STUDENT PAPER: Computational Analysis of SARS-CoV-2 Therapeutics Development

Samuel Biggerstaff
Department of Chemistry
Centre College
Danville, KY

Jennifer L. Muzyka
Department of Chemistry
Centre College
Danville, KY
jennifer.muzyka@centre.edu

David Toth
Department of Computer Science
Centre College
Danville, KY
david.toth@centre.edu

ABSTRACT

SARS-CoV-2 (also known as COVID-19) is a coronavirus that has recently emerged and impacted nearly every human on the planet. The nonstructural protein 12 (NSP 12) is an RNA-dependent RNA polymerase that replicates viral RNA in a cell to infect it. Interrupting this function should prohibit the virus from replicating within the body and would decrease the severity of the virus's effects in patients. The objective of this project is to identify potential inhibitors for NSP 12 that might be suitable as antiviral drugs. Thus, we obtained the structure of NSP 12 from RCSB's protein data bank. The protein structure was analyzed using computer software (Chimera and PyRx), and ligands obtained from the ZINC database and RCSB's protein data bank were docked to NSP 12. The resulting binding affinities were recorded, and binding geometries analyzed.

KEYWORDS

Virtual Screening, Drug Discovery, AutoDock Vina

1 INTRODUCTION

Since the end of 2019, the virus SARS-CoV-2, also known as Covid 19, has permeated throughout the cultural, political, and medical fields of nearly every country. The emergence of this virus has altered the day-to-day life of many as they attempt to avoid being infected by SARS-CoV-2. As a result, chemists, biochemists, biologists, and other medical scientists have directed their attention to SARS-CoV-2, its composition, effects, and treatments. Therapeutic treatments are currently of particular interest to the medical community, and two drugs, Molnupiravir and PF-07321332, show promising inhibitory effects against SARS-CoV-2 [12,13]. Additionally, Remdesivir, the only drug currently approved by the FDA for the treatment of SARS-CoV-2, is not as effective as desired [4]. As such, there is a great need for the further development of drugs that would inhibit SARS-CoV-2.

SARS-CoV-2 contains a variety of nonstructural proteins (NSP), each exhibiting their own form and function. These proteins, which are observed on the inside of the host cell, mediate the seven steps of viral replication [17]. Most of these proteins are essential for viral replication. Specifically, several of the 16 NSPs are exceptional drug targets. Protein targets were evaluated on necessity for the virus to replicate, uniqueness of structure from host cell proteins, and conservation of protein sequence for SARS-CoV to SARS-CoV-2. Assessment of the differences between the SARS-CoV-2 protein and host cell proteins is to help reduce side effects. Targeting a viral protein that has a similar structure to the host protein will result in high IC50 values and limited effectiveness. The basis for using the conservation of protein sequence between SARS-CoV to SARS-CoV-2 is that a protein that is mutating quickly will not be a good drug target because mutations can affect drug affinity and binding. Conservation between SARS-CoV to SARS-CoV-2 does not guarantee that there won't be mutations in the active site of the protein that will change binding affinity. It does give a better chance that there won't be a random mutation at any point, including the active site.

Our selected target is the nonstructural protein 12, or NSP 12. This NSP is the RNA-dependent RNA polymerase, meaning that NSP 12 uses RNA as a template to replicate the genome of SARS-CoV-2. Inhibiting NSP 12 would decrease the replication rate, reducing the symptoms of SARS-CoV-2 [4]. The NSPs of SARS-CoV-2 serve as good inhibition targets for therapeutics because of their significant roles in the function of the virus [14]. The NSPs in different variants of SARS-CoV-2 do not differ significantly, so drugs targeting the NSPs of SARS-CoV-2 will inhibit significant functions and will be effective across all observed variants [18]. More significantly, NSP 12 is highly conserved between SARS-CoV and SARS-CoV-2 [10]; thus, it is likely that inhibitors for SARS-CoV-2 will be effective for other coronaviruses [19].

Currently, few drugs have been proven to be successful in inhibiting this nonstructural protein. This is, in part, because there are millions of small molecules that could potentially be used as drugs. Deciding which is the best through experimentation alone would be an extremely long task that would not satisfy the urgent need for SARS-CoV-2 therapeutics. A solution to this dilemma is to take advantage of virtual screening to narrow down the list in a shorter period before beginning experimental trials. One computational program which aids in drug discovery is AutoDock Vina, an open-source program for molecular docking [16]. AutoDock Vina calculates binding affinity between proteins and small molecules in kilocalories per mol (kcal/mol) with a larger negative number indicating a greater binding affinity. For reference, Remdesivir is a nucleotide analog with a binding affinity of -7.8 kcal/mol in our calculations. Remdesivir is a delayed chain terminator that blocks transcription [4]. Since Remdesivir has been shown to be effective in treating patients infected with SARS-CoV-2, any small molecule with a stronger binding affinity might be at least somewhat effective in the treatment of SARS-CoV-2.

2 METHODS

First, the structure of the NSP 12 of SARS-CoV-2 (6YYT) was obtained from the RSCB Protein Data Bank and the structural file was analyzed and optimized in Chimera [4,11].

DOI: https://doi.org/10.22369/issn.2153-4136/14/1/9

The protein NSP12 crystal structure was selected from the Protein Data Bank (PDB). The structure selected was reported by Mariano et al. [9]. The structure selected also had NSP7 and NSP8 bound in the crystal structure. For modeling to get the protein small enough to be functionally useful for the software, NSP7 and NSP8 were manually removed. The ligands to be tested were selectively chosen from the PDB database. Ligands that were nucleotide triphosphates (NTP) or nucleotide monophosphate (NMP) derivatives were chosen for testing. The selected ligand file was downloaded as an SDF file from the PDB website.

Then, two studies were conducted: the targeted study and the general study. In the targeted study, small molecules with a similar structure to RNA were chosen and docked to NSP 12. The bounding box, which tells the docking software where the ligand should be placed, was determined by locating the binding site for RNA on NSP 12. PyRx, a front-end interface for AutoDock Vina, was used to dock these molecules [2]. The chosen molecules were either nucleotides or nucleotide derivatives because NSP 12 typically binds to nucleotides in the body. In the general study, small molecules from the ZINC Database were collected and docked. We downloaded .gz files containing multiple compounds from the ZINC database [5,20]. Then, we wrote a series of Python programs to automate the process [15]. The programs unzipped all the downloaded files, producing a series of text files, each containing multiple compounds. The programs then split the text files into the individual pdbqt files. Next, the programs assembled a new commands text file with one command to run AutoDock Vina per compound. The screening was then started in parallel on a multi-CPU server using xjobs [8]. When there were power failures at various points during the screening, another Python program was run to rebuild the commands text file without the compounds that had already been screened and run it again, so the screen could resume where it had left off. When the screening was complete, a final program was used to assemble a text file with all the compounds and their scores and sorted using the Linux sort command to produce a new file with all the compounds and their scores in the order from the best to worst.

3 RESULTS AND DISCUSSION

3.1 Targeted Study

In the targeted study, 45 nucleotide or nucleotide derivatives were docked to NSP 12 using PyRx, with many molecules demonstrating some compatibility. Of those results, 17 ligands had binding energies more negative than the value calculated for Remdesivir (-7.8 kcal/mol) and 2 met our desired target of -9 kcal/mol (more negative values indicate stronger binding). The best binding ligand was 7-methyl-guanosine-5'-triphosphate-5'-(2'-o-methyl)-adenosine (**V9G**) with a binding affinity of -9.1 kcal/mol (Figure 1). All of the ligands analyzed in the targeted study were nucleotides or nucleotide derivatives, so nearly all of the ligands contained a phosphate group, a 5-carbon sugar, and a nitrogenous base. As a result, the differences in conformations and additional atoms can be analyzed.

For instance, V9G and GTA are very similar in structure (Figure 2). The only difference between the two structures is the presence of an ether or alcohol. In V9G, there is an ether in the place of GTA's alcohol suggesting that a stronger electrostatic negative charge provided by the alcohol group in that location is detrimental to the binding affinity of the ligand. This accounts for a difference of 0.5 kcal/mol, resulting in a significant difference in binding affinity.

Most of the best ligands from the targeted study contained a triphosphate group, and six of the top ten ligands exhibited structures that were similar to those of V9G and GTA with minor differences. The length of the phosphate group may be important in the inhibition of NSP 12 because it creates a molecule of the appropriate size to fit into the active site.

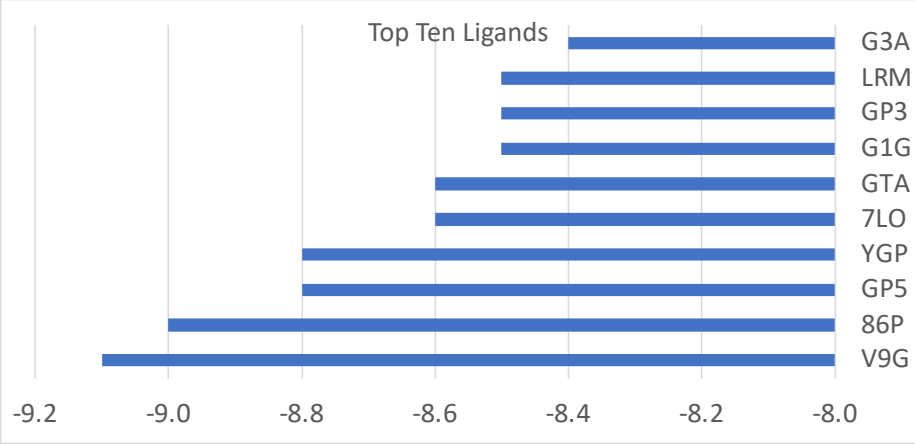


Figure 1. The top ten ligands with the best binding affinity to NSP-12 of SARS-CoV-2 calculated in PyRx from the targeted study.

3.2 General Study

10,582,294 molecules were screened using AutoDock Vina. Of them, about 3,000 were above the desired threshold for being desirable drugs in the inhibition of NSP 12. The best ligand was ZINC000004783172 with a binding affinity of -11.6 kcal/mol.

NSP 12 binds to RNA in the active site. As a result, it is expected that nucleotides and nucleotide derivatives are the ligands that would bind best. Surprisingly, very few of the top-ranking ligands in the general study had characteristics of a nucleotide, and the non-nucleotide molecules of the general study have stronger binding affinities than the nucleotides of the targeted study. There were no phosphate groups in any of the top-ranking ligands, as shown in Table 1. However, when considering that binding affinity is determined by many intermolecular forces such as electrostatic interactions, Van der Waals forces, and hydrogen bonding, small molecules could strongly bind to portions of the protein differently than the nucleotides that are normally found in the binding site.

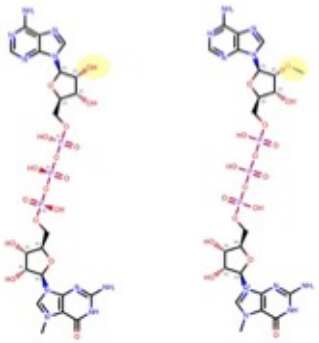
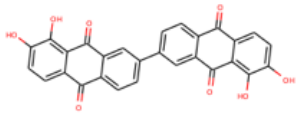
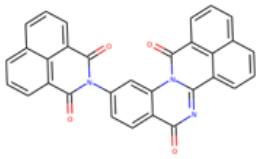
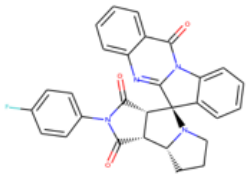
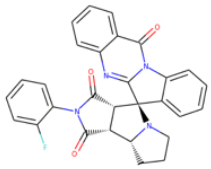
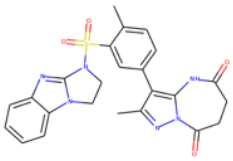
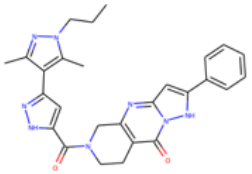
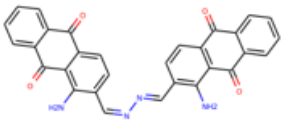
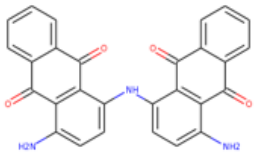
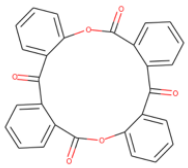


Figure 2. The molecular structures of the ligands GTA and V9G respectively, with their highlighted difference.

The ligand that bound the best, ZINC000004783172, can also be referred to as 7,7'-Bializarin. This molecule is currently being studied as an antibiotic.[20] If this molecule continues to show promising results, then it may be one of the best options for a future therapeutic. According to the calculations from AutoDock Vina, 7,7'-Bializarin should span the entire active site of NSP 12 (Figure 3). This is additionally promising as it ensures that the RNA does not have a location to bind to within the protein's active site, and therefore, cannot replicate.

Table 1. The top nine ligands with the best binding affinity to NSP-12 of SARS-CoV-2 calculated in Autodock Vina from the final general study.

ZINC ID, Binding affinity (kcal/mol)	Structure	ZINC ID, Binding affinity (kcal/mol)	Structure
ZINC000004783172 -11.6		ZINC000004015296 -11.4	
ZINC000035385140 -11.3		ZINC000101500434 -11.2	
ZINC000033122972 -11.2		ZINC000097137247 -11.2	
ZINC000004701175 -11.2		ZINC000003861401 -11.2	
ZINC000004974498 -11.2			

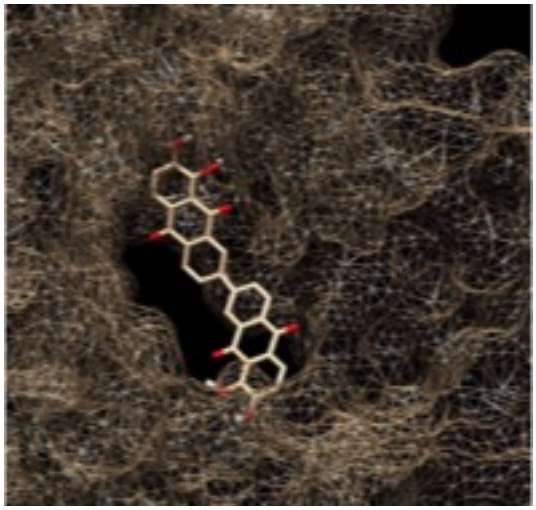


Figure 3. The three-dimensional representation of 7,7'-Bializarin binding to the active site of NSP 12

4 CONCLUSIONS

Using computational programs to find the binding affinities between ligands and the active sites of nonstructural proteins was successful in identifying a large range of ligands that could act as good drugs for SARS-CoV-2. Further research is planned into attempting to identify more ligands that could bind to NSP 12. Additionally, these ligands will be analyzed extensively in Chimera in order to identify the intermolecular forces and other potential causes of good binding affinity. Finally, since these nonstructural proteins are mostly conserved over viruses within the same families, the final drug produced from this research could be extremely effective in treating all SARS viruses, including those that may arise in the future.

5 REFLECTIONS

This project was my first experience with computational research. When I decided to join this project, I was already interested in computational chemistry but did not know the extent of how computational science could bring insights into the mechanics of the real world. At first, the research was intimidating. I was familiar with physical chemistry, but my knowledge of biology and biochemistry was lacking. Additionally, I had not taken any computer science classes, and I was not familiar with the software we used in the research. However, as we began to study the SARS-CoV-2 NSPs, I became fascinated by how different conformations and locations of small molecules could change binding affinities, and therefore have significant impacts on the protein's functionality. Of course, the research was not always easy, and we encountered many obstacles along the way. Still, I am thankful for experience, including the hardships, as I am now much more knowledgeable about proteins, SARS-CoV-2, and generally, how to do research. This project inspired me to pursue studying chemical systems using computational tools at the graduate level.

REFERENCES

[1] Juan Marcelo Carpio Arévalo and Juliana Carolina Amorim. 2021. An in-silico analysis reveals 7,7'-bializarin as a promising DNA gyrase B inhibitor on gram-positive and gram-negative bacteria. *Comput. Biol. Med. 135*, 104626. https://doi.org/10.1016/j.compbiomed.2021.104626

[2] Sargis Dallakyan and Arthur J. Olson. 2015. Small-molecule library screening by docking with PyRx. In *Chemical Biology. Methods in Molecular Biology Vol. 1263*. Humana Press, New York, NY, 243-250. https://doi.org/10.1007/978-1-4939-2269-7_19

[3] Richard T. Eastman, Jacob S. Roth, Kyle R. Brimacombe, Anton Simeonov, Min Shen, Samarjit Patnaik, and Matthew D. Hall. 2020. Remdesivir: A review of its discovery and development leading to emergency use authorization for treatment of COVID-19. *ACS Cent. Sci. 2020*, 6, 672-683. https://doi.org/10.1021/acscentsci.0c00489

[4] Hauke S. Hillen, Goran Kokic, Lucas Farnung, Christian Dienemann, Dimitry Tegunov, and Patrick Cramer. 2020. Structure of replicating SARS-CoV-2 polymerase. *Nature 584*, 154-156. https://doi.org/10.1038/s41586-020-2368-8

[5] John J. Irwin, Teague Sterling, Michael M. Mysinger, Erin S. Bolstad, and Ryan G. Coleman. 2012. ZINC: A free tool to discover chemistry for biology. *J. Chem. Inf. Model. 52*, 7, 1757-1768. https://doi.org/10.1021/ci3001277

[6] Lindsey S. Jung, Tamara M. Gund, and Mahesh Narayan. 2020. Comparison of binding site of remdesivir and its metabolites with NSP12-NSP7-NSP8, and NSP3 of SARS CoV-2 virus and alternative potential drugs for COVID-19 treatment. *Protein J. 39*, 619-630. https://doi.org/10.1007/s10930-020-09942-9

[7] Robert N. Kirchdoerfer and Andrew B. Ward. 2019. Structure of the SARS-CoV Nsp12 polymerase bound to Nsp7 and Nsp8 co-factors. *Nat. Commun. 10*, 2342. https://doi.org/10.1038/s41467-019-10280-3

[8] Thomas Maier-Komor. n.d. Home of the xjobs Utility. http://www.maier-komor.de/xjobs.html

[9] Giuseppina Mariano, Rebecca J. Farthing, Shamar L. M. Lale-Farjat, and Julien R. C. Bergeron. Structural characterization of SARS-CoV-2: Where we are, and where we need to be. *Front. Mol. Biosci. 7*, 605236. https://doi.org/10.3389/fmolb.2020.605236

[10] Ozal Mutlu, Osman Mutluhan Ugure, Emrah Sariyer, Oguz Ata, Tugba Gul Inci, and Erennur Ugurel. 2022. Targeting SARS-CoV-2 Nsp12/Nsp8 interaction interface with approved and investigational drugs: An *in silico* structure-based approach. *J. Biomol. Struct. Dyn. 40,* 2, 918-930. https://doi.org/10.1080/07391102.2020.1819882

[11] Eric F. Pettersen, Thomas D. Goddard, Conrad C. Huang, Gregory S. Couch, Daniel M. Greenblatt, Elaine C. Meng, and Thomas E. Ferrin. UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem. 25,* 13, 1605-1612. https://doi.org/10.1002/jcc.20084

[12] Carlos A. Ramos-Guzmán, J. Javier Ruiz-Pernía, and Iñaki Tuñón. Computational simulations on the binding and reactivity of a nitrile inhibitor of the SARS-CoV-2 main protease. *Chem. Commun. 57,* 72, 9096-9099. https://doi.org/10.1039/D1CC03953A

[13] Awadhesh Kumar Singh, Akriti Singh, Ritu Singh, and Anoop Misra. Molnupiravir in COVID-19: A systematic review of literature. *Diabetes Metab. Syndr. Clin. Res. Rev. 15,* 6, 102329. https://doi.org/10.1016/j.dsx.2021.102329

[14] E.J. Snijder, E. Decroly, and J. Ziebuhr. Chapter 3 - The nonstructural proteins directing coronavirus RNA synthesis and processing. In *Advances in Virus Research 96*, 59-126. https://doi.org/10.1016/bs.aivir.2016.08.008

[15] Dave Toth. n.d. Drug_discovery_python_scripts. https://github.com/DaveToth/drug_discovery_python_scripts

[16] Oleg Trott and Arthur J. Olson. 2010. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem. 31*, 2, 455-461. https://doi.org/10.1002/jcc.21334

[17] Philip V'kovski, Annika Kratzel, Silvio Steiner, Hanspeter Stalder, and Volker Thiel. 2021. Coronavirus biology and replication: Implications for SARS-CoV-2. *Nat. Rev. Microbiol. 19,* 3, 155-170. https://doi.org/10.1038/s41579-020-00468-6

[18] Deepa Vasireddy, Rachana Vanaparthy, Gisha Mohan, Srikrishna Varun Malayala, and Paavani Atluri. 2021. Review of COVID-19 variants and COVID-19 vaccine efficacy: What the clinician should know? *J. Clin. Med. Res. 13*, 6, 317-325. https://doi.org/10.14740/jocmr4518

[19] Francis K. Yoshimoto. The proteins of severe acute respiratory syndrome coronavirus-2 (SARS CoV-2 or n-COV19), the cause of COVID-19. *Protein J. 39,* 3, 198-216. https://doi.org/1007/s10930-020-09901-4

[20] ZINC[12]. n.d. Welcome to ZINC - A database of commercially-available compounds. https://zinc12.docking.org