Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 February 2020 doi:10.20944/preprints202002.0242.v2

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 Hsiao[1], Hung-Ching Chang[1], Tsung-Ting Hsieh[1], Chan-Hung Su[1], Su-Shia Wang[1], Jheng-Ying Yu[1], Shang-Hung Shih[1], Yu-Hsiang Lin[1], Yin-Hung Lin[1], Yi-Chin Ethan Tu[1], Chun-Hua Hsu[4], Hsueh-Fen Juan[2,5], Chun-Wei Tung[6,*], Chien-Yu Chen[1,7,*]

[†]These authors contributed equally to this work
*Correspondence: cwtung@tmu.edu.tw, chienyuchen@ntu.edu.tw

[1]Taiwan AI Labs, Taipei 10351, Taiwan
[2]Graduate Institute of Biomedical Electronics and Bioinformatics, National Taiwan University, Taipei 10617, Taiwan
[3]Genome and Systems biology degree program, Academia Sinica and National Taiwan University, Taipei 10617, Taiwan
[4]Department of Agricultural Chemistry, National Taiwan University, Taipei 10617, Taiwan
[5]Department of Life Science, National Taiwan University, Taipei 10617, Taiwan
[6]Graduate Institute of Data Science, College of Management, Taipei Medical University, Taipei 106, Taiwan
[7]Department of Biomechatronics Engineering, National Taiwan University, Taipei 10617, Taiwan

Abstract

The outbreak of novel coronavirus (COVID-19) infections in 2019 is in dire need of finding potential therapeutic agents. In this study, we used molecular docking to repurpose HIV protease inhibitors and nucleoside analogues for COVID-19, with evaluations based on docking scores calculated by AutoDock Vina and RosettaCommons. Our results suggest that Indinavir and Remdesivir possess the best docking scores, and comparison of the docking sites of the two drugs reveal a near perfect dock in the overlapping region of the protein pockets. After further investigation of the functional regions inferred from the proteins of SARS coronavirus, we discovered that Indinavir does not dock on any active sites of the protease, which may give rise to concern in regards to the efficacy of Indinavir. On the other hand, the docking site of Remdesivir is not compatible with any known functional regions, including template binding motifs, polymerization motifs and nucleoside triphosphate (NTP) binding motifs. However, when we tested the active form (CHEMBL2016761) of Remdesivir, the docking site revealed a perfect dock in the overlapping region of the NTP binding motif. This result suggests that Remdesivir could be a potential therapeutic agent. Clinical trials still must be done in order to confirm the curative effect of these drugs.

Introduction

In the concluding weeks of 2019, an outbreak of novel coronavirus (COVID-19) infections occurred in Wuhan, China. As of February 25, 2020, more than 80,000 cases and 2,700 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 certain 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 proteins[1]. Likewise, various clinical trials are also

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 February 2020 doi:10.20944/preprints202002.0242.v2

now being undergone on nucleoside analogue drugs such as Remdesivir, an antiviral drug proven to be effective against a wide range of RNA viruses in vitro[2]. This study aims to determine whether the protease of COVID-19 can be a target protein for Lopinavir and Ritonavir, and attempts 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

For our target receptors, we chose the 3-chymotrypsin-like protease (3CL-protease) and RdRp, the main protease used to cleave polyproteins into replication-related proteins, and the main protein for RNA replication respectively. The 3CL-protease structure (PDB ID: 6LU7)[3,4] was obtained from the RCSB Protein Data Bank[3], which was released on February 5th, 2020[4]. Because the structure of the COVID-19 polymerase is currently unavailable, we adapted an approximate model from the sequence of the patient, BetaCoV/Taiwan/2/2020|EPI_ISL_406031 obtained from the Global Initiative on Sharing All Influenza Data (GISAID)[5,6] using homology modeling. Our current model is based on the template of the SARS coronavirus polymerase (PDB ID: 6NUR)[7] built using the Swiss-model[8,9,10,11,12]. We achieved a root-mean-square deviation (RMSD) of 0.073 angstroms between the 793 aligned atom pairs, indicating that these two proteins share very similar structures (see Supplement Figure 1 and 2).

Based on a literature review[2,13,14,15,16], we tested ten FDA-approved HIV protease inhibiting drugs for potential treatment, including Amprenavir, Atazanavir, Darunavir, Fosamprenavir, Indinavir, Lopinavir, Nelfinavir, Ritonavir, Saquinavir, and Tipranavir[13]. As for nucleoside analogue drugs, we tested Favipiravir[14], Galidesivir[15], Remdesivir[2], and Ribavirin[16]. Each of the drug's 3D structures was obtained 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) formats 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 to dock the selected drugs to their corresponding protease and polymerase under the standard 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 in 5 Å of binding sites to represent the binding frequency of each ligand.

Being a valid drug, we anticipate the ligand to dock at either a protein pocket or a functional region. Therefore, we used the CASTp[22] tool to predict potential pockets of our target proteins, and confirmed whether the highest frequency binding sites of the heat map were located in a protein pocket. Additionally, due to the high similarity between the SARS and COVID-19, the catalytic sites of SARS protease obtained from the Uniprot database[23] and the functional motif of SARS polymerase obtained from previous study[24] were used to label the corresponding region of COVID-19. Finally, we superimposed the heat map, pocket sites and functional regions to visualize the binding poses.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 February 2020 doi:10.20944/preprints202002.0242.v2

Results

Listed in Table 1 are the results of the 10 ligands docked 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 best docking score, even outperforming 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 colored region highlights that the binding sites are mostly located at the protein pocket. However, one thing to note is that the active sites [His41, Cys145] of 3CL-protease does 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-protease[25]. Lastly, HIV protease belongs to the aspartic protease family, while the 3CL-like protease is of the cysteine protease family[25]. These inconsistencies may lead to unwanted noises in our docking results.

The docking scores of the four ligands to RdRp using AutoDock Vina and RosettaCommons are listed in Table 2. Remdesivir has the best docking performance by both AutoDock Vina and RosettaCommons. This suggests that Remdesivir has the most stable docked structure compared to that of the other drug candidates. Although Remdesivir has the lowest docking score, owing to Remdesivir's nucleoside analog prodrug properties, we have to further examine its different formations[26]. We applied AutoDock Vina and RosettaCommons on three formations of Remdesivir, including Remdesivir (Nucleoside Analogue Monophosphate Prodrug with Protect Group), GS441524 (Nucleoside Analogue) and CHEMBL2016761 (Nucleoside Analogue Triphosphate). Listed in Table 3, these results show that only CHEMBL2016761, the active form of nucleoside analogues that act by blocking viral replication, still has a good docking score. We therefore conclude that the formation of a drug will affect the docking score estimation, and if available, the active form should be used for more accurate estimation.

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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 February 2020 doi:10.20944/preprints202002.0242.v2

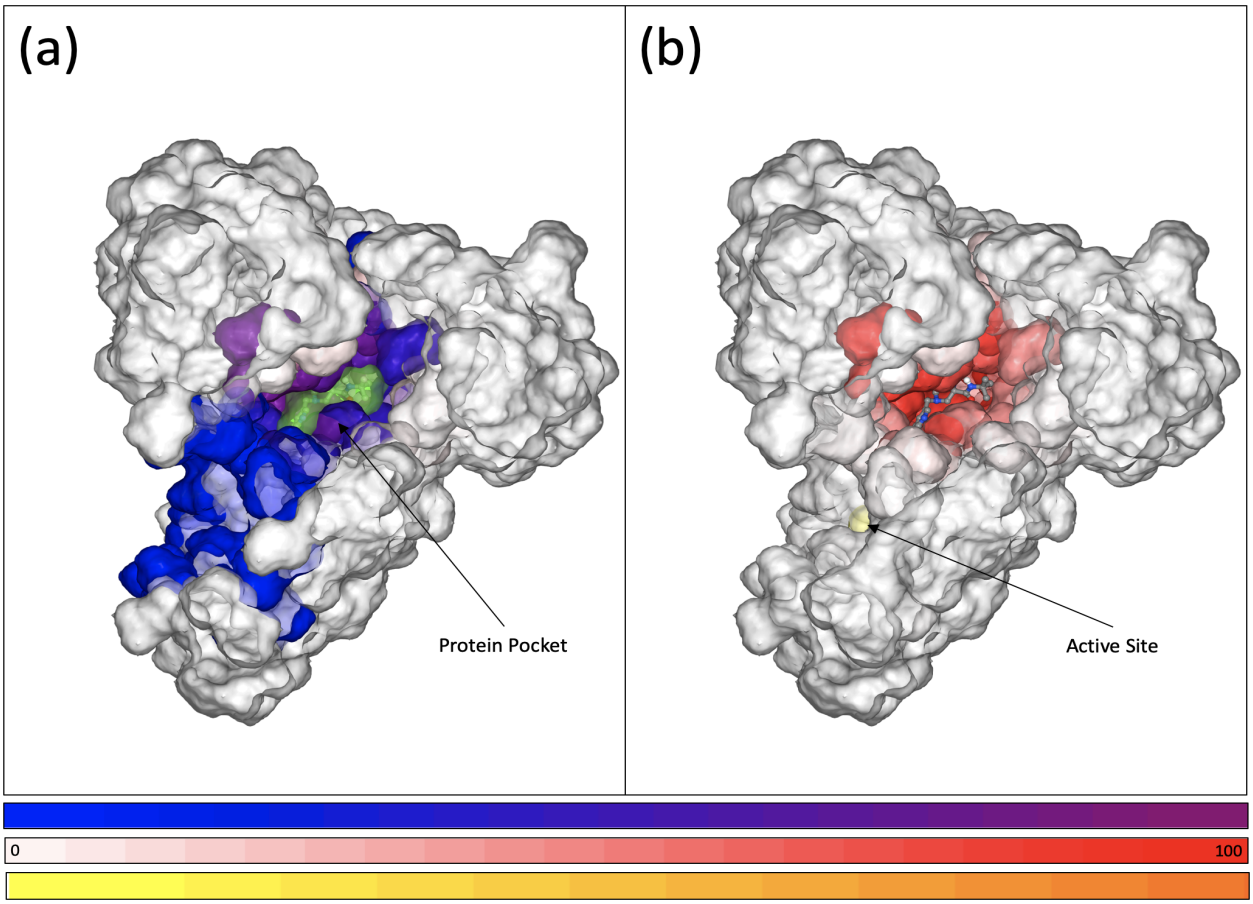


Figure 1. Visualization of the protease-Indinavir docking result. **a** | The overlap of the ligand heat map and the protein pocket. The red scale of the ligand heat map denotes the frequency of binding sites over one hundred trials. The residues in the protein pocket are noted through the blue-purple 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 contact 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 green. **b** | There is no overlap between the ligand heat map and the active site of 3CL-protease. The red scale of the ligand heat map denotes the frequency of binding sites over one hundred trials. The residues in the active site of 3CL-protease are noted through the yellow-orange scale, where orange represents a high frequency on the ligand heat map and yellow represents a low frequency.

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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 February 2020 doi:10.20944/preprints202002.0242.v2

We examined the docking results of the given structural complex between CHEMBL2016761, the active form of Remdesivir, and RdRp by looking at the contact frequencies between CHEMBL2016761 and RdRp, shown in Figure 2a. We observed that the docking orientation with the highest docking consensus lies in the binding pocket of the RdRp. After further investigation of the functional regions inferred from the proteins of SARS coronavirus, which include the template binding motif, polymerization motifs and the NTP binding motif[24], (See Supplement Table 1), we discovered that the docking site does not dock on template binding motifs and polymerization motifs but docks perfectly in the overlapping region of the NTP binding motif (Figure 2b). This result agrees with the previous study published in Cell Research[27], which suggests that Remdesivir is highly effective in controlling COVID-19 infection *in vitro*. Although our docking results agree with *in vitro* studies, the association between these docking results and the effectiveness of treating COVID-19 still needs further examination. We urge facilities with the appropriate equipment to continue this study *in vitro.*

In conclusion, our findings reveal that Indinavir and Remdesivir possess relatively low docking scores, as well as docking sites that strongly overlap with the protein pockets. We infer that this means better structural stability of their protein-ligand complexes. However, the docking site of Indinavir is not located in the catalytic sites of 3CL-protease. This may lead to concerns regarding it's inhibition ability. As for Remdesivir, the docking site of CHEMBL2016761 is perfectly located in the NTP binding motif, which is expected to block the replication of RNA sequence. 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 3. The docking score of different formations of Remdesivir to RdRp.

PubChem CID	Name	AutoDock Vina	RosettaCommons
121304016	Remdesivir	-7.803	-2413.7
56832906	CHEMBL2016761	-7.144	-2411.1
44468216	GS441524	-6.688	-2412.5

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 February 2020 doi:10.20944/preprints202002.0242.v2

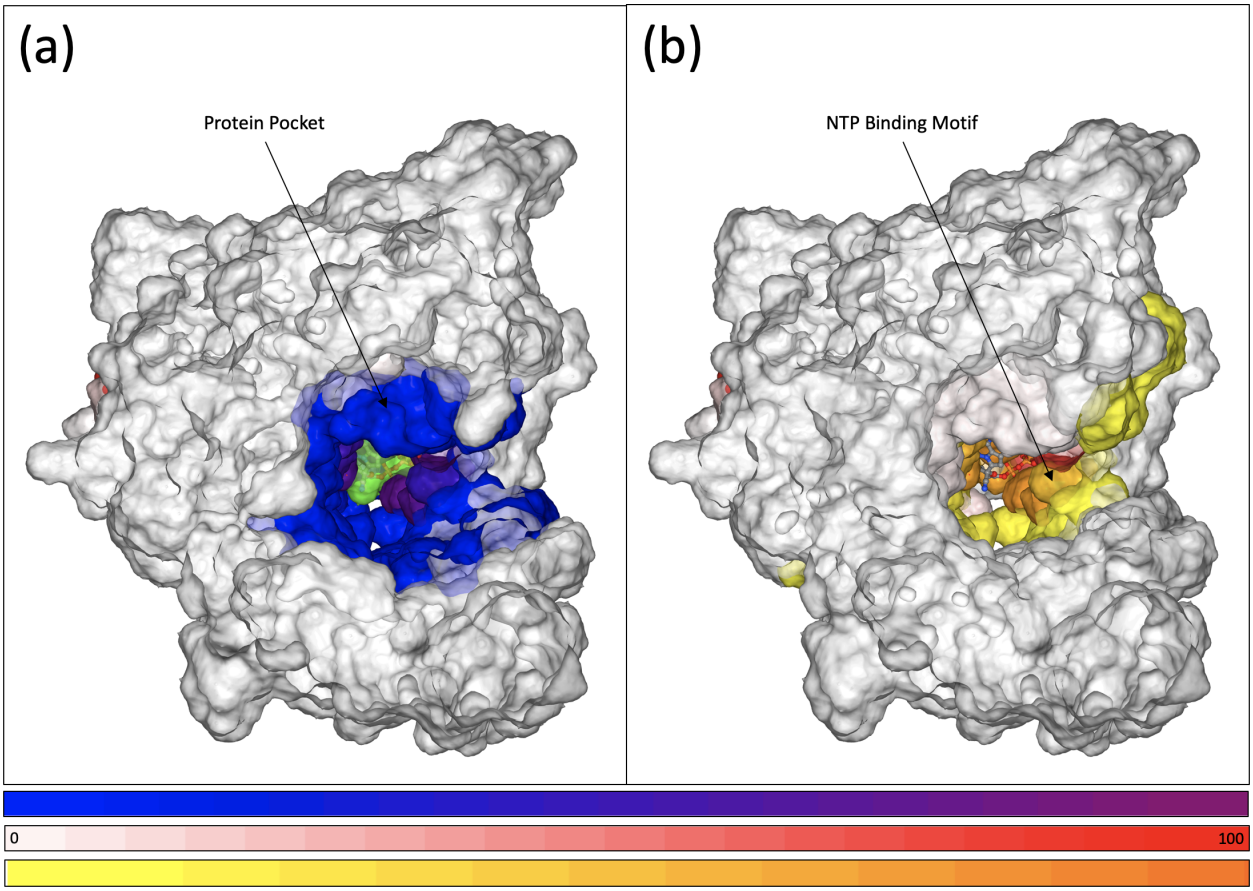


Figure 2. Visualization of the RNA dependent RNA polymerase-docking result. **a** | The overlap of the ligand heat map and the protein pocket. The red scale of the ligand heat map denotes the frequency of binding sites over one hundred trials. The residues in the protein pocket are noted through the blue-purple scale, where purple represents a high frequency on the ligand heat map and blue represents a low frequency. Among the docking results where the ligand has contact 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 green. **b** | The overlap between the ligand heat map and the NTP binding motif of RdRp. The red scale of the ligand heat map denotes the frequency of binding sites over one hundred trials. The residues in the functional regions of RdRp are noted through the yellow-orange scale, where orange represents a high frequency on the ligand heat map and yellow represents a low frequency.

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. Kirchdoerfer, R. N. & Ward, A. B. Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. *Nat. Commun.* **10**, 2342 (2019).
25. Li, G. & De Clercq, E. Therapeutic options for the 2019 novel coronavirus (2019-nCoV).

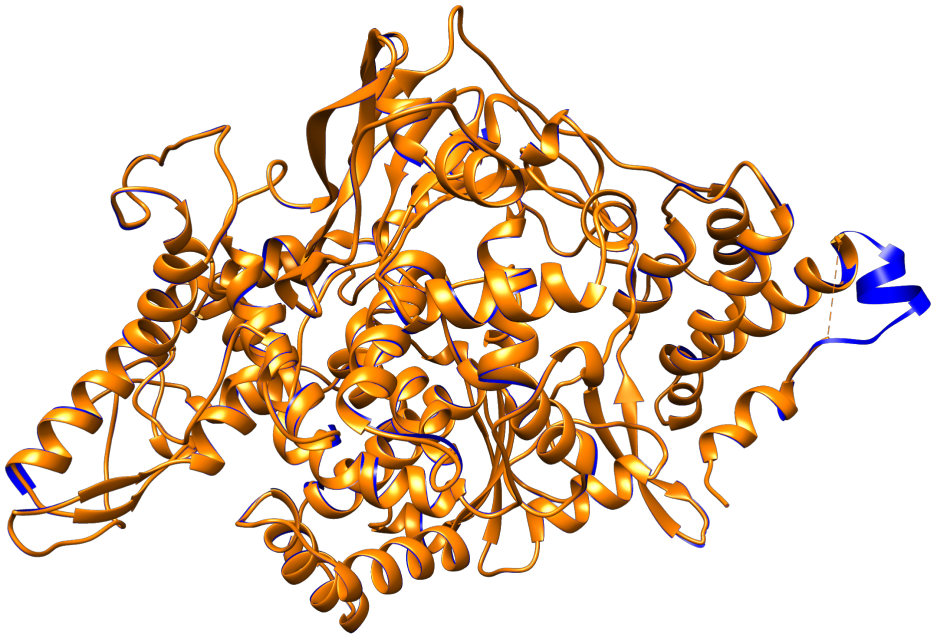
Nature Reviews Drug Discovery (2020) doi:10.1038/d41573-020-00016-0.

26. Eyer, L., Nencka, R., de Clercq, E., Seley-Radtke, K. & Růžek, D. Nucleoside analogs as a rich source of antiviral agents active against arthropod-borne flaviviruses. *Antivir. Chem. Chemother.* **26**, 2040206618761299 (2018).
27. 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.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 February 2020 doi:10.20944/preprints202002.0242.v2

Supplement Figure 1. The sequence alignment result of COVID-19 RNA polymerase (Target) and SARS RNA polymerase (6NUR). In the SWISS MODEL database, the SARS RNA polymerase model is the most similar structure with the COVID-19 RNA polymerase.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 February 2020 doi:10.20944/preprints202002.0242.v2



Supplement Figure 2. 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 RMSD between 793 aligned atom pairs is 0.073 angstroms, which indicates that these two proteins shared very similar structure.

Supplement Table 1. The sequence motifs of COVID-19 RdRp

Motif	Start	End	Sequence	Function
A	605	619	PHLMGWDYPKCDRAM	Polymerization
B	671	703	GGTSSGDATTAYANSVFNICQAVTANVNALLST	Polymerization
C	746	760	FSMMILSDDAVVCFN	Polymerization
D	764	789	ASQGLVASIKNFKSVLYYQNNVFMSE	Polymerization
E	803	813	HEFCSQHTMLV	Polymerization
F	537	548	LKYAISAKNRAR	NTP binding
G	492	506	DKSAGFPFNKWGKAR	Template binding