Prediction of the SARS-CoV-2 (2019-nCoV) 3C-like Protease (3CLpro) Structure: Virtual Screening Reveals Velpatasvir, Ledipasvir, and Other Drug Repurposing Candidates

Yu Wai Chen, Chin-Pang Yiu, Kwok-Yin Wong

Submitted date: 16/02/2020 • Posted date: 17/02/2020
Licence: CC BY-NC-ND 4.0

Citation information: Chen, Yu Wai; Yiu, Chin-Pang; Wong, Kwok-Yin (2020): Prediction of the SARS-CoV-2 (2019-nCoV) 3C-like Protease (3CLpro) Structure: Virtual Screening Reveals Velpatasvir, Ledipasvir, and Other Drug Repurposing Candidates. ChemRxiv. Preprint. https://doi.org/10.26434/chemrxiv.11831103.v2

We prepared the three-dimensional model of the 2019-nCoV 3C-like protease (3CL\textsuperscript{pro}) using the crystal structure of the highly-similar (96% identity) ortholog from the SARS-CoV. All residues involved in the catalysis, substrate binding and dimerisation are 100% conserved. Comparison of the polyprotein PP1AB sequences showed 86% identity. The 3C-like cleavage sites on the coronaviral polyproteins are highly conserved. Based on the near-identical substrate specificities and high sequence identities, we are in the opinion that some of the previous progress of specific inhibitors development for the SARS-CoV enzyme can be conferred on its 2019-nCoV counterpart. With the 3CL\textsuperscript{pro} molecular model, we performed virtual screening for purchasable drugs and proposed 16 candidates for consideration. Among these, the antivirals ledipasvir or velpatasvir are particularly attractive as therapeutics to combat the 2019-nCoV with minimal side effects, commonly fatigue and headache. The drugs Epclusa (velpatasvir / sofosbuvir) and Harvoni (ledipasvir / sofosbuvir) could be very effective owing to their dual inhibitory actions on two viral enzymes.

File list (2)

| File name                        | Size      | View on ChemRxiv | Download file |
|----------------------------------|-----------|-------------------|---------------|
| Prediction nCoV-2019 3CL v2.9T Main.pdf (6.47 MiB) | view on ChemRxiv | download file |
| Prediction nCoV-2019 3CL v2.9T Supp.pdf (4.56 MiB) | view on ChemRxiv | download file |
Prediction of the SARS-CoV-2 (2019-nCoV) 3C-like protease (3CL\textsuperscript{pro}) structure: virtual screening reveals velpatasvir, ledipasvir, and other drug repurposing candidates

Yu Wai Chen\textsuperscript{1*}, Chin-Pang Bennu Yiu\textsuperscript{2}, Kwok-Yin Wong\textsuperscript{1*}

\textsuperscript{1} Department of Applied Biology and Chemical Biology and the State Key Laboratory of Chemical Biology and Drug Discovery, The Hong Kong Polytechnic University, Hunghom, Hong Kong.

\textsuperscript{2} Flat E, 11/F, Block 1, La Costa, Ma On Shan, Hong Kong.

*Corresponding email: yu-wai.chen@polyu.edu.hk ; kwok-yin.wong@polyu.edu.hk

ORCID
YWC: 0000-0001-7833-7533
CPBY: 0000-0003-3965-3992
KYW: 0000-0003-4984-7109

Keywords: COVID-19, 2019-nCoV, 3C-like protease, drug repurpose, antiviral, coronavirus, virtual screening, molecular modelling, Hepatitis C virus, HCV, ledipasvir, velpatasvir
Abstract

We prepared the three-dimensional model of the SARS-CoV-2 (aka 2019-nCoV) 3C-like protease (3CLpro) using the crystal structure of the highly-similar (96% identity) ortholog from the SARS-CoV. All residues involved in the catalysis, substrate binding and dimerisation are 100% conserved. Comparison of the polyprotein PP1AB sequences showed 86% identity. The 3C-like cleavage sites on the coronaviral polyproteins are highly conserved. Based on the near-identical substrate specificities and high sequence identities, we are in the opinion that some of the previous progress of specific inhibitors development for the SARS-CoV enzyme can be conferred on its SARS-CoV-2 counterpart. With the 3CLpro molecular model, we performed virtual screening for purchasable drugs and proposed 16 candidates for consideration. Among these, the antivirals ledipasvir or velpatasvir are particularly attractive as therapeutics to combat the new coronavirus with minimal side effects, commonly fatigue and headache. The drugs Epclusa (velpatasvir / sofosbuvir) and Harvoni (ledipasvir / sofosbuvir) could be very effective owing to their dual inhibitory actions on two viral enzymes.
Introduction

On 7 January 2020, a new coronavirus, 2019-nCoV (now officially named SARS-CoV-2) was implicated in an alarming outbreak of a pneumonia-like illness COVID-19, originating from Wuhan City, Hubei, China. Human-to-human transmission was first confirmed in Guangdong, China. The World Health Organisation has declared this a global public health emergency — on 15 February 2020, there are more than 65,000 confirmed cases reported globally, and the death toll is over 1500. In the height of the crisis, this virus is spreading at a rate and scale far worse than previous coronaviral epidemics.

It was immediately evident from its genome that the coronavirus is evolutionarily related (80% identity) to the beta-coronavirus implicated in the severe acute respiratory syndrome (SARS), which has a bat origin and was causative of a global outbreak in 2003. The momentum of research on developing antiviral agents against the SARS-CoV carried on after the epidemic subsided. Despite that no SARS treatment has yet come to fruition, knowledge acquired from the extensive research and development efforts may be of use to inform the current therapeutic options.

The viral genome encodes more than 20 proteins, among them there are two proteases (PLpro and 3CLpro) that are vital to the virus’ replication: they cleave the two translated polyproteins (PP1A and PP1AB) into individual functional components. 3-chymotrypsin-like protease (3CLpro aka main protease, Mpro) is considered to be a promising drug target. Tremendous effort has been spent on studying this protein in order to identify therapeutics against the SARS-CoV in particular and other pathogenic coronaviruses (e.g. MERS-CoV, the Middle East respiratory syndrome coronavirus) in general because they share similar active sites and enzymatic mechanisms. The purpose of this study is to build a molecular model of the 3CLpro of the SARS-CoV-2 and to carry out virtual screening to identify readily usable therapeutics. It was not our intention, however, to comment on other structure-based drug design research as these will not be timely for the current epidemic.
Methods

Analysis of protein sequences

The translated polyprotein (PP1AB) sequence was obtained from the annotation of the respective GenBank entry of the SARS-CoV-2 genome. By comparing this sequence with the SARS-CoV PP1AB sequence, the protease cleavage sites and all mature protein sequences were obtained. Sequence comparison and alignment were performed with BLASTp (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins).

Preparation of structural model

The highest-resolution apo-enzyme structure of SARS-CoV 3CLpro (PDBID: 2duc) \(^2\) was employed as the template. The variant residues were “mutated” \(\text{in silico}\) by SCWRL4 \(^3\), followed by manual adjustment to ensure that the best side-chain rotamer was employed. The rebuilt model was subjected to steepest descent energy minimisation by Gromacs 2018.4 (www.gromacs.org) using the Gromos 54A7 forcefield, with a restraint force constant of 1000 kJ mol\(^{-1}\) nm\(^{-2}\) applied on all backbone atoms and all atoms of the vital residues (Table 1).

Virtual screening

MTiOpenScreen web service \(^4\) was used for screening against its library of 7173 purchasable drugs (Drugs-lib), with the binding site grid specified by the active-site residues. The active sites on chain A and chain B were screened independently with AutoDock Vina \(^5\). A list of 4,500 target:ligand docking combinations ranked by binding energies was produced for each screen.

Results

High sequence homology with SARS-CoV

The first available genome is GenBank MN908947, now NCBI Reference Sequence NC_045512. From it, the PP1AB sequence of SARS-CoV-2 is aligned with that of SARS-
The overall amino-acid sequence identity is very high (86%). The conservation is noticeable at the polyprotein cleavage sites. All the eleven 3CL\textsuperscript{pro} sites \textsuperscript{2} are highly conserved or identical (Table S1), inferring that their respective proteases have very similar specificities. The 3CL\textsuperscript{pro} sequence of SARS-CoV-2 has only 12 out of 306 residues different from that of SARS-CoV (identity = 96%).

**Conserved sequence identity among SARS-CoV-2**

We compared the polyprotein PP1AB and the 3CL\textsuperscript{pro} sequences among all 11 genomes (GenBank MN908947, MN938384, MN975262, MN985325, MN988668, MN988669, MN988713, MN994467, MN994468, MN996527 and MN996528) which were available on 1 February, 2020. With reference to MN908947 (NC_045512), among the 7096 residues, there is only one variable residue in each of MN975262 (in NSP-4), MN994467 (in NSP-2), MN994468 (in NSP-13), MN996527 (in NSP-16); and 2 in MN988713 (in NSP-1 and NSP-3). The remaining five have no difference. To summarise, all SARS-CoV-2 3CL\textsuperscript{pro} sequences and all their cleavage junctions on their polyproteins are 100% conserved.

**3D model of the SARS-CoV-2 3CL\textsuperscript{pro}**

The amino acids that are known to be important for the enzyme’s functions are listed in Table 1. Not unexpectedly, none of the 12 variant positions are involved in major roles. Therefore, we are confident to prepare a structural model of the SARS-CoV-2 3CL\textsuperscript{pro} by molecular modelling (Figure S1), which will be immediately useful for in silico development of targeted treatment. After we have submitted the first draft of this study, the crystal structure of SARS-CoV-2 3CL\textsuperscript{pro} was released (PDB ID 6lu7), which confirms that the predicted model is good within experimental errors (Figure S2).

**Virtual screening for readily-available drugs**

The top 10 or 11 (with a binding energy cut-off) hits of chains A and B (Table 3) were examined visually in PyMOL \textsuperscript{7} and all solutions were found to fit into their respective active sites convincingly. The binding energies of chain A complexes were generally higher than
those of chain B by approximately 1.4 kcal mol⁻¹. This presumably demonstrates the intrinsic conformational variability between the A- and B-chain active sites in the crystal structure (the average root-mean-square deviation (rmsd) in Cα atomic positions of active-site residues is 0.83Å). In each screen, the differences in binding energies are small, suggesting that the ranking is not discriminatory, and all top scorers should be examined. We combined the two screens and found 16 candidates which give promising binding models (etoposide and its phosphate count as one).

Assessment of the candidate drugs

We checked the actions, targets and side effects of the 16 candidates. Among these we first noticed velpatasvir (Figures 1A, 1D) and ledipasvir, which are inhibitors of the NS5A protein of the hepatitis C virus (HCV). Both are marketed as approved drugs in combination with sofosbuvir which is a prodrug nucleotide analogue inhibitor of RNA-dependent RNA polymerase (RdRp, or NS5B). Interestingly, sofosbuvir has recently been proposed as an antiviral for the SARS-CoV-2 based on the similarity between the replication mechanisms of the HCV and the coronavirus ⁸. Our results further strengthen that these dual-component HCV drugs, Epclusa (velpatasvir / sofosbuvir) and Harvoni (ledipasvir / sofosbuvir), may be attractive candidates to repurpose because they may inhibit two coronaviral enzymes. A drug that can target two viral proteins substantially reduces the ability of the virus to develop resistance. These direct-acting antiviral drugs are also associated with very minimal side effects and are conveniently orally administered (Table 4).

The flavonoid glycosides, diosmin (Figure 1B) and hesperidin (Figure 1E) from citrus fruits, fit very well into and block the substrate binding site. Yet, these compounds cause mild adverse reactions (Table 4). Hesperidin hits showed up multiple times suggesting it has many modes of binding (Figure 1A), Teniposide and etoposide (and its phosphate) are chemically related and turned up in multiple hits with good binding models (Figure 1F). However, these chemotherapy drugs have a lot of strong side effects and need intravenous administration (Table 4). The approved venetoclax (Figure 1C) and investigational drugs MK-3207 and R428 scored well in both screens. Venetoclax is another chemotherapy drug that is burdened by side effects including upper respiratory tract infection (Table 4). Not a lot are disclosed about MK-3207 and R428.
We subjected the crystal structure with an empty active site to the same virtual screening procedures. The same list of candidates showed up consistently (Table S2) with high scores although ledipasvir is not found.

We noticed that most of the compounds on the list have molecular weights (MW) over 500, except lumacaftor (MW=452). The largest one is ledipasvir (MW=889). This is because the size of the peptide substrate and the deeply buried protease active site demand a large molecule that has many rotatable dynamics to fit into it.

Discussion

We performed a search on the USFDA (clinicaltrials.gov) and identified five trials involving antiviral and immunomodulatory drug treatments for SARS (Table 5), all without reported results; i.e. at present, there are no safe and effective drug candidates against SARS-CoV. This is because once the epidemic is over, there are no patients to recruit for clinical trials. Only the study with streptokinase succeeded to complete phase 3. It is disappointing that little progress in SARS drug development has been made in the past 17 years. After the 2003 outbreak, numerous inhibitors for the $3\text{CL}^{\text{pro}}$ enzyme have been proposed\textsuperscript{9,10}, yet no new drug candidates have succeeded to enter the clinical phase 1.

One record which receives a lot of attention amid the current outbreak is the lopinavir / ritonavir combination. They are protease inhibitors originally developed against HIV. During the 2003 SARS outbreak, despite lacking a clinical trial, they were tried as an emergency measure and found to offer improved clinical outcome\textsuperscript{11}. However, some scientists did express scepticism\textsuperscript{12}. By analogy, these compounds were speculated to act on SARS-CoV $3\text{CL}^{\text{pro}}$ specifically but there is as yet no crystal structure to support that, although docking studies were carried out to propose various binding modes\textsuperscript{13-16}. The IC\textsubscript{50} value of lopinavir is 50 μM ($K_i = 14 \mu$M) and that for ritonavir cannot be established\textsuperscript{17}. Although this is far from a cure, based on our results that the two CoV $3\text{CL}^{\text{pro}}$ enzymes are identical as far as protein sequences and substrate specificities are concerned, we are in the opinion that this is still one of the recommended routes for immediate treatment at the time of writing (early February, 2020).
If we look beyond the 3CL\textsuperscript{pro}, an earlier screen produced 27 candidates that could be repurposed against both SARS-CoV and MERS-CoV\textsuperscript{18}. In addition, the other coronaviral proteins could be targeted for screening. Treatment of COVID-19 with remdesivir (a repurposed drug in development targeting the RdRp) showing improved clinical outcome has just been reported and clinical trial is now underway\textsuperscript{19}.

We consider this work part of the global efforts responding in a timely fashion to fight this deadly communicable disease. We are aware that there are similar modelling, screening and repurposing exercises targeting 3CL\textsuperscript{pro} reported or announced\textsuperscript{13,20-26}. Our methods did not overlap and we share no common results with these studies.

**Data Availability**

Open Science Framework: “SARS-CoV-2 (2019-nCoV) 3CLpro Model and Screening”\textsuperscript{6}, 10.17605/OSF.IO/7X9A2

This data project (dataset) contains the following data, under the “Virtual Screening” folder:

1. “2019-nCoV-3CLpro.pdb” (3D model of the 3CL\textsuperscript{pro}: A and B chains)

2. “A-screen4500.pdbqt, B-screen4500.pdbqt, X-screen4500.pdbqt”

   (Virtual screening 3D results of Model A chain, Model B chain and the crystal-structure (A chain) in PDBQT format (can be viewed by any text editor). Use the software PyMOL to open these files. Each result file contains 4500 drug-to-protein docking hits ranked by AutoDock Vina binding energies in kcal mol\textsuperscript{-1}).

3. “A-screen1500.table.csv, B-screen1500.table.csv, X-screen1500.table.csv”

   (Virtual screening results (names only) of Model A chain, Model B chain and the crystal-structure (A chain) in CSV format (can be opened by Excel or any text editor). This is a summary of the top 1500 drug-to-protein docking hits ranked by AutoDock Vina binding energies in kcal mol\textsuperscript{-1}).

Data are available under the terms of the Creative Commons Zero "No rights reserved" data waiver (CC0 1.0 Public domain dedication).
Acknowledgement

We acknowledge support from the Innovation and Technology Commission of Hong Kong, the Hong Kong Polytechnic University and the Life Science Area of Strategic Fund 1-ZVH9.

Author Contributions

Y.W.C. and C.-P.B.Y. conducted the research and wrote the paper. K.-Y.W. read the manuscript and contributed to the funding.

Conflict of Interest

The authors declare no competing interests.

References

1 Chan, J. F. et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet* 10.1016/S0140-6736(20)30154-9 (2020).
2 Muramatsu, T. et al. SARS-CoV 3CL protease cleaves its C-terminal autoprocessing site by novel subsite cooperativity. *Proc. Natl. Acad. Sci. U. S. A.* 113, 12997-13002 (2016).
3 Krivov, G. G., Shapovalov, M. V. & Dunbrack, R. L., Jr. Improved prediction of protein side-chain conformations with SCWRL4. *Proteins* 77, 778-795 (2009).
4 Labbé, C. M. et al. MTiOpenScreen: a web server for structure-based virtual screening. *Nucleic Acids Res.* 43, W448-W454 (2015).
5 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).
6 Chen, Y. W. SARS-CoV-2 (2019-nCoV) 3CLpro Model & Screening. *Open Science Framework* 10.17605/OSF.IO/7X9A2 (2020).
7 The PyMOL molecular graphics system (Schrödinger, LLC).
8 Ju, J., Kumara, S., Li, X., Jockusch, S. & Russo, J. J. Nucleotide analogues as inhibitors of viral polymerases. *bioRxiv* 10.1101/2020.01.30.927574 (2020).
An overview of severe acute respiratory syndrome-coronavirus (SARS-CoV) 3CL protease inhibitors: peptidomimetics and small molecule chemotherapy. *J. Med. Chem.* **59**, 6595-6628 (2016).

Characterization and inhibition of the main protease of severe acute respiratory syndrome coronavirus. *ChemBioEng Reviews* **2**, 118-132 (2015).

Role of lopinavir/ritonavir in the treatment of SARS: initial virological and clinical findings. *Thorax* **59**, 252-256 (2004).

SARS: systematic review of treatment effects. *PLoS Med.* **3**, e343 (2006).

Wuhan coronavirus 2019-nCoV - what we can find out on a structural bioinformatics level https://innophore.com/2019-ncov/ (2020).

Role of lopinavir/ritonavir in the treatment of coronavirus infection: insight from molecular docking study. *Arch. Clin. Infect. Dis.* **12**, e13823 (2017).

Molecular dynamic simulations analysis of ritonavir and lopinavir as SARS-CoV 3CL\(^\text{pro}\) inhibitors. *J. Theor. Biol.* **254**, 861-867 (2008).

Old drugs as lead compounds for a new disease? Binding analysis of SARS coronavirus main proteinase with HIV, psychotic and parasite drugs. *Bioorg. Med. Chem.* **12**, 2517-2521 (2004).

Small molecules targeting severe acute respiratory syndrome human coronavirus. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 10012-10017 (2004).

Repurposing of clinically developed drugs for treatment of Middle East respiratory syndrome coronavirus infection. *Antimicrob. Agents Chemother.* **58**, 4885-4893 (2014).

First case of 2019 novel coronavirus in the United States. *N. Engl. J. Med.* 10.1056/NEJMoa2001191 (2020).

Nelfinavir was predicted to be a potential inhibitor of 2019-nCov main protease by an integrative approach combining homology modelling, molecular docking and binding free energy calculation. *bioRxiv* 10.1101/2020.01.27.921627 (2020).

Potential inhibitors for 2019-nCoV coronavirus M protease from clinically approved medicines. *bioRxiv* 10.1101/2020.01.29.924100 (2020).

Homology models of Wuhan coronavirus 3CL\(^\text{pro}\) protease. *ChemRxiv* (2020).

Potential inhibitors for 2019-nCoV coronavirus M protease from clinically approved medicines. *bioRxiv* 10.1101/2020.01.29.924100 (2020).
Global Health Drug Discovery Institute. Targeting 2019-nCoV Portal https://ghddi-ailab.github.io/Targeting2019-nCoV/nCov_Structural_Modeling/ (2020).

Beck, B. R., Shin, B., Choi, Y., Park, S. & Kang, K. Predicting commercially available antiviral drugs that may act on the novel coronavirus (2019-nCoV), Wuhan, China through a drug-target interaction deep learning model. bioRxiv 10.1101/2020.01.31.929547 (2020).

Gao, K., Nguyen, D. D., Wang, R. & Wei, G. W. Machine intelligence design of 2019-nCoV drugs. bioRxiv 10.1101/2020.01.30.927889 (2020).

Li, Y. et al. Therapeutic drugs targeting 2019-nCoV main protease by high-throughput screening. bioRxiv 10.1101/2020.01.28.922922 (2020).

Huang, C., Wei, P., Fan, K., Liu, Y. & Lai, L. 3C-like proteinase from SARS coronavirus catalyzes substrate hydrolysis by a general base mechanism. Biochemistry 43, 4568-4574 (2004).

Hsu, M. F. et al. Mechanism of the maturation process of SARS-CoV 3CL protease. J. Biol. Chem. 280, 31257-31266 (2005).

Barrila, J., Bacha, U. & Freire, E. Long-range cooperative interactions modulate dimerization in SARS 3CLpro. Biochemistry 45, 14908-14916 (2006).

Barrila, J., Gabelli, S. B., Bacha, U., Amzel, L. M. & Freire, E. Mutation of Asn28 disrupts the dimerization and enzymatic activity of SARS 3CLpro. Biochemistry 49, 4308-4317 (2010).

Hu, T. et al. Two adjacent mutations on the dimer interface of SARS coronavirus 3C-like protease cause different conformational changes in crystal structure. Virology 388, 324-334 (2009).

Chen, S. et al. Residues on the dimer interface of SARS coronavirus 3C-like protease: dimer stability characterization and enzyme catalytic activity analysis. J Biochem 143, 525-536 (2008).

Lovell, S. C., Word, J. M., Richardson, J. S. & Richardson, D. C. The penultimate rotamer library. Proteins 40, 389-408 (2000).
| Function           | Residue Number                  | Reference |
|--------------------|--------------------------------|-----------|
| Catalytic          | 41, 145                         | 27        |
| Substrate binding  | 41, 49, 143-144, 163-167, 187-192 | 2, 28     |
| Dimerisation       | 10, 11, 14, 28, 139, 140, 147, 298 | 29-32     |
| 2019-nCoV variants | 35, 46, 65, 86, 88, 94, 134, 180, 202, 267, 285, 286 | This work |

**Table 1.** SARS-CoV 3CL\textsuperscript{pro} important residues and SARS-CoV-2 variant residues.
| Residue | Rotamer | ASA, Å\(^2\) (%) | Remarks on replacement |
|---------|---------|-------------------|------------------------|
| T35V    | AB: (t-), top | 19 (14%) | conservative |
| A46S    | A: (t-), 3rd; B: (p-), top | 73 (63%) | A chain disordered, rotamer chosen to minimise steric clash |
| S65N    | AB: (m-20), top | 38 (28%) |
| L86V    | A: (m), 2nd; B: (t), top | 0 (0%) | A chain rotamer to avoid clash |
| R88K    | A: (mtpt), 9th; B: (mtpp), 19th | 81 (33%) | AB: real-space refined with good fit to R densities |
| S94A    | not applicable | 64 (51%) |
| H134F   | AB: (m-85), top | 57 (29%) | occupy similar but larger space |
| K180N   | AB: (m-20), top | 102 (50%) |
| L202V   | AB: (p), 3rd | 22 (12%) | avoid steric clash |
| A267S   | AB: (m), 2nd | 0 (0%) | avoid steric clash |
| T285A   | not applicable | 68 (44%) | at dimeric interface |
| I286L   | (mt), top | 75 (46%) | at dimeric interface |

**Table 2.** *In silico* mutagenesis of the SARS-CoV-2 3CL\(^{\text{pro}}\). The 12 variant residues with reference to the SARS-CoV enzyme are shown with the respective treatment of rotamer. “A” and “B” refers to the individual chains of the dimeric model. Both chains are in the crystal asymmetric unit and are not identical. The rotamer symbol (bracketed) is defined according to the conventions of Richardson \(^{33}\), followed by its respective rank of popularity. ASA: accessible surface area (average of A and B chains) of the residue in the SARS-CoV 3CL\(^{\text{pro}}\) structure, in Å\(^2\) and in % relative to the ASA of a residue X in the Gly-X-Gly conformation, calculated with *areaimol* of the CCP4 suite (www.ccp4.ac.uk).
| A Chain | B Chain |
|---------|---------|
| **A Top scorers** | **B Top scorers** |
| diosmin | etoposide |
| -10.1 | -8.7 |
| hesperidin | R428 |
| -10.1 | -8.6 |
| MK-3207 | MK-3207 |
| -10.1 | -8.6 |
| venetoclax | teniposide |
| -10.0 | -8.5 |
| dihydroergocristine | UK-432097 |
| -9.8 | -8.5 |
| bolazine | eluxadoline |
| -9.8 | -8.4 |
| R428 | venetoclax |
| -9.8 | -8.4 |
| ditercalinium | ledipasvir |
| -9.8 | -8.4 |
| etoposide-phosphate | irinotecan |
| -9.8 | -8.4 |
| (B Top scorers) | (A Top scorers) |
| teniposide | hesperidin |
| -9.7 | -8.3 |
| etoposide | etoposide-phosphate |
| -9.7 | -8.3 |
| UK-432097 | bolazine |
| -9.6 | -8.3 |
| irinotecan | dihydroergocristine |
| -9.5 | -8.1 |
| lumacaftor | diosmin |
| -8.9 | -7.9 |
| velpatasvir | ditercalinium |
| -8.5 | -7.7 |
| eluxadoline | |
| -8.0 | |
| ledipasvir | |
| 0 | |

**Table 3.** The results of virtual screening of drugs on the active sites of SARS-CoV-2 3CL\(^{pro}\) model. The left and right columns are the results of A and B chains, respectively. The top scorers are listed first, then the equivalent top scorers of the other chain listed at the lower half. B.E.: AutoDock Vina binding energy in kcal mol\(^{-1}\). The number of hits of a drug is the times it appears among all results within a screen regardless of rank, only the binding energy of the top-ranking hit was shown. Etoposide and its phosphate are listed separately in the screens. Approved and pre-approved drugs are shown in green and orange, respectively. Except dihydroergocristine and ditercalinium, all approved drugs have undergone post-market surveillance i.e. Phase 4.
| Drug                  | Possible side effects (adverse reactions)                                                                 | Admin.         |
|----------------------|-----------------------------------------------------------------------------------------------------------|----------------|
| diosmin<sup>a,b</sup> | mild gastrointestinal disorders; skin irritations; nausea; heart arrhythmias                               | topical; oral  |
| hesperidin<sup>a,d</sup> | stomach pain and upset; diarrhea; headache                                                               | oral           |
| MK-3207<sup>c</sup>  | no information                                                                                           | oral           |
| venetoclax<sup>a,b</sup> | neutropenia; nausea; anaemia, diarrhea; upper respiratory tract infection                                   | oral           |
| dihydroergocristine<sup>a</sup> | no information                                                                                           | oral           |
| bolazine<sup>b</sup>   | no information                                                                                           | intramuscular  |
| R428<sup>b</sup>       | no information                                                                                           | oral           |
| diitercalinium        | no information                                                                                           | no info        |
| etoposide<sup>a,b</sup> | alopecia; constipation; diarrhea; nausea; vomiting; secondary malignancies                               | intravenous    |
| teniposide<sup>a,b</sup> | gastrointestinal toxicity; hypersensitivity reactions; reversible alopecia                                | intravenous    |
| UK-432097<sup>c</sup>  | no information                                                                                           | inhaled        |
| irinotecan<sup>a,b</sup> | gastrointestinal complication                                                                            | intravenous    |
| lumacaftor<sup>a</sup> | dyspnea; nasopharyngitis; nausea; diarrhea; upper respiratory tract infection                             | oral           |
| velpatasvir<sup>a,b</sup> | headache; fatigue; nausea                                                                                | oral           |
| eluxadoline<sup>a,b</sup> | constipation; nausea; fatigue, bronchitis, viral gastroenteritis; pancreatitis                            | oral           |
| ledipasvir<sup>a</sup>  | fatigue; headache                                                                                        | oral           |

**Table 4.** Possible side effects and routes of administration of the drugs identified from virtual screening for SARS-CoV-2 3CL<sup>pro</sup>. Sources of information: <sup>a</sup> DrugBank.ca (main), <sup>b</sup> Wikipedia.org, <sup>c</sup> clinicaltrials.gov and <sup>d</sup> WebMD.com.
| Drug                                      | Condition       | Phase   | Status   | From   | To     | Location |
|-------------------------------------------|-----------------|---------|----------|--------|--------|----------|
| Lopinavir / Ritonavir + Ribavirin         | SARS            | Unknown | Unknown  |        |        | Hong Kong|
| Alferon LDO                               | SARS            | Phase 2 | Completed| Nov 04 | Apr 06 | Hong Kong|
| Poly-ICLC                                 | Respiratory viruses<sup>a</sup> | Phase 1 | Completed| Mar 08 | Dec 09 | USA      |
| Streptokinase                             | SARS, ARDS      | Phase 3 | Completed| Feb 16 | Jan 18 |          |
| Glucocorticoid (methylprednisolone) therapy | Coronavirus infections<sup>b</sup> | Phase 2, Phase 3 | Unknown  | Jan 20 (Est.) | Dec 20 (Est.) | China |

**Table 5.** Drugs targeting SARS that are registered for the U.S. Food and Drug Administration (USFDA) clinical trials. <sup>a</sup>This covers unknown respiratory viruses. <sup>b</sup>This includes the COVID-19. Est. = estimated.
Figure 1. Virtual screening results for the SARS-CoV-2 3CL\textsuperscript{pro} protease. Docking of representative drugs into the active sites of A chain (A, B, C) and that of B chain (D, E, F). The catalytic residue surfaces are coloured in yellow. Atom colours of drug: C: cyan; O: red; N: blue; H: white; S: yellow; only polar hydrogens are shown. Prepared with PyMOL.
Prediction of the SARS-CoV-2 (2019-nCoV) 3C-like protease (3CL\textsuperscript{pro}) structure: virtual screening reveals velpatasvir, ledipasvir, and other drug repurposing candidates

Yu Wai Chen\textsuperscript{1*}, Chin-Pang Bennu Yiu\textsuperscript{2}, Kwok-Yin Wong\textsuperscript{1*}

\textsuperscript{1} Department of Applied Biology and Chemical Biology and the State Key Laboratory of Chemical Biology and Drug Discovery, The Hong Kong Polytechnic University, Hunghom, Hong Kong.

\textsuperscript{2} Flat E, 11/F, Block 1, La Costa, Ma On Shan, Hong Kong.

*Corresponding email: yu-wai.chen@polyu.edu.hk ; kwok-yin.wong@polyu.edu.hk

ORCID
YWC: 0000-0001-7833-7533
CPBY: 0000-0003-3965-3992
KYW: 0000-0003-4984-7109

Keywords: COVID-19, 2019-nCoV, 3C-like protease, drug repurpose, antiviral, coronavirus, virtual screening, molecular modelling, Hepatitis C virus, HCV, ledipasvir, velpatasvir
Table S1. Sequence homology of the 3CL<sup>pro</sup> cleavage junctions of PP1AB between SARS-CoV-2 and SARS-CoV. The different residues are highlighted. ID: identical residues out of shown 18-residues. NSP: non-structural proteins. Substrate residue positions P<sub>1</sub>' to P<sub>3</sub>' are represented by P<sub>1</sub> to P<sub>3</sub> (underlined).
| A Chain                                                                 |
|------------------------------------------------------------------------|
| Top scorers                | B.E. | Hits |
| teniposide                | -10.4| 30   |
| etoposide                 | -10.3| 32   |
| orvepitant                | -9.8 | 1    |
| dihydroergocristine       | -9.8 | 6    |
| R428                      | -9.7 | 2    |
| MK-3207                   | -9.6 | 4    |
| tadalafil                 | -9.6 | 4    |
| bolazine                  | -9.5 | 1    |
| tivantinib                | -9.5 | 1    |

(A,B model scorers)

ditercalinium              | -9.4 | 1    |
| venetoclax                | -9.2 | 1    |
| hesperidin                | -9.1 | 38   |
| UK-432097                 | -9.1 | 2    |
| lumacaftor                | -8.3 | 1    |
| eluxadoline               | -8.3 | 1    |
| velpatasvir               | -8.1 | 3    |
| diosmin                   | -8.1 | 1    |
| etoposide-phosphate       | -8.0 | 19   |
| irinotecan                | -8.0 | 1    |
| ledipasvir                | 0    |      |

Table S2. The results of virtual screening of drugs on the active site of SARS-CoV-2 3CL\textsuperscript{pro} crystal structure. The top scorers of the crystal structure are listed first, then the equivalent top scorers found in the predicted model chains A and B listed at the lower half. The top scorers which are unique to the crystal structure are coloured in cyan. B.E.: AutoDock Vina binding energy in kcal mol\textsuperscript{-1}. The number of hits of a drug is the times it appears among all results within a screen regardless of rank, only the binding energy of the top-ranking hit was shown.

Docking of drugs into the active sites of SARS-CoV-2 3CL\textsuperscript{pro} crystal structure. (A) Teniposide, the top scorer, and (B) velpatasvir. The catalytic residue surfaces are coloured in yellow. Atom colours of drug: C: cyan; O: red; N: blue; H: white; S: yellow; only polar hydrogens are shown. Prepared with PyMOL.
Figure S1. The structural model of the SARS-CoV-2 3CL\textsuperscript{pro} protease. The variant residues with reference to SARS-CoV 3CL\textsuperscript{pro} are coloured in cyan and labelled. The residues playing critical roles in catalysis and substrate binding are in yellow. Those involved in dimerisation are in pale green. The two views are related by an approximate 90° rotation along the vertical axis.
Comparison with the crystal structure (PDB ID 6lu7)

The overall structural differences in terms of average rmsd of Cα atomic positions (residues 1-300) is 1.2Å for A chain and 0.8Å for B chain. While both chains of the model are similar to the crystal structure, the B chain is almost identical within experimental errors. At the active site, the catalytic dyad and the substrate binding residues have an average all-atom rmsd for A and B chains of 1.4Å and 1.5Å, respectively (Figure 2). These values indicate that the active sites of the two proteases are well preserved.

**Figure S2.** Structural comparison of the model of SARS-CoV-2 3CL\textsubscript{pro} protease with its crystal structure. The model is coloured in orange (chain A) and pale green (chain B); the crystal structure is in grey. The residues playing critical roles in catalysis and substrate binding are shown in sticks. The overall structures are shown as transparent secondary-structure cartoons.
