioRxiv.

Vankadari, N. (2020). "Arbidol: A potential antiviral drug for the treatment of SARS-CoV-2 by blocking trimerization of the spike glycoprotein." International journal of antimicrobial agents **56**(2): 105998.

Veeramachaneni, G. K., et al. (2020). "Structural and simulation analysis of hotspot residues interactions of SARS-CoV 2 with human ACE2 receptor." Journal of Biomolecular Structure and Dynamics: 1-11.

Verdonk, M. L., et al. (2003). "Improved protein–ligand docking using GOLD." Proteins: Structure, Function, and Bioinformatics **52**(4): 609-623.

Volz, E., et al. (2021). "Transmission of SARS-CoV-2 Lineage B. 1.1. 7 in England: Insights from linking epidemiological and genetic data." medRxiv: 2020.2012. 2030.20249034.

Waller, J. V., et al. (2020). "Diagnostic tools for coronavirus disease (COVID-19): comparing CT and RT-PCR viral nucleic acid testing." American Journal of Roentgenology **215**(4): 834-838.

Walls, A. C., et al. (2020). "Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein." Cell **181**(2): 281-292. e286.

Wang, G., et al. (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., et al. (2020). "Structural and functional basis of SARS-CoV-2 entry by using human ACE2." Cell **181**(4): 894-904. e899.

Wang, R., et al. (2002). "Further development and validation of empirical scoring functions for structure-based binding affinity prediction." Journal of computer-aided molecular design **16**(1): 11-26.

Wang, X., et al. (2020). "The anti-influenza virus drug, arbidol is an efficient inhibitor of SARS-CoV-2 in vitro." Cell discovery **6**(1): 1-5.

Wang, Y., et al. (2020). "Coronaviruses: an updated overview of their replication and pathogenesis." Coronaviruses: 1-29.

Weiner, S. J., et al. (1984). "A new force field for molecular mechanical simulation of nucleic acids and proteins." Journal of the American Chemical Society **106**(3): 765-784.

Wibmer, C. K., et al. (2021). "SARS-CoV-2 501Y. V2 escapes neutralization by South African COVID-19 donor plasma." Nature medicine: 1-4.

Xiu, S., et al. (2020). "Inhibitors of SARS-CoV-2 entry: current and future opportunities." Journal of medicinal chemistry **63**(21): 12256-12274.

Yan, R., et al. (2020). "Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2." Science **367**(6485): 1444-1448.

Ye, Z.-W., et al. (2020). "Zoonotic origins of human coronaviruses." International journal of biological sciences **16**(10): 1686.

Zhao, L., et al. (2012). "Antagonism of the interferon-induced OAS-RNase L pathway by murine coronavirus ns2 protein is required for virus replication and liver pathology." Cell host & microbe **11**(6): 607-616.

Appendix A

A.1 Iteration Script

```
denizyo@LAPTOP-IVCUV7P9: ~/scripts
#!/bin/bash

export PATH=~/.autodock/autodock_vina_1_1_2_linux_x86/bin:$PATH

nameOfTheCompound=$1
numberOfIterations=$2

END=100

rm -rf $nameOfTheCompound
mkdir -p $nameOfTheCompound
cp -af conf_$nameOfTheCompound.txt $nameOfTheCompound/conf.txt
cp -af 6m0j.pdb $nameOfTheCompound
cp -af 6m0j.pdbqt $nameOfTheCompound
cp -af $nameOfTheCompound.pdbqt $nameOfTheCompound
cd $nameOfTheCompound

touch outputAll
for (( i=1; i<=$END; i++ ))
do
    echo "Iteration: $i / $NumberOfIterations"
    vina --config conf.txt --log logSO_$i.txt
    mv vina_outSO.pdbqt out$i.pdbqt
    cat logSO_$i.txt >> outputAll
done

~
```

A.2 Rescoring Script

```
denizyo@LAPTOP-IVCUV7P9: ~/scripts
#!/bin/bash

trap 'kill -9 $(jobs -p %1)' 2

currentPath=$PWD
dockoutputFile="$currentPath/status.log"
dockresultFile="$currentPath/result.csv"

rm -rf "$dockoutputFile"
touch "$dockoutputFile"
rm -rf "$dockresultFile"
touch "$dockresultFile"

autoscore() {
    if [ -d "$folder" ]; then
        echo "Autoscore: $folder"
        echo "===== "
        cd $folder
        for ligand in *_out_ligand*; do
            ligandName=$(echo "$ligand" | cut -f1 -d".")
            if [ "$ligandName" == "$ligand" ]; then
                continue
            fi
            echo "Folder: $folder Ligand: $ligandName ligand"
            rm -rf *.gpf *.fld *.xyz *.dpf *.log *.map *.bak 'out-$ligandName'
            mkdir -p 'out-$ligandName'
            cd 'out-$ligandName'
            cp -af ../$ligand ./
            cp -af ../6m0j.* ./
            pythonsh ~/autodock/mgltools_x86_64Linux2_1.5.6/MGLToolsPckgs/AutoDockTools/Utilities24/prepare_gpf4.py -l $ligand -r 6m0j.pdbqt -y \
                || (echo -e "$ligandName \t\t\t FAIL" >> "$dockoutputFile" && continue)
            sed -i '/^npts/d' 6m0j.gpf
            sed -i '1s/^/npts 60 60 60\n/' 6m0j.gpf
            autogrid4 -p 6m0j.gpf -l gridmaps.log \
                || (echo -e "$ligandName \t\t\t FAIL" >> "$dockoutputFile" && continue)
            pythonsh ~/autodock/mgltools_x86_64Linux2_1.5.6/MGLToolsPckgs/AutoDockTools/Utilities24/prepare_dpf4.py -l $ligand -r 6m0j.pdbqt \
                || (echo -e "$ligandName \t\t\t FAIL" >> "$dockoutputFile" && continue)
            sed -i '/seed pid time/d' $ligandName_6m0j.dpf'
            sed -i '/small molecule center/{q}' $ligandName_6m0j.dpf'
        done
    fi
}

"run_autoscore.sh" 76L, 2586C
```

1,1

Top

A.3 Parsing Script

```
denizyo@LAPIOP-IVCUI7P9: ~/scripts
#!/bin/bash

tmpfolder=/tmp/dockresults
targetfile=outputAll
export PATH=$PWD:$PATH
currentdir=$PWD

rm -rf $tmpfolder
mkdir $tmpfolder
cp -af $1 $tmpfolder/$targetfile

cd $tmpfolder
dos2unix $targetfile && /dev/null && /dev/null
cat $targetfile | grep " 1 " | awk '{ print $2 }' > $tmpfolder/1.txt
cat $targetfile | grep " 2 " | awk '{ print $2 }' > $tmpfolder/2.txt
cat $targetfile | grep " 3 " | awk '{ print $2 }' > $tmpfolder/3.txt
cat $targetfile | grep " 4 " | awk '{ print $2 }' > $tmpfolder/4.txt
cat $targetfile | grep " 5 " | awk '{ print $2 }' > $tmpfolder/5.txt
cat $targetfile | grep " 6 " | awk '{ print $2 }' > $tmpfolder/6.txt
cat $targetfile | grep " 7 " | awk '{ print $2 }' > $tmpfolder/7.txt
cat $targetfile | grep " 8 " | awk '{ print $2 }' > $tmpfolder/8.txt
cat $targetfile | grep " 9 " | awk '{ print $2 }' > $tmpfolder/9.txt
cat $targetfile | grep " 1 " | awk '{ print $3 }' > $tmpfolder/1x.txt
cat $targetfile | grep " 2 " | awk '{ print $3 }' > $tmpfolder/2x.txt
cat $targetfile | grep " 3 " | awk '{ print $3 }' > $tmpfolder/3x.txt
cat $targetfile | grep " 4 " | awk '{ print $3 }' > $tmpfolder/4x.txt
cat $targetfile | grep " 5 " | awk '{ print $3 }' > $tmpfolder/5x.txt
cat $targetfile | grep " 6 " | awk '{ print $3 }' > $tmpfolder/6x.txt
cat $targetfile | grep " 7 " | awk '{ print $3 }' > $tmpfolder/7x.txt
cat $targetfile | grep " 8 " | awk '{ print $3 }' > $tmpfolder/8x.txt
cat $targetfile | grep " 9 " | awk '{ print $3 }' > $tmpfolder/9x.txt
paste -d, 1.txt 2.txt 3.txt 4.txt 5.txt 6.txt 7.txt 8.txt 9.txt 1x.txt 2x.txt 3x.txt 4x.txt 5x.txt 6x.txt 7x.txt 8x.txt 9x.txt > output.csv
cd - && /dev/null
cp -af $tmpfolder/output.csv ./
rm -rf $tmpfolder

~
~
~
~
~
"parse.sh" 35L, 1751C 1,1 All
```