

Potential therapeutic agents for COVID-19 based on the analysis of protease and RNA polymerase docking

Yu-Chuan Chang^{1,2,†}, Yi-An Tung^{1,3,†}, Ko-Han Lee^{1,†}, Ting-Fu Chen^{1,†}, Yu-Chun Hsiao¹, Hung-Ching Chang¹, Tsung-Ting Hsieh¹, Chan-Hung Su¹, Su-Shia Wang¹, Jheng-Ying Yu¹, Shang-shung Shih¹, Yu-Hsiang Lin¹, Yin-Hung Lin¹, Yi-Chin Ethan Tu¹, Chun-Wei Tung^{1,4,*}, Chien-Yu Chen^{1,5,*}

¹Taiwan AI Labs, Taipei 10351, Taiwan

²Graduate Institute of Biomedical Electronics and Bioinformatics, National Taiwan University, Taipei 10617, Taiwan

³Genome and Systems biology degree program, Academia Sinica and National Taiwan University, Taipei 10617, Taiwan

⁴Graduate Institute of Data Science, College of Management, Taipei Medical University, Taipei 106, Taiwan

⁵Department of Biomechatronics Engineering, National Taiwan University, Taipei 10617, Taiwan

[†]These authors contributed equally to this work

*Correspondence: cwtung@tmu.edu.tw, chienyuchen@ntu.edu.tw

Abstract

The outbreak of novel coronavirus (COVID-19) infections occurring in 2019 is in dire need of finding potential therapeutic agents. In this study, we used molecular docking strategies to repurpose HIV protease inhibitors and nucleotide analogues for COVID-19. The evaluation was made on docking scores calculated by AutoDock Vina and RosettaCommons. Preliminary results suggested that Indinavir and Remdesivir have the best docking scores and the comparison of the docking sites of these two drugs shows a near perfect dock in the overlap region of the protein pocket. However, the active sites inferred from the proteins of SARS coronavirus are not compatible with the docking site of COVID-19, which may give rise to concern in the efficacy of drugs.

Introduction

In the concluding weeks of 2019, an outbreak of novel coronavirus (COVID-19) infections occurred in Wuhan, China. As of February 14, 2020, more than 60,000 cases and 1,000 deaths have been reported. With no proven antiviral agent available, medical professionals have resorted to supportive care to contain the infection. However, current research now suggests that selecting drugs with the appropriate viral restraining mechanisms can yield promising results. At the Rajavithi Hospital in Thailand, the infectious disease team used a combination of oseltamivir (anti-influenza agent) and lopinavir/ritonavir (anti-HIV agent) to successfully improve patients with severe conditions. Lopinavir and ritonavir are both HIV protease inhibitors that suppress the cleaving of a polyprotein into multiple functional proteins¹. Likewise, various clinical trials are also now being undergone on nucleotide analogue drugs such as Remdesivir, an antiviral drug proven to be effective against a wide range of RNA viruses in vitro². This study aims to determine whether the protease of COVID-19 can be a target protein of lopinavir and ritonavir, and to attempt to identify other HIV protease inhibitors with even stronger binding affinities. Additionally, we also tested a set of RNA virus agents for potential binding with RNA-dependent RNA polymerase (RdRp) of COVID-19 in this study.

Methods

In this study, we chose 3-chymotrypsin-like protease (3CL-protease), the main protease used to cleave polyproteins into replication-related proteins, and RdRp, the main protein for RNA replication, as the target receptors. The 3CL-protease structure (PDB ID: 6LU7)^{3,4} was obtained from the RCSB Protein Data Bank³, which was recently released on February 5th, 2020⁴. The structure of the COVID-19 polymerase is currently not available, so we used homology modeling on the sequence of the patient, BetaCoV/Taiwan/2/2020|EPI_ISL_406031, from the Global Initiative on Sharing All Influenza Data (GISAID)^{5,6} to build an approximate structure based on the template of the SARS coronavirus polymerase (PDB ID: 6NUR)⁷ using Swiss-model^{8,9,10,11,12}. The Root-mean-square deviation (RMSD) between 793 aligned atom pairs is 0.073 angstroms, which indicates that these two proteins shared very similar structure (see Supplement Figure 1).

Based on a literature survey^{2,13,14,15,16}, we tested ten FDA-approved drugs of HIV protease inhibitor for potential treatment, including Amprenavir, Atazanavir, Darunavir, Fosamprenavir, Indinavir, Lopinavir, Nelfinavir, Ritonavir, Saquinavir, and Tipranavir¹³. As for nucleotide analogue, we tested four related drugs, Favipiravir¹⁴, Galidesivir¹⁵, Remdesivir², and Ribavirin¹⁶. We downloaded the 3D structure of each drug from the PubChem database in Structure-data file (SDF) format.

To simulate binding affinity between protein and ligands, two docking tools, AutoDock Vina (version 1.1.2)¹⁷ and RosettaCommons (version 3.11)^{18–20} were used. Since AutoDock Vina only takes Protein Data Bank, Partial Charge (Q), & Atom Type (T) (PDBQT) format as input, we used OpenBabel (version 3.0.0)²¹ to convert SDF to PDBQT. The entire protein is taken as the search space. We also used RosettaCommons ligand docking utilities to dock the selected drugs to their corresponding protease and polymerase under default settings. This process was repeated 100 times for each ligand, and the final mean affinity score was taken. From this, we built a heat map for the residues of binding sites to represent the binding frequency of each ligand.

Being a valid drug, we anticipate the ligand to dock at a protein pocket or active site. Therefore, we used the CASTp²² tool to predict potential pockets of our target proteins, and confirmed whether the highest frequency binding sites of the heat map is located in a protein pocket. Additionally, due to the high similarity between the SARS and COVID-19, the catalytic sites of SARS protease and polymerase obtained from the Uniprot database²³ were used to label the corresponding region of COVID-19. Finally, we superimposed the heat map, pocket sites, and active regions to visualize the binding poses.

Results

In Table 1, we listed the results of 10 ligands docking with the 3CL-protease. These scores, which are the original raw outputs from both docking tools, represent the relative binding affinity. For AutoDock Vina, Indinavir has the lowest docking score, even outperforming the results of Lopinavir and Ritonavir, the two drugs currently in use in Thailand. As for the results of RosettaCommons, all ten drugs displayed similar scores, though Amprenavir, Atazanavir, and Darunavir have slightly better performances. We therefore chose Indinavir for further investigation. Visualization of Indinavir docking is shown on Figure 1. Figure 1a shows the overlap

of the ligand heat map and the protein pocket. The purple color highlights that the binding sites are mostly located at the protein pocket. However, one thing to note is that the active sites of 3CL-protease do not overlap with the binding site (Figure 1b), which may be a concern in the case of inhibition. Guangdi Li and Erik De Clercq also noted that the C2-symmetric site, which is the optimized fitting site of HIV protease inhibitors, was absent in 3CL-protease²⁴. Lastly, HIV protease belongs to the aspartic protease family, while the 3CL-like protease is of the cysteine protease family²⁴. These inconsistencies may lead to unwanted noises in selecting potential therapeutic agents.

The docking scores of the four ligands to RdRp using AutoDock Vina and RosettaCommons are listed on Table 2. As shown in Table 2, Remdesivir has the best docking performance by both AutoDock Vina and RosettaCommons. This result suggests that Remdesivir has the most stable docked structure compared to that of the other drug candidates. We further examined the docking results of the given structural complex between Remdesivir and the RdRp through the heat map. The heat map of the contact frequencies between the Remdesivir and RdRp docking results is illustrated in Figure 2a. The docking orientation with the docking consensus lies in the binding pocket of the RdRp. This result agrees with the previous study published in Cell Research²⁵, which suggests that Remdesivir is highly effective in controlling COVID-19 infection *in vitro*. Although our docking results agreed with *in vitro* studies, we discovered that the docking event is not located around the known catalytic domain of the RdRp (Figure 2b). This indicates that the association between these docking results and the effectiveness of treating COVID-19 still needs further examination.

In conclusion, our findings revealed that Indinavir and Remdesivir possess docking sites that strongly overlap with the protein pockets, and could be potential therapeutic agents. Because both drugs have been used in clinical practices with limited toxicity, we recommend that they should be taken into consideration while treating for COVID-19.

Table 1. The docking score of candidate ligands to 3CL-protease

PubChem CID	Name	AutoDock Vina	Rosetta
5362440	Indinavir	-10.424	-1,542.25
92727	Lopinavir	-9.434	-1,549.45
148192	Atazanavir	-9.186	-1,575.17
441243	Saquinavir	-9.174	-1,559.60
392622	Ritonavir	-8.927	-1,563.98
64143	Nelfinavir	-8.702	-1,539.37
213039	Darunavir	-8.505	-1,574.33
54682461	Tipranavir	-8.415	-1,541.84
65016	Amprenavir	-8.023	-1,575.98
131536	Fosamprenavir	-7.637	-1,549.66

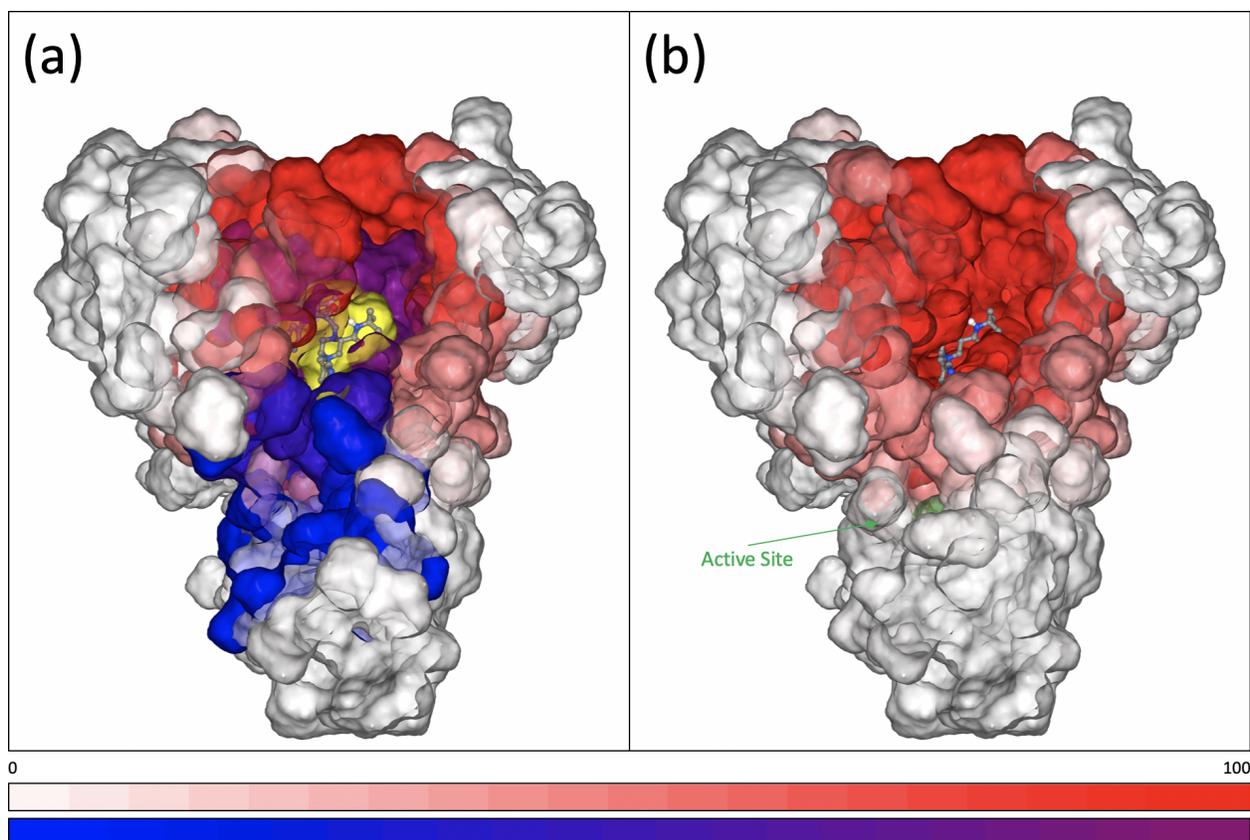


Figure 1. Visualization of the protease-Indinavir docking result. **a** | The overlap of the ligand heat map and the protein pocket. The red color scale of the ligand heat map denotes the frequency of binding sites over one hundred times. The residues in the protein pocket were colored by a blue-purple color scale, where purple represents a high frequency on the ligand heat map and blue represents a low frequency. A hot spot is defined as a set of residues with the highest frequency in the ligand heat map. Among the docking results where the ligand has contacts with the hot spot, the ligand with the lowest docking score is shown in sticks (CPK color schema) and the space it occupies is colored in yellow. **b** | The active site of 3CL-protease (green).

Table 2. The docking score of candidate ligands to RNA-dependent RNA polymerase

PubChem CID	Name	AutoDock Vina	RosettaCommons
121304016	Remdesivir	-7.803	-2413.7
10445549	Galidesivir	-6.806	-2411.3
37542	Ribavirin	-6.233	-2409.7
492405	Favipiravir	-5.495	-2409.9

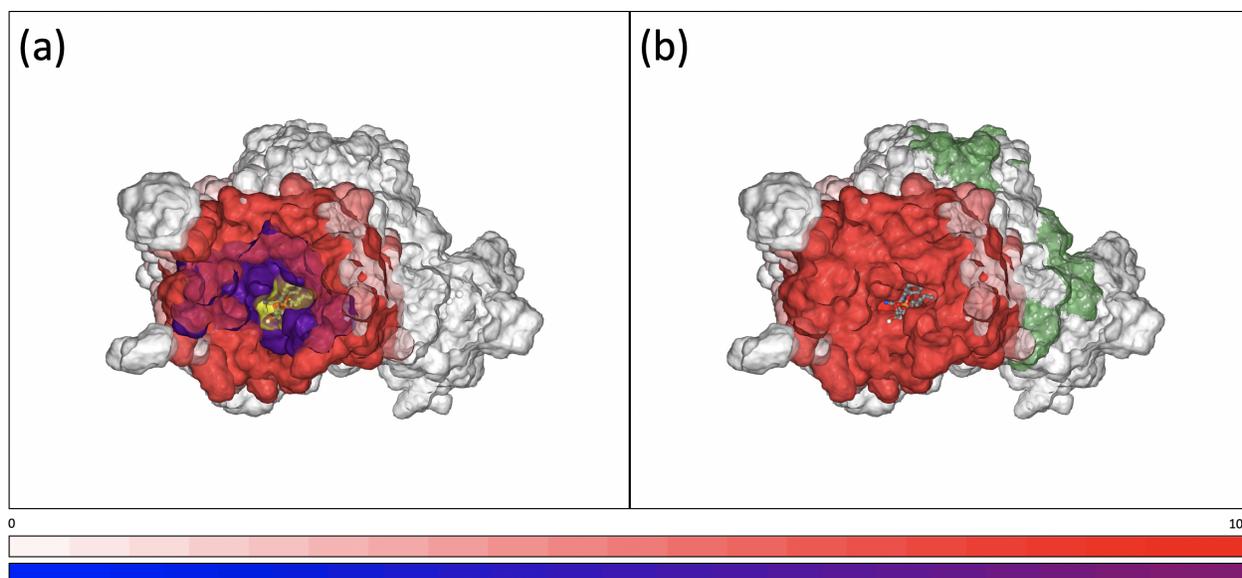


Figure 2. Visualization of the RNA dependent RNA polymerase-docking result. a | The overlap of the ligand heat map and the protein pocket. The red color scale of the ligand heat map denotes the frequency of binding sites over one hundred times. The residues in the protein pocket were colored by a blue-purple color scale, where purple represents a high frequency on the ligand heat map and blue represents a low frequency. A hot spot is defined as a set of residues with the highest frequency in the ligand heat map. Among the docking results where the ligand has contacts with the hot spot, the ligand with the lowest docking score is shown in sticks (CPK color schema) and the space it occupies is colored in yellow. b | The active sites of RdRp (green).



Supplement Figure 1. The structural alignment result of the homology modeled COVID-19 (colored in dark blue) and the SARS polymerase structure (RdRp, PDB ID: 6NUR, colored in orange). The Root-mean-square deviation (RMSD) between 793 aligned atom pairs is 0.073 angstroms, which indicates that these two proteins shared very similar structure.

References

1. Shuter, J. Lopinavir/ritonavir in the treatment of HIV-1 infection: a review. *Therapeutics and Clinical Risk Management* vol. 4 1023–1033 (2008).
2. Agostini, M. L. *et al.* Coronavirus Susceptibility to the Antiviral Remdesivir (GS-5734) Is Mediated by the Viral Polymerase and the Proofreading Exoribonuclease. *MBio* **9**, (2018).
3. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
4. Liu, X., Zhang, B., Jin, Z., Yang, H. & Rao, Z. The crystal structure of 2019-nCoV main protease in complex with an inhibitor N3. (2020) doi:10.2210/pdb6lu7/pdb.
5. Elbe, S. & Buckland-Merrett, G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob Chall* **1**, 33–46 (2017).
6. Shu, Y. & McCauley, J. GISAID: Global initiative on sharing all influenza data - from vision to reality. *Euro Surveill.* **22**, (2017).
7. Kirchdoerfer, R. N. & Ward, A. B. SARS-Coronavirus NSP12 bound to NSP7 and NSP8 co-factors. (2019) doi:10.2210/pdb6nur/pdb.
8. Waterhouse, A. *et al.* SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).
9. Bienert, S. *et al.* The SWISS-MODEL Repository-new features and functionality. *Nucleic Acids Res.* **45**, D313–D319 (2017).
10. Guex, N., Peitsch, M. C. & Schwede, T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. *Electrophoresis* **30 Suppl 1**, S162–73 (2009).
11. Benkert, P., Biasini, M. & Schwede, T. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics* **27**, 343–350 (2011).
12. Bertoni, M., Kiefer, F., Biasini, M., Bordoli, L. & Schwede, T. Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. *Sci. Rep.* **7**, 10480 (2017).
13. Lv, Z., Chu, Y. & Wang, Y. HIV protease inhibitors: a review of molecular selectivity and toxicity. *HIV AIDS* **7**, 95–104 (2015).
14. Favipiravir (T-705), a novel viral RNA polymerase inhibitor. *Antiviral Res.* **100**, 446–454 (2013).
15. Lim, S.-Y. *et al.* Galidesivir, a Direct-Acting Antiviral Drug, Abrogates Viremia in Rhesus Macaques Challenged with Zika Virus. *Open Forum Infect Dis* **4**, S55–S55 (2017).
16. Ribavirin and interferon- β synergistically inhibit SARS-associated coronavirus replication in animal and human cell lines. *Biochem. Biophys. Res. Commun.* **326**, 905–908 (2005).
17. Trott, O. & Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31**, 455–461 (2010).
18. Meiler, J. & Baker, D. ROSETTALIGAND: protein-small molecule docking with full side-chain flexibility. *Proteins* **65**, 538–548 (2006).
19. Davis, I. W. & Baker, D. RosettaLigand docking with full ligand and receptor flexibility. *J. Mol. Biol.* **385**, 381–392 (2009).
20. Lemmon, G. & Meiler, J. Rosetta Ligand docking with flexible XML protocols. *Methods Mol. Biol.* **819**, 143–155 (2012).

21. Yoshikawa, N. & Hutchison, G. R. Fast, efficient fragment-based coordinate generation for Open Babel. *J. Cheminform.* **11**, 1–9 (2019).
22. Tian, W., Chen, C., Lei, X., Zhao, J. & Liang, J. CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic Acids Res.* **46**, W363–W367 (2018).
23. UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506–D515 (2019).
24. Li, G. & De Clercq, E. Therapeutic options for the 2019 novel coronavirus (2019-nCoV). *Nature Reviews Drug Discovery* (2020) doi:10.1038/d41573-020-00016-0.
25. Wang, M. *et al.* Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. *Cell Res.* (2020) doi:10.1038/s41422-020-0282-0.