College of Saint Benedict and Saint John's University
DigitalCommons@CSB/SJU

CSBSJU Distinguished Thesis | Undergraduate Research

4-2022

The development of inhibitors for SARS-CoV-2 ORF8

My Thanh Thao Nguyen
College of Saint Benedict/Saint John's University, MNGUYEN001@CSBSJU.EDU

Follow this and additional works at: https://digitalcommons.csbsju.edu/ur_thesis

 Part of the Amino Acids, Peptides, and Proteins Commons, Biochemistry Commons, Bioinformatics Commons, Biophysics Commons, Medicinal and Pharmaceutical Chemistry Commons, Medicinal Chemistry and Pharmaceutics Commons, Medicinal-Pharmaceutical Chemistry Commons, Molecular Biology Commons, Other Biochemistry, Biophysics, and Structural Biology Commons, Pharmaceutics and Drug Design Commons, and the Structural Biology Commons

Recommended Citation

Nguyen, My Thanh Thao, "The development of inhibitors for SARS-CoV-2 ORF8" (2022). *CSBSJU Distinguished Thesis*. 19.
https://digitalcommons.csbsju.edu/ur_thesis/19

This Thesis is brought to you for free and open access by DigitalCommons@CSB/SJU. It has been accepted for inclusion in CSBSJU Distinguished Thesis by an authorized administrator of DigitalCommons@CSB/SJU. For more information, please contact digitalcommons@csbsju.edu.

The development of inhibitors for SARS-CoV-2 ORF8

Anna (My) Nguyen

The College of Saint Benedict and Saint John's University

mnguyen001@csbsju.edu

Abstract

An unexpected outbreak of SARS-CoV-2 caused a worldwide pandemic in 2020. Many repurposed drugs were tested, but there are currently only three FDA approved antivirals (Merck's antiviral Molnupiravir, Pfizer's antiviral Paxlovid, and Remdisivir).[1] Most of the antiviral drugs tested SARS-CoV-2 main protease and RNA-dependent RNA polymerase. However, it is important to explore different drug targets of SARS-CoV-2 to prepare for the virus mutations of the future. This research looks at an alternative approach in which SARS-CoV-2 Open Reading Frame 8 (ORF8), which has been shown to be a rapidly evolving hypervariable gene, was chosen to be the protein of interest. A series of computational strategies were developed to generate pharmacophores and identify lead compounds. In addition to the lead compound identification pipeline, this thesis also presents an automated method for generating pharmacophore models from molecular docking output. The pharmacophore-based models resulted in four potential FDA-approved compounds. This thesis focuses on the binding activity of one of the four compounds, novobiocin. To collect data on the binding activity, the ORF8 protein was expressed and purified. Results suggest that novobiocin binds to ORF8, and it might be a potential inhibitor for further developments of an antiviral for SARS-CoV-2.

Approved by

Lisa Gentile

Thesis Advisor

Professor of Chemistry

Edward McIntee

Faculty Reader

Professor of Chemistry

Lisa Engstrom

Faculty Reader

Professor of Chemistry

Heather Amthauer

Faculty Reader

Professor of Computer Science

Imad Rahal

Faculty Reader

Professor of Computer Science

Peter Ohmann

Faculty Reader

Professor of Computer Science

Contents

List of Figures

List of Algorithms

1 Background

1.1 SARS-CoV-2 and currently available antivirals

As of March of 2022, the pandemic caused by severe acute respiratory syndrome coronavirus (SARS-CoV-2) has resulted in more than four hundred and forty-six million infected cases worldwide, six million deaths, and these numbers are still increasing.[2] The substantial number of infected cases demands an effective antiviral medication to fight the virus, but there is a shortage of therapeutic methods for severe cases of COVID-19.

The SARS-CoV-2 genome is one of the largest among RNA viruses. It encodes for 29 proteins: 4 structural, 16 nonstructural, and 9 accessory proteins. The structural proteins include the spike, envelope, membrane, and nucleocapsid proteins. The envelope and membrane proteins are involved in virus packaging and assembly, and the spike protein interacts with human ACE2 receptors on the host cell surface and helps the virus enter the cell during viral infection.[3] All the structural proteins as well as the protease are great drug targets due to their known structures and their vital role in virus replication.

At the time of writing, there are three antiviral pills that are approved by the FDA to treat SARS-CoV-2 under emergency use authorization. Specifically, Pfizer's novel oral antiviral, ritonavir, targets the SARS-CoV-2 main protease (Mpro).[4] Merck's drug, molnupiravir, aims to create multiple deleterious errors in the viral genome and to terminate the replication of the virus.[5] Molnupiravir achieves this by targeting the RNA-dependent RNA polymerase (RdRp) of the virus. Similarly, remdesivir also targets the virus polymerase RdRp.[1] There are many other in-progress antiviral drugs also targeting the SARS-CoV-2 Mpro and RdRp.[1] Even though there is an available antiviral drug for SARS-CoV-2, it is crucial to expand the drug targets of the virus. This will offer an opportunity to fight back against SARS-CoV-2 when key structural viral targets mutate, which can potentially cause the current antivirals to become ineffective.

1.2 Current understanding of SARS-CoV-2 ORF8 protein

The ORF8 protein is a unique accessory protein to SARS-CoV-2. It is one of the proteins encoded by nine open reading frames (ORFs) of SARS-CoV-2. The crystal structure of ORF8 reveals a dimer structure, with an Ig-like (immunoglobulin) fold (**Figure 1**). Additionally, it also reveals the

overall similarity of the core fold to ORF7a.[6] Even though ORF8 is not necessary for protein replication, it plays a role in viral infections.[6] In Young et al., the authors emphasize the correlation of ORF8 and the severity of viral symptoms.[7] The expression of ORF8 is high in infected patients, and it is used as a marker for SARS-CoV-2 diagnostics.[8] Due to the multiple functionalities of ORF8 in the viral infection process, it presents as a promising drug target. Specifically, ORF8 is suspected to weaken immune surveillance by down-regulating the expression of class I major histocompatibility complex (MHC-I).[9] Additionally, ORF8 is reported to form intracellular aggregates in human lung cells and weaken the host's innate immune response.[10] It is also believed that ORF8 contributes to the cytokine storm during infection,[11]and disrupts the host cell response to infection.[12] In Kee et al., they discuss the function of the ORF8 protein as a chromatin disruptor in early infection. This is achieved by ORF8's association with chromatin and its binding to a histone-associated protein which disrupts multiple histone post-translational modifications.[12] Specifically, there is a linkage between the high level of histone H3 and the severity of SARS-CoV-2 symptoms.[13] This mechanism of mimicking critical regions of human histones is also identified in other viral proteins because histone regulation is important for cells to control transcription and respond to viral threats.[12] However, there is still a lack of understanding of the specific working mechanism of ORF8, so this research does help shed light on the mechanism and the general function of ORF8. The development of ORF8's inhibitors will aim to recover the host's innate immune response, relieve the cytokine storm, and decrease the viral infection.

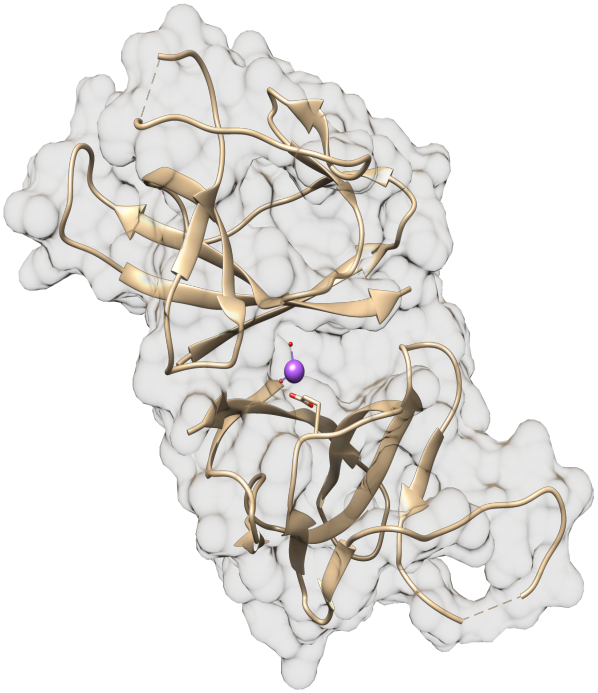


Figure 1: ORF8 (7jtl) visualized in UCSF Chimera

1.3 Current understanding of pharmacophores and virtual screening

Computer-aided drug design (CADD) has strongly contributed to identifying lead compounds and speeding up the drug discovery process. CADD includes several methods such as cheminformatics, bioinformatics, molecular docking, artificial intelligence, structure-based pharmacophore modeling, and virtual screening.[14] Molecular docking tools are used to predict the binding orientation of the ligand with a protein in a three-dimensional structure, which aids in the process of developing pharmacophore models.[15] Pharmacophore models describe the three-dimensional arrangement of generalized features such as intermolecular forces (IMFs): hydrogen bonding, hydrophobic interactions, and ionic interactions.[16] Pharmacophore generation is crucial to the process of finding similar small molecules that share the same molecular characteristics to yield similar IMFs. The purpose of using molecular docking and virtual screening is to reduce the number of compounds to

test *in vitro*. These two methods take advantage of powerful computational technologies which aid in the process of predicting the best matching pair of a ligand to a macromolecule.[14] Virtual screening has proven to be an effective tool to discover lead compounds by searching large databases for chemical structures.[17] Pharmacophore-based virtual screening uses a set of desired features received from molecular docking to search for the lead compounds. In this thesis, both molecular docking and pharmacophore-based virtual screening methods were utilized to search for optimal lead inhibitors of ORF8.

1.4 Limitation of current early drug discovery pipeline

During this research, various cheminformatic tools were utilized to analyze information and visualize results between different software programs which created a significant barrier to the early drug discovery pipeline. Even though the different software programs speed up the process of finding lead compounds, the shifting between software programs/libraries/packages is time consuming and might lead to a loss of information. It is crucial to investigate the automation between different cheminformatics tools such as molecular docking and virtual screening. The generation of a pharmacophore from molecular docking is usually a manual process. Consequently, visualizing the molecular docking results and identifying intermolecular interactions delays the discovery of lead compounds. Many studies work on automated approaches from the starting target PDB compelxes,[18] or ligand poses.[19,20] However, most of these tools still require complex manual steps, difficult input file preparation, and they are not open-source tools, which means that the program's code is not public to modify. Herein a Python script was created to automate the process of identifying IMF interactions from molecular docking results. Additionally, the Python script takes input as the molecular docking's output and produces potential pharmacophores.

1.5 Current understanding of lead compounds

This research focuses on the potential to repurpose novobiocin as an inhibitor of SARS-CoV-2 ORF8. Novobiocin acts as an anti-bacterial agent, but it is also shown that it has antiviral activity *in vitro* and *in vivo* against Zika and vaccinia viruses.[21,22] Novobiocin is also known as an inhibitor of prokaryotic DNA gyrase and eukaryotic type II topoisomerase enzymes. It interacts with histones which disrupt histone-histone associations which cause histones to precipitate from both nucleoplasmin-histone and histone-

DNA complexes.[23]

2 Methods and Results

2.1 Computational studies

2.1.1 Computational methods summary

Computational methods were utilized to identify lead compounds and validate their binding *in silico*. The general outline of the computational method is presented in **Figure 2**. POCASA[24] was used to identify the binding pocket. Molecular docking was carried out using Autodock Vina,[25] and LigandScout 4.4[20] was used to generate structure-based pharmacophores. Virtual screening was performed through ZINCPharmer.[26] Open Babel[27] was used to convert data file types from protein data bank (PDB) to protein data bank, partial charge, and atom type (PDBQT). PDBQT file type is mostly used by Autodock Vina, and it provides more information about an atom's properties compare to the PDB file type. Visualization software such as UCSF Chimera[28] was also utilized. Additionally, an automated pharmacophore generation was written to facilitate this work, and the code is public on GitHub https://github.com/annanguyen99/ThesisDockingAnalysis, which is available for anyone who wants to continue to improve the code.

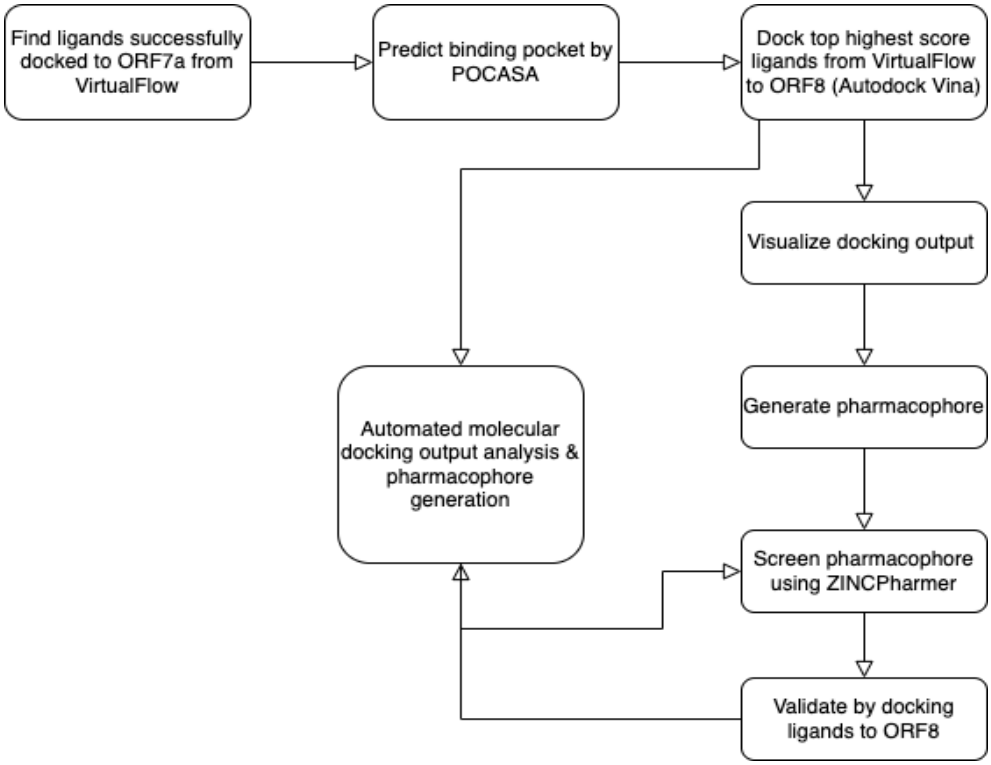


Figure 2: Computational methods used to generate pharmacophores

2.1.2 Structure-based pharmacophore modeling and virtual screening

Since ORF8 has 16% sequence identity with the SARS-CoV ORF7a protein and it is also believed to originate from ORF7a, the best docking score ligands of ORF7a from Virtual Flow were used to dock to ORF8.[29] The binding pocket in ORF8 was identified by POCASA[24] and was utilized to set up docking parameters by AutoDock Vina. The center was set at the center of the binding pocket ($center\_x = 51.512$, $center\_y = 22.11$, $center\_z = 104.839$), and the grid box size of (x = 38, y = 48, z = 34) was enough to capture the whole binding pocket. All the ligands and the receptors had hydrogen at pH = 7.4. UCSF Chimera was used to visual docking results and identify intermolecular force interactions. After performing semi-flexible docking of the best score ORF7a's ligands to ORF8 through AutoDock Vina, three pharmacophores were generated (**Table 1**). In all the pharmacophores, the strongest interactions were conserved to maximize the inter-

molecular forces of ligand to the receptor. However, the weaker interactions such as hydrophobic interaction were different between all pharmacophores to maximize the number of hit compounds.

Features	Pharmacophore 1	Pharmacophore 2	Pharmacophore 3
Hydrophobic with ILE 76A	No	Yes	Yes
Hydrophobic with TYR 73A	No	Yes	No
Hydrogen bonding acceptor with ILE 74A (backbone)	Yes	Yes	Yes
Hydrogen bonding donor with ASP 75A	Yes	Yes	Yes
Hydrogen bonding acceptor with LYS 94A	Yes	Yes	Yes
Hydrogen bonding acceptor with LEU 95A (backbone)	Yes	No	No

Table 1: Chemical features of the three pharmacophores generated from molecular docking.

All of the pharmacophores were used to perform virtual screening against the ZINC drug database through ZINCPharmer.[26] The result of virtual screening is included in supplemental information **(Supplemental Table 3)**. All hit compounds were validated by docking with AutoDock Vina to confirm virtual binding to ORF8. Four compounds from those identified were purchased to test for binding activity. Specifically, the four compounds include kaempferol 7-O-glucoside which is a flavonoid,[30] lercanidipine hydrochloride which is a calcium channel blocker,[31] lohexol which is known as a radioopaque medium,[32] and novobiocin which is an anti-bacterial agent.[22] As seen in **Figure 3**, all four compounds have relatively similar sizes and similar functional groups which include multiple aromatics rings, and hydrogen bond donors and acceptors. The *in silico* binding analysis and ex-

perimental binding activity of novobiocin are the focus of this thesis.

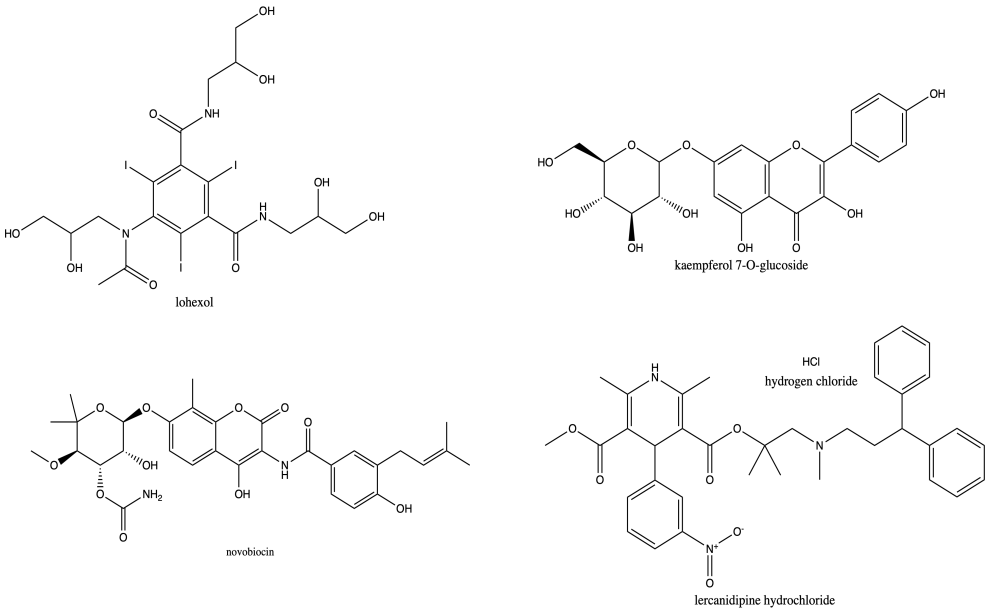


Figure 3: Structure of lohexol, kaempferol 7-O-glucoside, novobiocin, lercanidipine hydrochloride (computationally generated lead compounds)

2.1.3 Automated molecular docking output analysis

To speed up the process of analyzing molecular docking output and pharmacophore generation, an automated molecular docking analysis program was designed. The goal of this program was to speed up the process of identifying ligand-receptor interactions and automate the process of pharmacophore generation. The general flowchart of the program is included in **Figure** 4.

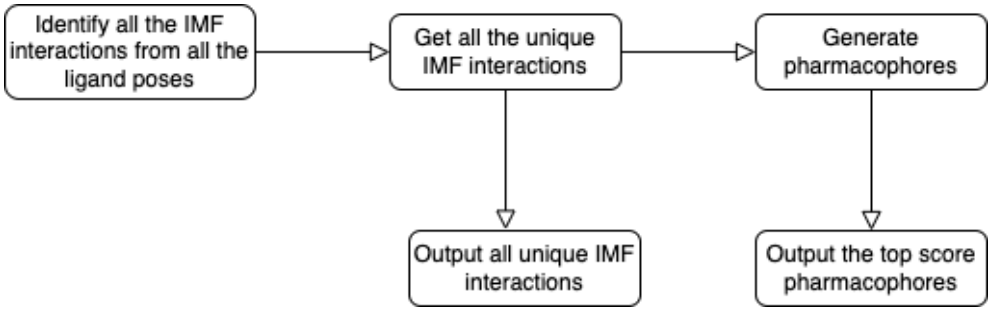


Figure 4: Python automated analysis program

The input of the program is the molecular docking output in the Protein Data Bank, Partial Charge Atom Type (PDBQT) file type, and the receptor is also in the PDBQT file. The input is required to be in the PDBQT file because that is the file type of Autodock Vina's molecular docking output. The interactions between the ligands and the receptor were identified and categorized into four different types of interactions: ion-dipole, hydrogen bonding, dipole-dipole, and hydrophobic interactions. Due to a large number of repetitive IMF interactions from all the docking poses, only unique interactions are kept. Unique interactions are identified by the unique receptor's atom and ligand's atom pair. If a similar receptor atom and ligand pair is encountered, the pair with a smaller interaction distance is kept, and the old pair is removed. To minimize the unique interactions even further, only one hydrophobic interaction per receptor residue is kept. Then, combinations of pharmacophore attributes were randomly generated based on the unique interactions and scored. The combination of pharmacophores is generated by utilizing the permutation method from Python. Reducing the number of unique interactions by eliminate repetition interactions improves the run time of the program.

Generated pharmacophores were scored on the strength of interactions. The scoring of each interaction was purely based on the strength of the interactions as seen in **Algorithm 1**. Specifically, the strongest bond is ion-dipole which has the highest score (4), and the weakest interaction is hydrophobic which has the lowest score (1). The score of a pharmacophore is equal to the total score of all the interactions in the pharmacophore. Then, all the generated pharmacophores were sorted based on their score. The top ten scoring pharmacophores were returned as output. The output of the program was an xlsx (Excel) file. The program returns two outputs: an xlsx

file that includes all the interactions from all the poses of the ligand with the target protein, and another xlsx file that includes the generated pharmacophores. The program was used to generate all the interactions of novobiocin with ORF8 and potential pharmacophores. The program identified seventy-five interactions between novobiocin and ORF8. Additionally, the program produced ten potential pharmacophores with three attributions for each pharmacophore (**Table 2**). There are normally three to five attributes in a pharmacophore. The number of attributes was arbitrary chosen to achieve an ideal number of hit compounds. For this automated program, three attributes were chosen to minimize the number of permutations and the total run time of the program.

Algorithm 1 The scoring function of each attribute of a pharmacophore

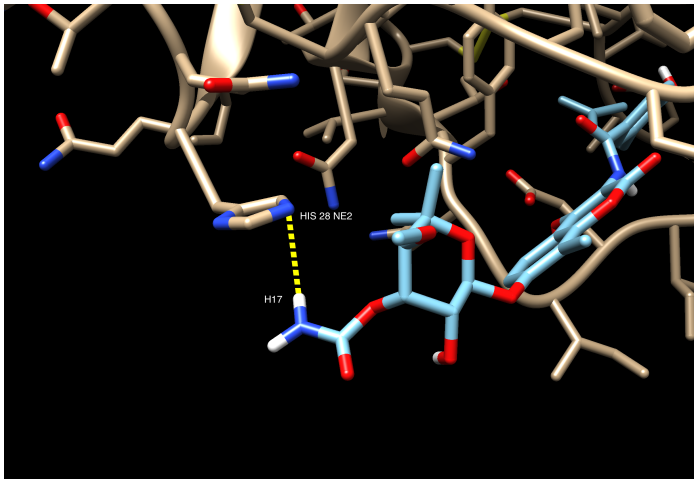
Input: $A_1 \dots A_N$ ▷ The list of pharmacophore interactions
Output: *score* (the total score of the pharmacophore)

```
function scoring_function(A[ ])
    score ← 0
    for k ← 0 to N do
        if A[N] is ion-dipole then
            score ← score + 4
        else if A[N] is hydrogen bonding then
            score ← score + 3
        else if A[N] is dipole-dipole then
            score ← score + 2
        else if A[N] is hydrophobic then
            score ← score + 1
        end if
    end for
    return score
end function
```

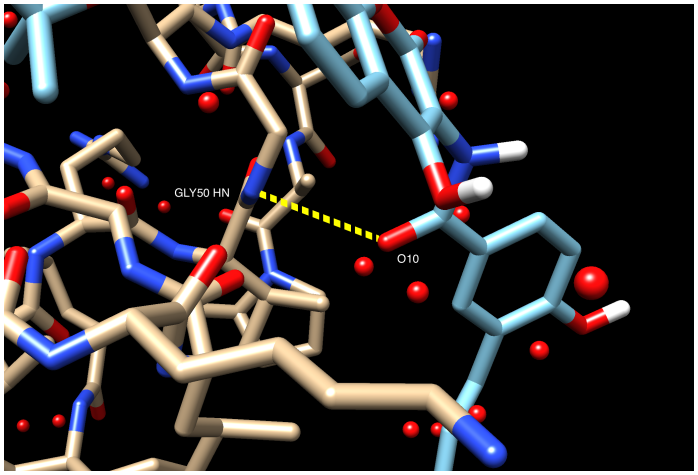
PharmacophoreID	Pharmacophore's score	Receptor side chain	Receptor atom	Ligand atom	Ligand atom serial number	Pose ID	Interaction
1	10	HIS28	NE2	H17	28	4	hydrogen bonding
		GLY50	HN	O10	36	8	hydrogen bonding
		ASP75	OD1	H30	50	9	neg ion - pos par
2	10	HIS28	NE2	H17	28	4	hydrogen bonding
		ARG52	HH1	O9	2	8	hydrogen bonding
		ASP75	OD1	H30	50	9	neg ion - pos par
3	10	HIS28	NE2	H17	28	4	hydrogen bonding
		GLU59	O	H23	33	6	hydrogen bonding
		ASP75	OD1	H30	50	9	neg ion - pos par
4	10	HIS28	NE2	H17	28	4	hydrogen bonding
		CYS61	HN	O7	1	2	hydrogen bonding
		ASP75	OD1	H30	50	9	neg ion - pos par
5	10	HIS28	NE2	H17	28	4	hydrogen bonding
		VAL62	HN	O9	2	2	hydrogen bonding
		ASP75	OD1	H30	50	9	neg ion - pos par
6	10	HIS28	NE2	H17	28	4	hydrogen bonding
		GLU64	HN	O2	22	1	hydrogen bonding
		ASP75	OD1	H30	50	9	neg ion - pos par
7	10	HIS28	NE2	H17	28	4	hydrogen bonding
		SER67	HN3	O6	26	1	hydrogen bonding
		ASP75	OD1	H30	50	9	neg ion - pos par
8	10	HIS28	NE2	H17	28	4	hydrogen bonding
		ILE71	N	H17	28	7	hydrogen bonding
		ASP75	OD1	H30	50	9	neg ion - pos par
9	10	HIS28	NE2	H17	28	4	hydrogen bonding
		TYR73	N	H23	33	7	hydrogen bonding
		ASP75	OD1	H30	50	9	neg ion - pos par
10	10	HIS28	NE2	H17	28	4	hydrogen bonding
		ILE74	N	H23	33	1	hydrogen bonding
		ASP75	OD1	H30	50	9	neg ion - pos par

Table 2: Output of the Python automated program with novobiocin and ORF8. The receptor side chain includes the IUPAC abbreviation of amino acid side chains[33] and the residue ID number. The ligand's atom includes the IUPAC abbreviation of the atom and its ID number.

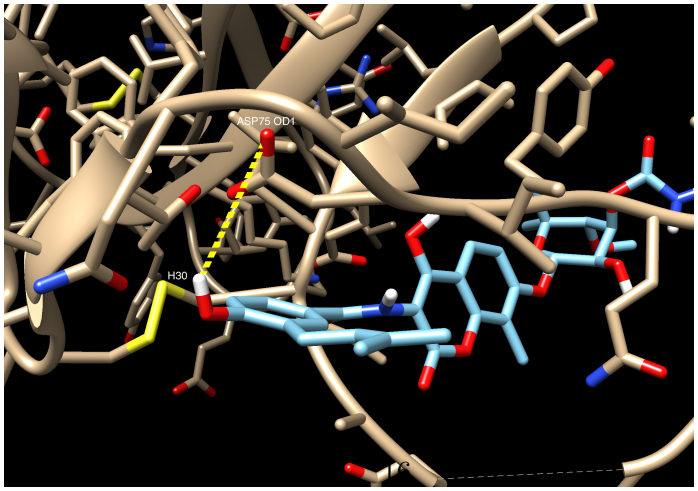
The interactions in the generated pharmacophore can be verified by visualizing each ligand pose in UCSF Chimera. For example, the first pharmacophore includes three interactions: nitrogen (NE2) from histidine (HIS28) hydrogen bonding with hydrogen (H17) of novobiocin (pose 4), nitrogen (NH) from glycine (GLY50) hydrogen bonding with oxygen (O10) (pose 8), and oxygen (OD1) from aspartic acid (APS75) ion-dipole interactions with hydrogen (H30) (**Table 2**). The success in visualizing three interactions of the proposed pharmacophore 1 in UCSF Chimera confirms the ability of the program to identify IMF interactions (**Figure 5**). Additionally, the first pharmacophore is also visualized in ZINCPharmer (**Figure 6**). All the pharmacophores produced from the automated Python program can be used to perform virtual screening with free computational tools such as ZINCPharmer.



(a) Interaction of ligand atom H17 to receptor residue HIS28



(b) Interaction of ligand atom O10 to receptor residue GLY50



(c) Interaction of ligand atom H30 to receptor residue ASP75

Figure 5: Interactions of pharmacophore 1 are visualized in UCSF Chimera (the receptor carbon chain is colored in yellow, and the ligand carbon chain is colored in blue)

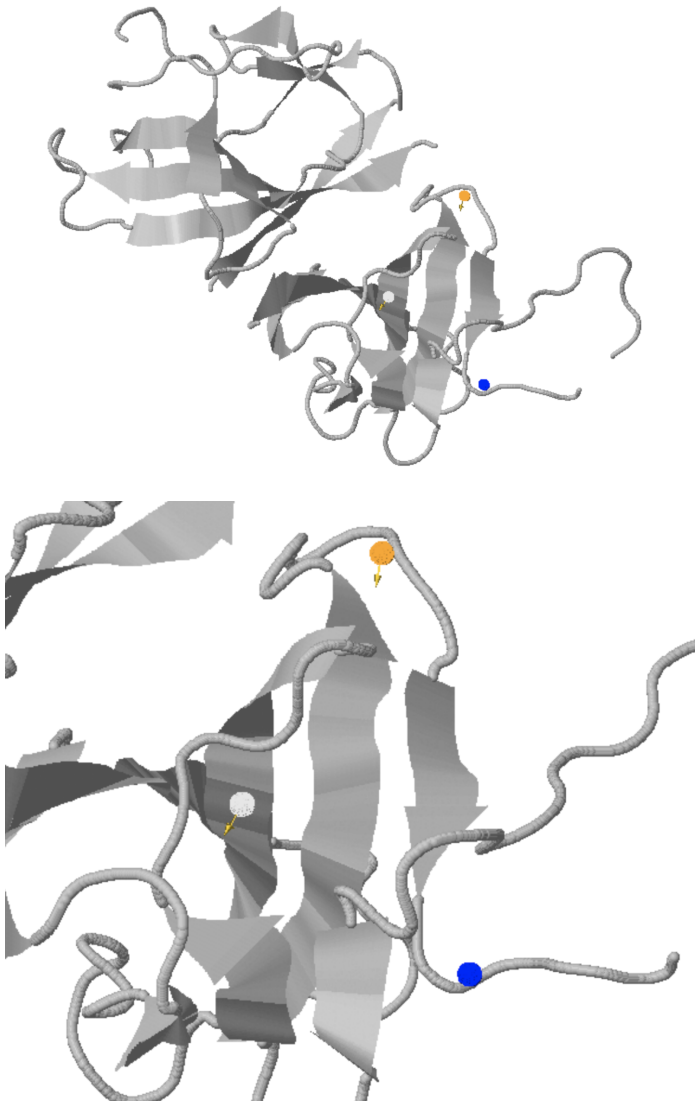


Figure 6: Visualization of the first pharmacophore in ZINCPharmer. ORF8 is colored in gray. Hydrogen donor atom (H17 atom) is colored white, hydrogen acceptor (O10 atom) is colored yellow, and partial positive dipole atom (H30) is colored dark blue)

2.2 *In vitro* studies

The *in silico* binding results were validated by *in vitro* binding experiments.

2.2.1 Protein expression and purification

Protein expression and purification were carried out following a procedure similar to Flower et al.[6] The gene for wild-type SARS-CoV-2 ORF8 in DH5-α cells was obtained from Addgene. The plasmid was isolated and transferred to RosettaTM (DE3) pLysS competent cells. A single colony was selected and expressed overnight at 37^{o}C in 15 mL of Luria broth (LB) containing ampicillin. The next day, the overnight solution was added into 235 mL of similar LB solution. When the OD600 was approximately at 1 and 1mM of IPTG was added. The solution was moved to room temperate and grown overnight at 20^{o}C. The cells were pelleted and resuspended in lysis buffer (50 mM Tris pH 8.0, 2 mM EDTA, 100 mM NaCl, 1 mM DTT, 0.5% Triton-X100) and lysed by sonication. Then, the cells were clarified by centrifugation for 20 minutes at 5^{o}C and at 10000 rpm. The cells were resuspended in lysis buffer, sonicated, and centrifuged again as above. The pellet was resuspended in solubilization buffer (100 mM Tris pH 8.5, 6M guanidine hydrochloride, 10 mM reduced glutathione) and incubated at 20^{o}C with rocking for one hour. The pellet was centrifuged again for 20 minutes at 5^{o}C and at 10000 rpm to remove insoluble material. The supernatant was applied to a nickel-charged IMAC (immobilized metal affinity chromatography) column which was preequilibrated in solubilization buffer. The resin was washed with solubilization buffer with a pump at the rate of 1 mL/minute. The resin and the supernatant were incubated at 4^{o}C with rocking overnight. The bound His-tagged ORF8 was eluted with elution buffer (100 mM Tris pH 8.5, 6M guanidine hydrochloride, 10 mM reduced glutathione, 350 mM imidazole) with a pump at the rate of 1 mL/ minute. The elution solution was added to the refolding buffer dropwise at the rate of 1 mL/minute for a two-hour period with stirring. Then, the refolding solution was incubated with stirring at 4^{o}C overnight.

2.2.2 Protein size, purity, aggregation state

SDS-PAGE

The refolded protein was concentrated and run on SDS-PAGE. ORF8 protein molecular weight is 12,212 Da, cytochrome c 's molecular weight is

12,327 Da (cytochrome c was used as protein markers) **Figure 7**.

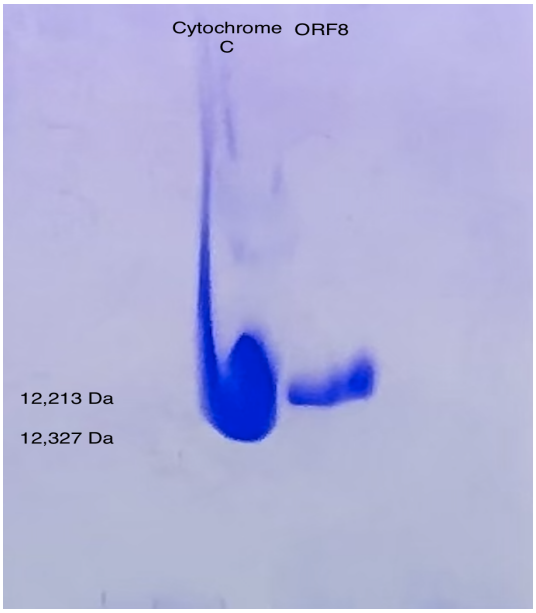


Figure 7: SDS-PAGE of ORF8 (12,212 Da) and marker cytochrome c (12,327 Da)

Size exclusion chromatography

Size exclusion chromatography was performed to confirm that the protein was not aggregated. The size exclusion chromatography was carried out with 5 ml of Superdex 75 resin. The flow rate was 0.3 mL/ minute. Carbonic anhydrase was used as a molecular weight marker (carbonic anhydrase's molecular weight is 29,200 Da). The molecular weight of a monomer of ORF8 protein is 12,212 Da, so its aggregation would be 24,424 Da which is around the molecular weight of carbonic anhydrase. Fractions of 200 μL of carbonic anhydrase were collected in a 96-well plate and absorbance at 280 nm was collected using Molecular Devices Spectramax M2. Similarly, fractions of 200 μL of ORF8 were collected in a 96-well plate and excited at 295 nm and the fluorescence emission at 332 nm was collected using the same instrument.

The absorbance for carbonic anhydrase showed up at 600 μL, and the absorbance for ORF8 showed up at 1,200 μL. This indicates that the reten-

tion time of ORF8 was longer than the retention time of carbonic anhydrase (**Figure 8**). Since the size exclusion chromatography separates molecule by size, the data indicates that ORF8 has a smaller size than carbonic anhydrase. This confirms the prediction that ORF8 is a monomer.

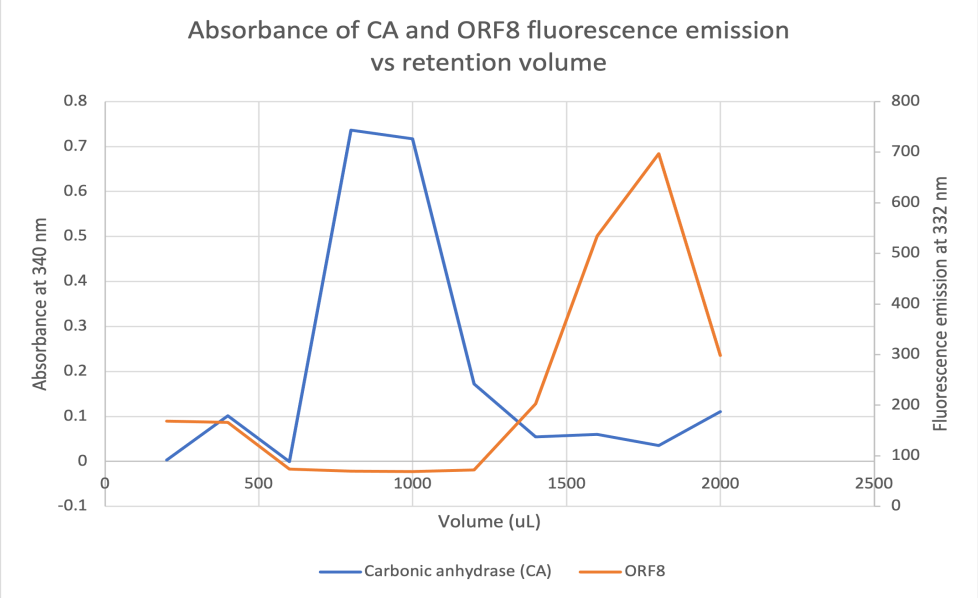


Figure 8: Size exclusion chromatography of ORF8. Absorbance of carbonic anhydrase excited at 280 nm versus its retention volume from size exclusion chromatography, and the fluorescence emission of ORF8 at 332 nm which was excited at 295 nm versus its retention volume from size exclusion chromatography

2.2.3 Binding activity assay for Novobiocin

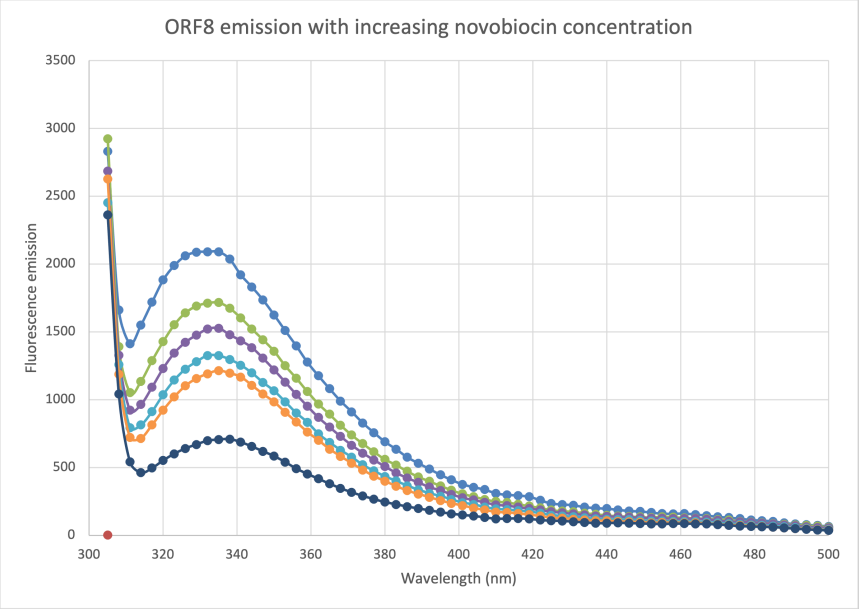
Ligand preparation. Novobiocin was purchased from Sigma Aldrich. It was dissolved in H_2O (pH = 7.0). Different concentration of novobiocin were prepared (1.29 mM, 0.796 mM, 0.573 mM, 0.369 mM, 0.215 mM, 0.117 mM).

***In vitro* binding assay.** In a 96-well plate, 12 μL of the ligand solution was added to 120 μL of ORF8 (1.17×10^{-5} M in its refolding buffer). The solution was excited at 295 nm and fluorescence absorbance was collected from 305 nm to 500 nm using Molecular Devices Spectramax M2.

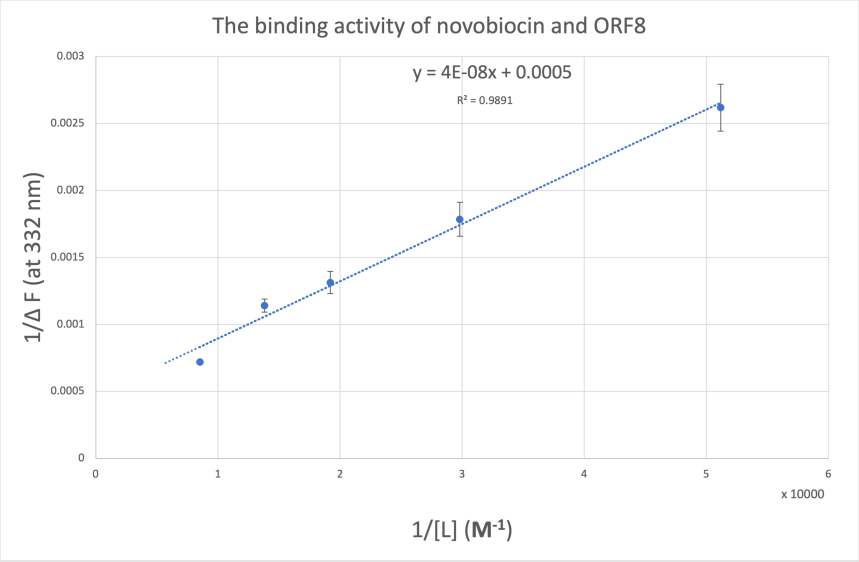
The fluorescence emission of tryptophan was recorded.

Dissociation constant (Kd) determination. The dissociation constant (Kd) was determined through plotting a double reciprocal plot derived from the Benesi-Hildebrand relationship.[34] F is the fluorescence intensity at some particular ligand concentration. The value ΔF is (F – F_o) where F_o is the fluorescence intensity in the absence of ligand. The equation $\frac{1}{\Delta F} = \frac{1}{(F_\infty - F_o) * K_d * [L]} + \frac{1}{F_\infty - F_o}$ was used to determine K_d. The intercept on the x-axis is $\frac{-1}{K_d}$.

The fluorescence emission of the one tryptophan amino acid in ORF8 protein decreased when the concentration of novobiocin increased (**Figure 9**). There is only one tryptophan amino acid in ORF8 which is located in the binding pocket, so the decrease in fluorescence emission could indicate that either the ligand bound directly to the tryptophan or that there was a conformational change by binding. Additionally, the λ_{max} of emission (332 nm) slightly shifted to longer wavelength and lower intensity as the concentration of the novobiocin increases. Both of the observations could indicate that there is an increase in polarity around the ORF8 tryptophan. The dissociation constant (Kd) of novobiocin was calculated to be 91 μM $\pm$ 26 μM based on the double reciprocal plot (**Figure 9**).



(a) ORF8 fluorescence with increasing novobiocin concentration (concentration of novobiocin from top to bottom: no ligand, 1.95 x 10e-5 M, 3.35 x 10e-5 M, 5.21 x 10e-5 M, 7.23 x 10e-5 M, 1.2 x 10e-4 M). The concentration of ORF8 was 1.17 x 10-5 M



(b) The double reciprocal plot of 1/ F and 1/[L]

Figure 9: ORF8 fluorescence and Kd determination

3 Discussion and Conclusions

In the process of COVID-19 drug discovery, all the structural proteins have gained attention as novel antiviral or vaccine targets.[29] This thesis focuses on an unpopular SARS-CoV-2 protein, ORF8. The research had two main goals which were to develop an automated program to generate a pharmacophore from molecular docking results and to utilize the combination of computational and *in vitro* techniques to identify potential SARS-CoV-2 ORF8 inhibitors.

Firstly, the automated program offers the ability to automate the process of identifying IMF interactions from Autodock Vina output. This feature has not been included in any popular docking visualization tools such as UCSF Chimera or LigandScout. This new program will provide chemists with a tool to speed up the process of analyzing docking visualization and the ability to evaluate a large number of molecular docking outputs. In this thesis, the program was used to analyze the molecular docking output of ORF8's potential inhibitor, novobiocin, and generate ten pharmacophores from the molecular docking outputs. In future studies, this program could be expanded to generate pharmacophores from multiple molecular docking output. The program could also incorporate a multi-threaded implementation to score the pharmacophore, and this could improve the run time of the program. Additionally, this feature could potentially be incorporated into molecular docking visualization tools to aid in the process of generating pharmacophores.

Secondly, this thesis presented a potential inhibitor for the SARS-CoV-2 ORF8 protein, novobiocin. Novobiocin was one of the hit compounds from the computational studies, and its binding activity was validated by an *in vitro* binding assay. Novobiocin could act as a great start to perform lead optimization to improve its binding to ORF8 protein. The dissociation constant, 91 mM $\pm$ 26 mM, only shows there is interaction between novobiocin and ORF8, but it does not indicate tight binding. Modification of the functional groups of novobiocin to improve the IMF interactions could be the potential next step to strengthen the inhibition ability of novobiocin. The binding of novobiocin to ORF8 *in vitro* is interesting considering its mechanism of binding to histone or histone-like molecules. A recent study shows that SARS-CoV-2 ORF8 functions as a histone mimic of the ARKS motif in histone 3.[12] Histone proteins are essential for the cell to control gene expression and they achieve that by wrapping DNA into complex structures which

are modified by post-translation modifications. By mimicking histones, the virus may disrupt the host cell's regulation of gene expression and weaken the cell response to infection.[12] Aiming to repurpose novobiocin to target ORF8 could potentially block its binding to histone-associated proteins and improve host cell virus regulation. In future studies, *in vivo* test could be performed to test the ability of novobiocin to cross cell membrane as well as its cytotoxicity.

4 Supplemental information

Pharmacophore 1	**Pharmacophore 2**	**Pharmacophore 3**
ZINC40368514	ZINC14879999	ZINC03830945
ZINC04096097		ZINC03830407
ZINC05742780		ZINC03830396
ZINC03977952		ZINC03830426
ZINC04262249		ZINC03977816
ZINC38664850		ZINC03830489
ZINC38664850		ZINC03830633
ZINC38664850		ZINC03830429
ZINC36046230		ZINC03830430
ZINC04348951		ZINC14879999
ZINC12494320		ZINC03831159
		ZINC19685788
		ZINC03830946
		ZINC03831506
		ZINC0383043

Table 3: The virtual screening hits and the hit compounds ZINC ID

References

(1) Cully, M. *Nature Reviews Drug Discovery* **2021**, *21*, Bandiera_abtest: a Cg_type: News Number: 1 Publisher: Nature Publishing Group, 3–5.

(2) WHO Coronavirus (COVID-19) Dashboard https://covid19.who.int (accessed 02/22/2022).

(3) Saxena, A. *Journal of Biosciences* **2020**, *45*, 87.

(4) Owen, D. R. et al. *Science* **2021**, Publisher: American Association for the Advancement of Science, DOI: 10.1126/science.abl4784.

(5) Jayk Bernal, A. et al. *New England Journal of Medicine* **2021**, *0*, Publisher: Massachusetts Medical Society, null.

(6) Thomas G. Flower; Cosmo Z. Buffalo; Richard M. Hooy; Marc Allaire; Xuefeng Ren; James H. Hurley Structure of SARS-CoV-2 ORF8, a rapidly evolving immune evasion protein — PNAS, PNAS https://www.pnas.org/content/118/2/e2021785118 (accessed 07/30/2021).

(7) Young, B. E. et al. *The Lancet* **2020**, *396*, 603–611.

(8) Hachim, A.; Kavian, N.; Cohen, C. A.; Chin, A. W. H.; Chu, D. K. W.; Mok, C. K. P.; Tsang, O. T. Y.; Yeung, Y. C.; Perera, R. A. P. M.; Poon, L. L. M.; Peiris, J. S. M.; Valkenburg, S. A. *Nature Immunology* **2020**, *21*, Bandiera_abtest: a Cg_type: Nature Research Journals Number: 10 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Antibodies;Viral infection Subject_term_id: antibodies;viral-infection, 1293–1301.

(9) Zhang, Y. et al. *Proceedings of the National Academy of Sciences* **2021**, *118*, Publisher: National Academy of Sciences Section: Biological Sciences, DOI: 10.1073/pnas.2024202118.

(10) Geng, H.; Subramanian, S.; Wu, L.; Bu, H.-F.; Wang, X.; Du, C.; De Plaen, I. G.; Tan, X.-D. *Frontiers in Immunology* **2021**, *12*, 679482.

(11) Lin, X. et al. *iScience* **2021**, *24*, 102293.

(12) Kee, J. et al. *SARS-CoV-2 ORF8 encoded protein contains a histone mimic, disrupts chromatin regulation, and enhances replication*, Section: New Results Type: article; bioRxiv, 2021, p 2021.11.10.468057.

(13) Huckriede, J.; de Vries, F.; Hultström, M.; Wichapong, K.; Reutelingsperger, C.; Lipcsey, M.; Garcia de Frutos, P.; Frithiof, R.; Nicolaes, G. A. F. *Frontiers in Cellular and Infection Microbiology* **2021**, *11*.

(14) Frontiers — Bridging Molecular Docking to Molecular Dynamics in Exploring Ligand-Protein Recognition Process: An Overview — Pharmacology https://www.frontiersin.org/articles/10.3389/fphar.2018.00923/full (accessed 09/09/2021).

(15) Meng, X.-Y.; Zhang, H.-X.; Mezei, M.; Cui, M. *Current computer-aided drug design* **2011**, *7*, 146–157.

(16) Jiang, S.; Feher, M.; Williams, C.; Cole, B.; Shaw, D. E. *Journal of Chemical Information and Modeling* **2020**, *60*, Publisher: American Chemical Society, 4326–4338.

(17) Chen, Z.; Li, H.-l.; Zhang, Q.-j.; Bao, X.-g.; Yu, K.-q.; Luo, X.-m.; Zhu, W.-l.; Jiang, H.-l. Pharmacophore-based virtual screening versus docking-based virtual screening: a benchmark comparison against eight targets — Acta Pharmacologica Sinica, Acta Pharmacologica https://www.nature.com/articles/aps2009159 (accessed 09/09/2021).

(18) Ortuso, F.; Langer, T.; Alcaro, S. *Bioinformatics* **2006**, *22*, 1449–1455.

(19) Hu, B.; Lill, M. A. *Journal of Chemical Information and Modeling* **2013**, *53*, Publisher: American Chemical Society, 1179–1190.

(20) Wolber, G.; Langer, T. *Journal of Chemical Information and Modeling* **2005**, *45*, Publisher: American Chemical Society, 160–169.

(21) Abrams, R. P. M. et al. *Proceedings of the National Academy of Sciences* **2020**, *117*, Publisher: National Academy of Sciences Section: Biological Sciences, 31365–31375.

(22) Sekiguchi, J.; Shuman, S. *Virology* **1997**, *235*, 129–137.

(23) Sealy, L.; Cotten, M.; Chalkley, R. *The EMBO Journal* **1986**, *5*, 3305–3311.

(24) Yu, J.; Zhou, Y.; Tanaka, I.; Yao, M. *Bioinformatics (Oxford, England)* **2010**, *26*, 46–52.

(25) Trott, O.; Olson, A. J. *Journal of computational chemistry* **2010**, *31*, 455–461.

(26) Koes, D. R.; Camacho, C. J. *Nucleic Acids Research* **2012**, *40*, W409–W414.

(27) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. *Journal of Cheminformatics* **2011**, *3*, 33.

(28) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Meng, E. C.; Couch, G. S.; Croll, T. I.; Morris, J. H.; Ferrin, T. E. *Protein Science : A Publication of the Protein Society* **2021**, *30*, 70–82.

(29) Zinzula, L. *Biochemical and Biophysical Research Communications* **2021**, *538*, 116–124.

(30) Lu, C. H.; Li, Y. Y.; Li, L. J.; Liang, L. Y.; Shen, Y. M. *Drug Discoveries & Therapeutics* **2012**, *6*, 194–197.

(31) Chen, Y.; Yuan, J.; Jiang, G.; Zhu, J.; Zou, Y.; Lv, Q. *Molecular Medicine Reports* **2017**, *16*, 4545–4552.

(32) PubChem Iohexol https://pubchem.ncbi.nlm.nih.gov/compound/3730 (accessed 03/07/2022).

(33) *European Journal of Biochemistry* **1984**, *138*, 9–37.

(34) Raabe, R.; Gentile, L. *Journal of Biophysics* **2008**, *2008*, 474205.