Virtual Screening For Potential SARS-CoV-2 Main Protease Inhibitors Through Molecular Docking In Schrodinger

Tyrone Xue [1], Khaled M. Elokely [2, 3]

[1] Hamilton College, Clinton, NY
[2] Institute for Computational Molecular Science, Temple University, Philadelphia, PA
[3] Department of Chemistry, Temple University, Philadelphia, PA

ICMS Institute for Computational Molecular Science
HPC High-Performance Computing
TEMPLE UNIVERSITY®
U.S. National Science Foundation

ABSTRACT

This study aims to discover new leads will be able to successfully inhibit the main protease (M^{pro}) of SARS-CoV-2. We obtained 151 3D structures of the M^{pro} which were downloaded from the RCSB Protein Data Bank (PDB). 150 of which in complex with ligands that we categorized them by the ligands' interactions with M^{pro} (covalent/noncovalent) and the location of the binding sites (allosteric/orthosteric). Afterwards, both the protein (M^{pro}) and the ligands were prepared in the Schrodinger software. Preparation involved removing soaps/solvents, refining the protein for missing loops, replacing hydrogens, assigning optimized hydrogen bonding, generating the correct protonation states and tautomers, capping the protein termini, removing waters having less than 2 hydrogen bonds with non-water residues, etc. This preparation ensures that the 3D structures do not have experimental remnants that resulted from the crystallography which could affect the ligand-protein interaction. We docked the ligands to their respective receptors for the redocking process to ensure that the interactions shown by the PDB were reasonable. Comparing the ligand interaction diagrams between the original and redocked ligands showed that most of the PDBs were reasonable. After redocking, we virtually screened ligands obtained from a ligand database built from ChemDiv libraries in which we obtain potential ligands that inhibits M^{pro}.

BACKGROUND

- SARS-CoV-2 (otherwise known as COVID 19) has non-structural proteins (NSPs) which are cleaved from polyproteins by two proteases.
- The main protease (M^{pro}) cleaves the pp1a and pp1ab polyproteins to release NSP4 to NSP16.[1]
- Released NSPs are mainly responsible for crucial functions for SARS-CoV-2 such as viral replication.[1]
- M^{pro} has a catalytic dyad formed by a Cysteine (Cys145) and nearby histidine (His41) which catalyzes covalent carbon to sulfur bonds between Cys145 and certain ligands.[1]
- Structure-based drug design utilizes the 3D structure of a target protein in order to find ligands that binds to the protein.
- Virtual screening (VS) utilizes computation in order to screen for potential ligands that has a high probability of binding to the target.
- VS can be done through molecular docking which mimics the binding interactions a ligand would have to a protein.

DATABASES

RCSB PDB

- RCSB PDB was used to obtain our M^{pro} receptors
- 3D structures of M^{pro} (PDBs) were queried using a sequence similarity (E = 0.00001, 95% identity cutoff) to 7GEF and grouped by PDB Deposit ID
- 151 PDBs were categorized based on: ligand presence, number of chains, ligand interaction (non/covalent), and binding site (orthosteric/allosteric)
- Only PDBs with a resolution of less than 2 Å were selected

Fig. 1. Our base PDB for M^{pro} 7GEF found through basic text search on the RCSB.

ChemDiv

- ChemDiv was used as our ligand database that total to 69587 compounds
- The ligand database is built off of four libraries: CORONAVIRUS, Antiviral, PLpro, Protease

DOCKING

PROCEDURE

Preparation

Protein Preparation

- Remove waters showing less than 2 hydrogen bonds
- Fill in missing hydrogen atoms
- Assigning correct protonation states and tautomerization
- Cap protein termini with ACE and NMA residues
- Fill in missing loops
- Convert selenomethionine to methionine
- Optimize hydrogen bonds
- Set pH levels to 7.4 pH
- Minimize the structure

Ligand Preparation

- Ligands are prepared in a similar fashion such as assigning protonation state and tautomeric form
- Correct resonance structures

Docking

- Grids of the binding site are generated from PDBs
- Ligands are docked using glide in Schrödinger
- Poses are generated to find the best conformation
- We used partial flexibility to balance between speed and docking accuracy

Scoring

- The interactions of poses are quantified based on a docking score using a force-field scoring function
- The lower the docking score the better
- We used standard precision (SP) terms

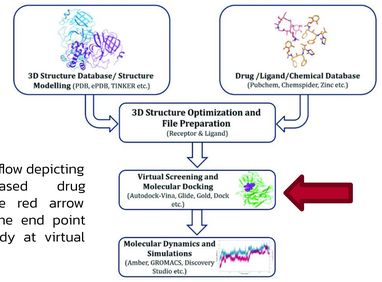
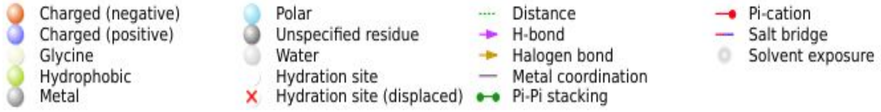


Fig. 2. Workflow depicting structure-based drug design. The red arrow indicates the end point of this study at virtual screening. [2]

2D Interaction Diagram Legend



REDOCKING

All of the PDBs that had a ligand were flexibly docked to itself to determine that the PDBs ligand-receptor interactions were reasonable. Differences between original and docked interactions of the ligand does not indicate inaccuracy of the 3D structure.

PDB ID	Prime MMGBSA Ligand Efficiency	Docking Score	MMGBSA dG Bind	Lig Strain Energy	MMGBSA dG Bind(NS)	Rec Strain Energy
5RHC	-4.685	-4.857	-32.798	0.696	-26.893	-6.601
5RFZ	-3.875	-5.744	-42.624	0.329	-39.485	-3.469
7GRM	-3.872	-7.650	-58.085	0.866	-51.979	-6.971
7GEL	-3.782	-7.598	-75.636	4.606	-69.090	-11.152
7GB5	-3.738	-6.190	-63.546	1.095	-69.878	5.238

Fig. 3. (Above) The top 5 PDBs sorted by Prime MMGBSA ligand efficiency is listed with its energies and docking score.

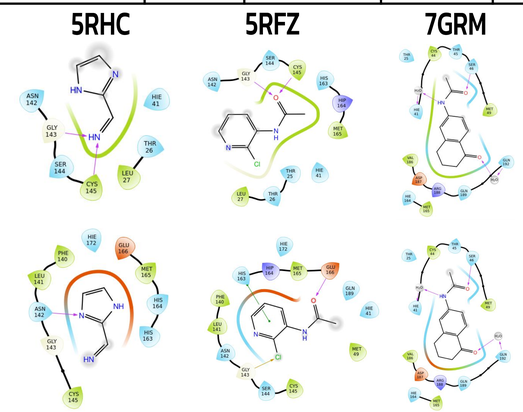


Fig. 4. (Left) 2D interaction diagrams shows the ligand to receptor interactions of the top 3 PDBs. It shows both the original and docked ligands. Refer to the legend.

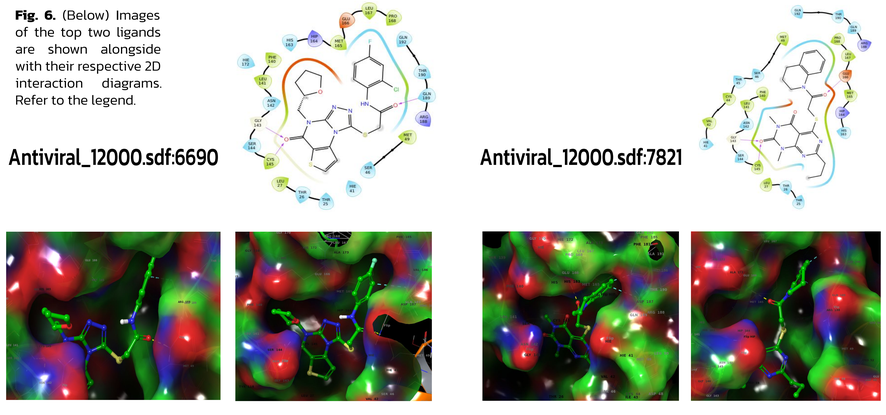
VIRTUAL SCREENING

Using the Virtual Screening Workspace tool, we ensemble-docked the grids from the top 3 PDBs from our redocking: 5RHC, 5RFZ, and 7GRM with our ligand database. Prior to docking, the ligands are filtered under Lipinski's rule and its reactivity (to unintended targets).

Ligand	ChemDiv Library	Docking Score	Receptor PDB
Antiviral_12000.sdf:6690	Antiviral	-9.801	5RFZ
Antiviral_12000.sdf:7821	Antiviral	-9.396	5RFZ
Antiviral_12000.sdf:2018	Antiviral	-9.341	5RFZ
Antiviral_12000.sdf:7820	Antiviral	-9.265	5RFZ
E017-0689	Protease	-9.263	5RFZ

Fig. 5. The top 5 ligands from the virtual screening were sorted by docking score and are reported in this chart. All of which docked with the orthosteric grid from 5RFZ.

Fig. 6. (Below) Images of the top two ligands are shown alongside with their respective 2D interaction diagrams. Refer to the legend.



FUTURE DIRECTIONS

- As there was limited time, the results would be more accurate if our ligand database was ensemble-docked with the grids of all of the PDBs.
- For the same reason, our selected ChemDiv libraries were small and tailored towards SARS-CoV-2 in some way. For drug discovery, it would be best to use larger compound libraries.

CONCLUSIONS

- Our molecular docking shows that compounds 6690 and 7821 from the ChemDiv Antiviral library are strong candidates for ligands to SARS-CoV-2 M^{pro}
- Both compounds binds to the S1 and S1' subpockets of the orthosteric binding site with hydrogen bonds to Cys145 and Gly143

Aknowledgements / References

We acknowledge the NSF for funding this REU and making this research possible. We acknowledge Aleksandra Khotimchenko for providing the ChemDiv libraries, and Ahmed M. Abdelwaly for helping in the initial setup.

1. Li, G., Hilgenfeld, R., Whitley, R. et al. Therapeutic strategies for COVID-19: progress and lessons learned. Nat Rev Drug Discov 22, 449–475 (2023). https://doi.org/10.1038/s41573-023-00672-y
2. Sharma PP, Bansal M, Sethi A, Poonam, Pena L, Goel VK, Grishina M, Chaturvedi S, Kumar D, Rathi B. Computational methods directed towards drug repurposing for COVID-19: advantages and limitations. RSC Adv. 2021 Nov 10;11(57):36181–36198. doi: 10.1039/d1ra05320e. PMID: 35492747; PMCID: PMC9043418.