Prediction of the 2019-nCoV 3C-like protease ($3CL^{pro}$) structure: virtual screening reveals velpatasvir, ledipasvir, and other drug repurposing candidates

Yu Wai Chen[1]*, Chin-Pang Bennu Yiu[2], Kwok-Yin Wong[1]*

[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.

[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

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

Abstract

We prepared the three-dimensional model of the 2019-nCoV 3C-like protease ($3CL^{pro}$) 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 $3CL^{pro}$ 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.

Introduction

On 7 January 2020, a new coronavirus, 2019-nCoV, was implicated in an alarming outbreak of a pneumonia-like illness originating from Wuhan City, Hubei, China. Human-to-human transmission was first confirmed in Guangdong, China [1]. On 6 February 2020, there are more than 28,000 confirmed cases reported globally, with 565 deaths — the World Health Organisation has declared this a global public health emergency. 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 (PL^{pro} and $3CL^{pro}$) that are vital to the virus' replication: they cleave the two translated polyproteins (PP1A and PP1AB) into individual functional components. 3-chymotrypsin-like protease ($3CL^{pro}$ aka main protease, M^{pro}) 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 $3CL^{pro}$ of the 2019-nCoV 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 2019-nCoV 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.nim.nih.gov/ /Blast.cgi?PAGE=Proteins).

Preparation of structural model

The highest-resolution apo-enzyme structure of SARS-CoV $3CL^{pro}$ (PDBID: 2duc) [2] was employed as the template. The variant residues were "mutated" *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 2019-nCoV is aligned with that of SARS-CoV.

The overall amino-acid sequence identity is very high (86%). The conservation is noticeable at the polyprotein cleavage sites. All the eleven $3CL^{pro}$ sites [2] are highly conserved or identical (Table S1), inferring that their respective proteases have very similar specificities. The $3CL^{pro}$ sequence of 2019-nCoV has only 12 out of 306 residues different from that of SARS-CoV (identity = 96%).

Conserved sequence identity among 2019-nCoV

We compared the polyprotein PP1AB and the $3CL^{pro}$ 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 2019-nCoV $3CL^{pro}$ sequences and all their cleavage junctions on their polyproteins are 100% conserved.

3D model of the 2019-CoV $3CL^{pro}$

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 2019-nCoV $3CL^{pro}$ by molecular modelling (Figure S1), which will be immediately useful for *in silico* development of targeted treatment.

Virtual screening for readily-available drugs

The top 10 or 11 (with a binding energy cut-off) hits of chains A and B (Table 2) were examined visually in PyMOL [6] 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 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 2019-nCoV based on the similarity between the replication mechanisms of the HCV and the coronavirus [7]. 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. These direct-acting antiviral drugs are also associated with very minimal side effects and are conveniently orally administered (Table 3).

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 3). 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 3). 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 3). Not a lot are disclosed about MK-3207 and R428.

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 S3), 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 $3CL^{pro}$ enzyme have been proposed [8,9], 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 [10]. However, some scientists did express skepticism [11]. By analogy, these compounds were speculated to act on SARS-CoV $3CL^{pro}$ specifically but there is as yet no crystal structure to support that, although docking studies were carried out to propose various binding modes [12-15]. The IC_{50} value of lopinavir is 50 μM (K_i = 14 μM) and that for ritonavir cannot be established [16]. Although this is far from a cure, based on our results that the two CoV $3CL^{pro}$ 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 $3CL^{pro}$, an earlier screen produced 27 candidates that could be repurposed against both SARS-CoV and MERS-CoV [17]. In addition, the other coronaviral proteins could be targeted for screening. Treatment of the 2019-nCoV with remdesivir (a repurposed drug in development targeting the RdRp) showing improved clinical outcome has just been reported [18].

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 $3CL^{pro}$ reported or announced [12,19-25]. Our methods did not overlap and we share no common results with these studies.

Availability of Materials

The model of 2019-nCoV $3CL^{pro}$ (.pdb) and the full lists of screening results of both active sites (.pdbqt) are available as supplementary materials or upon request from the authors.

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 The PyMOL molecular graphics system (Schrödinger, LLC).

7 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).

8 Pillaiyar, T., Manickam, M., Namasivayam, V., Hayashi, Y. & Jung, S. H. 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).

9 Kuo, C.-J. & Liang, P.-H. Characterization and inhibition of the main protease of severe acute respiratory syndrome coronavirus. *ChemBioEng Reviews* **2**, 118-132 (2015).

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

11 Stockman, L. J., Bellamy, R. & Garner, P. SARS: systematic review of treatment effects. *PLoS Med.* **3**, e343 (2006).

12 Gruber, C. & Steinkellner, G. *Wuhan coronavirus 2019-nCoV - what we can find out on a structural bioinformatics level* https://innophore.com/2019-ncov/ (2020).

13 Dayer, M. R., Taleb-Gassabi, S. & Dayer, M. S. Lopinavir; a potent drug against coronavirus infection: insight from molecular docking study. *Arch. Clin. Infect. Dis.* **12**, e13823 (2017).

14 Nukoolkarn, V., Lee, V. S., Malaisree, M., Aruksakulwong, O. & Hannongbua, S. Molecular dynamic simulations analysis of ritonavir and lopinavir as SARS-CoV $3CL^{pro}$ inhibitors. *J. Theor. Biol.* **254**, 861-867 (2008).

15 Zhang, X. W. & Yap, Y. L. 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).
16 Wu, C. Y. *et al.* Small molecules targeting severe acute respiratory syndrome human coronavirus. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 10012-10017 (2004).
17 Dyall, J. *et al.* Repurposing of clinically developed drugs for treatment of Middle East respiratory syndrome coronavirus infection. *Antimicrob. Agents Chemother.* **58**, 4885-4893 (2014).
18 Holshue, M. L. *et al.* First case of 2019 novel coronavirus in the United States. *N. Engl. J. Med.* 10.1056/NEJMoa2001191 (2020).
19 Xu, Z. *et al.* 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).
20 Liu, X. & Wang, X.-j. Potential inhibitors for 2019-nCoV coronavirus M protease from clinically approved medicines. *bioRxiv* 10.1101/2020.01.29.924100 (2020).
21 Stoermer, M. J. Homology models of Wuhan coronavirus 3CL^{pro} protease. *ChemRxiv* (2020).
22 Global Health Drug Discovery Institute. *Targeting 2019-nCoV Portal* https://ghddi-ailab.github.io/Targeting2019-nCoV/nCov_Structural_Modeling/ (2020).
23 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).
24 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).
25 Li, Y. *et al.* Therapeutic drugs targeting 2019-nCoV main protease by high-throughput screening. *bioRxiv* 10.1101/2020.01.28.922922 (2020).
26 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).
27 Hsu, M. F. *et al.* Mechanism of the maturation process of SARS-CoV 3CL protease. *J. Biol. Chem.* **280**, 31257-31266 (2005).
28 Barrila, J., Bacha, U. & Freire, E. Long-range cooperative interactions modulate dimerization in SARS 3CL^{pro}. *Biochemistry* **45**, 14908-14916 (2006).
29 Barrila, J., Gabelli, S. B., Bacha, U., Amzel, L. M. & Freire, E. Mutation of Asn28 disrupts the dimerization and enzymatic activity of SARS 3CL^{pro}. *Biochemistry* **49**, 4308-4317 (2010).

30 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).

31 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).

32 Lovell, S. C., Word, J. M., Richardson, J. S. & Richardson, D. C. The penultimate rotamer library. *Proteins* **40**, 389-408 (2000).

Tables

Function	Residue Number	Reference
Catalytic	41, 145	[26]
Substrate binding	41, 49, 143-144, 163-167, 187-192	[2,27]
Dimerisation	10, 11, 14, 28, 139, 140, 147, 298	[28-31]
2019-nCoV variants	35, 46, 65, 86, 88, 94, 134, 180, 202, 267, 285, 286	This work

Table 1. SARS-CoV $3CL^{pro}$ important residues and 2019-nCoV variant residues.

A Chain		
A Top scorers	B.E.	Hits
diosmin	-10.1	1
hesperidin	-10.1	38
MK-3207	-10.1	4
venetoclax	-10.0	1
dihydroergocristine	-9.8	6
bolazine	-9.8	1
R428	-9.8	2
ditercalinium	-9.8	1
etoposide-phosphate	-9.8	21
(B Top scorers)		
teniposide	-9.7	34
etoposide	-9.7	32
UK-432097	-9.6	2
irinotecan	-9.5	1
lumacaftor	-8.9	1
velpatasvir	-8.5	2
eluxadoline	-8.0	1
ledipasvir		0

B Chain		
B Top scorers	B.E.	Hits
etoposide	-8.7	32
R428	-8.6	2
MK-3207	-8.6	4
teniposide	-8.5	34
UK-432097	-8.5	2
eluxadoline	-8.4	1
venetoclax	-8.4	1
ledipasvir	-8.4	1
irinotecan	-8.4	1
lumacaftor	-8.4	1
velpatasvir	-8.4	5
(A Top scorers)		
hesperidin	-8.3	38
etoposide-phosphate	-8.3	21
bolazine	-8.3	1
dihydroergocristine	-8.1	6
diosmin	-7.9	1
ditercalinium	-7.7	1

Table 2. The results of virtual screening of drugs on the active sites of 2019-nCoV 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.

Compound	Possible side effects (adverse reactions)	Admin.
diosmin [a,b]	mild gastrointestinal disorders; skin irritations; nausea; heart arrhythmias	topical; oral
hesperidin [a,d]	stomach pain and upset; diarrhea; headache	oral
MK-3207 [c]	no information	oral
venetoclax [a,b]	neutropenia; nausea; anaemia, diarrhea; upper respiratory tract infection	oral
dihydroergocristine [a]	no information	oral
bolazine [b]	no information	intramuscular
R428 [b]	no information	oral
ditercalinium	no information	no info
etoposide [a,b]	alopecia; constipation; diarrhea; nausea; vomiting; secondary malignancies	intravenous
teniposide [a,b]	gastrointestinal toxicity; hypersensitivity reactions; reversible alopecia	intravenous
UK-432097 [c]	no information	inhaled
irinotecan [a,b]	gastrointestinal complication	intravenous
lumacaftor [a]	dyspnea; nasopharyngitis; nausea; diarrhea; upper respiratory tract infection	oral
velpatasvir [a,b]	headache; fatigue; nausea	oral
eluxadoline [a,b]	constipation; nausea; fatigue, bronchitis, viral gastroenteritis; pancreatitis	oral
ledipasvir [a]	fatigue; headache	oral

Table 3. Possible side effects and routes of administration of the drugs identified from virtual screening for 2019-nCoV 3CL^{pro}. Sources of information: [a] DrugBank.ca (main), [b] Wikipedia.org, [c] clinicaltrials.gov and [d] WebMD.com.

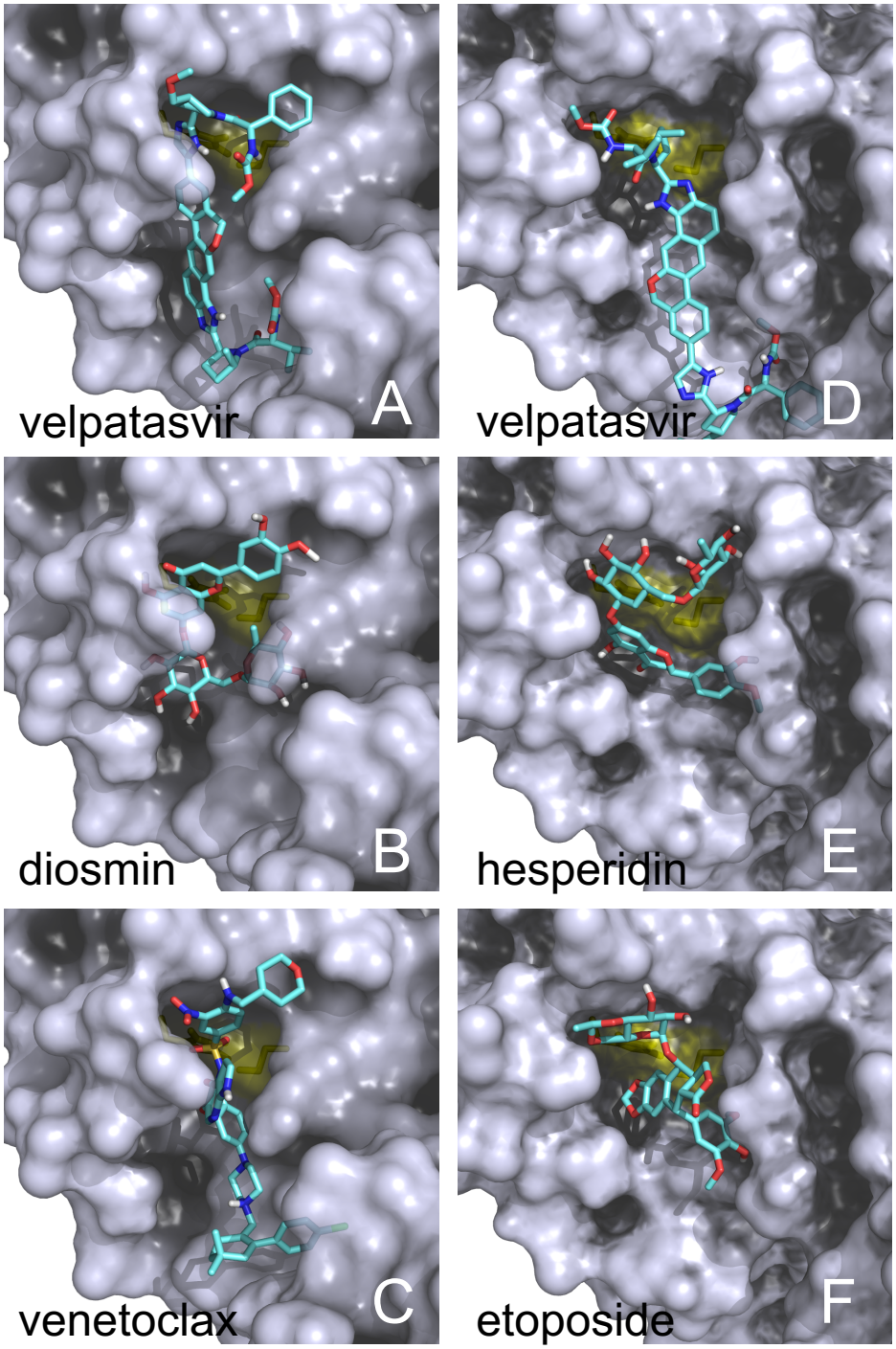

Figure 1. Virtual screening results for 2019-nCoV 3CL^{pro} protease. Docking of representative drugs into the active sites of A chain (A, B, C) and that of B chain (D, E, F). The catalytic residue surfaces are coloured in yellow. Atom colours of drug: C: cyan; O: red; N: blue; H: white; S: yellow; only polar hydrogens are shown. Prepared with PyMOL.