

Potential inhibitors for targeting Mpro and Spike of SARS-CoV-2 based on sequence and structural pharmacology analysis

Chuanjun SHU^{1#}, Xuan HUANG^{2#}, Ting HUANG³, Li CHEN², Bing YAO^{2*}, Jianwei ZHOU^{4*}, Cheng DENG^{5*}

¹ Department of Bioinformatics, School of Biomedical Engineering and Informatics, Nanjing Medical University, Nanjing, 211166, China.

² Reproductive Medical Center, Jinling Hospital Affiliated to The Medical School of Nanjing University, Nanjing, 210002, China.

³ Department of Pharmacy, Xuzhou Maternity and Child Health Care Hospital, Xuzhou, 221000, China.

⁴ Department of Molecular Cell Biology & Toxicology, Center for Global Health, School of Public Health, Nanjing Medical University, Nanjing 211166, China.

⁵ Jiangsu Key Laboratory for Biodiversity and Biotechnology, College of Life Sciences, Nanjing Normal University, Nanjing, 210023, China.

#Contributed equally.

*Correspondence: dengcheng@njnu.edu.cn; jwzhou@njmu.edu.cn; yaobing@nju.edu.cn

<https://doi.org/10.37175/stemedicine.v1i2.41>

ABSTRACT

The SARS-CoV-2 outbreak has spread rapidly and widely since December 2019, and the effective drugs are urgently needed. The two key proteins, Mpro and Spike, are attractive therapy targets for developing drugs against SARS-CoV-2 infection. In this study, we searched for the potential inhibitors targeting Mpro and Spike based on protein sequences and structural pharmacological analysis. We found that both Mpro and Spike of SARS-CoV-2 were homologous with bat SARS-like-CoV. SARS-CoV-2 Mpro showed high conservation (sequence similarities >99%), and the existing few point mutants in different patients from diverse cities suggested that SARS-CoV-2 probably underwent adaptive evolution when the virus infection transmitted from Wuhan patients to other non-Wuhan patients. Moreover, some inhibitors for SARS-CoV Mpro could probably inhibit the activity of SARS-CoV-2 Mpro, because they do not target conserved mutated sites of SARS-CoV-2 Mpro, such as SDJ, ACE-THR-VAL-ALC-HIS-H, B4Z inhibitor, Beclabuvir, Saquinavir, and Lopinavir. In contrast, Spike of SARS-CoV-2 had more mutations and some mutant sites were distributed in the interaction domain between Spike and ACE2. A new peptide FRKSNLKPFERDISTEIYQAGSTPC, based on interactions between Spike and ACE2, could be a potential drug to treat SARS-CoV-2 patients. In summary, our study provided potential new inhibitors for targeting Mpro and Spike in SARS-CoV-2 virus-infected patients based on sequence and structural pharmacology analysis.

Keywords: SARS-CoV-2 · Mpro · Spike · Structural pharmacology analysis · New peptide inhibitor

Introduction

In December 2019, a pneumonia associated with the 2019 novel coronavirus (SARS-CoV-2) occurred in Wuhan, Hubei Province, China (1). And the new coronavirus pneumonia (NCP) named as COVID-19 has spread rapidly over many countries around the world (2, 3). As of 30 March 2020, >730,000 cases have been confirmed and the number of patients is still increasing, with an estimated mortality risk of ~ 4.6% (3, 4).

However, the source of the virus, the effective drugs, and the pathogenesis are still not clear (5, 6).

SARS-CoV-2 has a single-stranded RNA, with distinct clade from the β -coronaviruses associated with the human severe acute respiratory syndrome (SARS) and was classified in the β -coronavirus 2b lineage (7, 8). Similar to SARS, SARS-CoV-2 genome encodes non-structural proteins, structural proteins and accessory proteins (7). Non-structural protein, such as 3-chymotrypsin-like protease (3CLpro, also known as main protease Mpro), is one of the key enzymes for the viral life cycle (9). Structural protein-Spike protein, responsible for viral entry, binds to the cellular receptor angiotensin-converting enzyme 2 (ACE2) and mediates the fusion between the viral and cellular membranes (10). There are two regions

Received: Mar 30, 2020; Accepted: Mar 31, 2020.

© The Author(s). 2019 This is an **Open Access** article distributed under the terms of the Creative Commons License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and reproduction in any medium or format, provided the original work is properly cited.

(S1 and S2) in the Spike protein (11). In the S1 region, there is a receptor binding domain (Spike-RBD) that interacts with ACE2 (12). The functional importance of the Mpro and Spike in the viral life cycle has attracted a lot of interests for developing drugs against SARS-CoV-2.

Many scientists have envisaged that vaccines, monoclonal antibodies, peptides, interferon therapies and small-molecule drugs might be used to control and prevent emerging infections (13-15). However, there is no evidence to support specific drug treatment against the NCP in suspected or confirmed cases. Hence, our objectives were to search for potential drugs based on sequence and structural pharmacological analysis from Mpro and Spike of SARS-CoV-2.

Materials and Methods

Data preparation

Whole genome sequences of SARS-CoV-2 were downloaded from CNCB/BIG (<https://bigd.big.ac.cn/ncov>) and NCBI. According to gene locations, the nucleotide sequences of Mpro, Spike and Spike-S1 of SARS-CoV-2 and their corresponding amino acids were acquired by BioEdit software (16). To explore the origin of proteins, each of Mpro, Spike and Spike-S1 in SARS-CoV-2 was employed in BLAST search. The 100 sequences with the highest similarity to Mpro/Spike/Spike-S1 of SARS-CoV-2 were downloaded from the BLAST results, redundant sequences were deleted (17). Subsequently, 93 Mpro sequences of SARS-CoV-2 and 87 Mpro sequences of SARS/SARS-like, and 82 Spike/Spike-S1 sequences of SARS-CoV-2 and 98 Spike/Spike-S1 sequences of SARS/SARS-like were obtained.

Three-dimensional (3D) spatial structure of Mpro of SARS-CoV-2 was downloaded from the PDB database (18). And 72 Mpro structures in the PDB database were download for comparing the Mpro structure of SARS-CoV-2 with other coronavirus. Meanwhile, structures of Mpro-ligand complex and Spike-ACE2 complex for SARS-CoV were downloaded from the PDB database (PDB ID: 5i08) (11).

Genetic and phylogenetic analysis

Weblogo was implemented to find the conserved sites/area in Mpro/Spike/Spike-S1 (19). The sequences of proteins and the similarities between sequences were aligned by the BioEdit software. Unrooted tree topology based on multiple alignments of amino acids was established with the neighbor-joining method in MEGA 6.06 (20). Consistency of branching was tested using a bootstrap analysis with 500 resamplings of the data in MEGA 6.06.

Structural pharmacology analysis

The I-TASSER (Iterative Threading Assembly Refinement) algorithm was utilized to predict the structures of Spike and Spike-S1 of SARS-CoV-2 (21). The RMSD (root-mean-square deviation) between two

structures was computed by the Rosetta software. Physical and chemical parameters for a protein were predicted by the ProtParam tool (<https://web.expasy.org/protparam/>) (22).

To investigate differences in the electrostatic properties between proteins, adaptive Poisson-Boltzmann solver (APBS) and PDB2PQR were applied to each protein (http://nbc-222.ucsd.edu/pdb2pqr_2.1.1/) (23). The pqr file of each structure was generated using the PDB2PQR program. The dx file of each structure was generated by utilizing APBS. The pqr file and dx file were then uploaded in VMD to render the molecular surface electrostatic potential map.

The largest possible binding pocket of these proteins, i.e., Mpro and Spike-S1, was predicted by Discovery Studio 3.0, respectively. These predicted pockets were utilized to construct an initial coarse model of the Mpro-ligand and Spike-S1-ACE2 complexes. Then, structures of complexes were refined by the Rosetta software (RosettaDock and FlexPepDock module), respectively (24). The final structure was obtained based on energy scores. The interactions between proteins and molecular ligands were calculated by Discovery Studio 3.0 (25). Meanwhile, interactions between proteins were computed based on distances between atoms and type of residues. Expression levels for ACE2 in human tissues were obtained from The Genotype-Tissue Expression (26). High quality 3D images of the proteins were drawn by PyMOL (27).

Results

Sequence analysis could be helpful to evaluate the repurposing of existing antiviral agents to treat SARS-CoV-2. Phylogenetic trees were built (neighbor-joining, bootstrap = 500) for three proteins (Mpro, Spike, Spike-S1) based on selected 185 virus sequences obtained from CNCB/BIG (<https://bigd.big.ac.cn/ncov>) and NCBI blast results (**Figure 1A, B and C**, access date 7 March 2020). The results indicated that all of these three proteins of SARS-CoV-2 probably originated from bat SARS-like-CoV, and pangolin as a mammal is probably a potential intermediate host (**Figure 1A, B and C**). According to homologous analysis, Mpro sequences of SARS-CoV-2 had very high conservation (100% identify for 93 Mpro), and they were remarkably close to corresponding proteins of bat SARS-CoV and SARS-CoV (sequence similarities > 95%) (**Figure 1D**). However, Spike and Spike-S1 protein sequences of SARS-CoV-2 already displayed point mutants (SARS-CoV-2 to SARS-CoV-2 sequence similarities > 99%), and all their corresponding sequence similarity values for bat SARS-like-CoV and SARS-CoV are between 0.6 and 0.8 (**Figure 1D, F**). These results indicated that Mpro is much more conserved than Spike proteins in SARS-CoV-2. Compared to SARS-CoV, twelve conserved amino acids mutations, i.e. 35V, 46S, 65N, 86V, 88K, 94A, 134F, 180N, 202V, 267S, 285A, and 286L, were detected in the Mpro sequences of SARS-CoV-2 (**Figure 1E, Figure 2A and Figure 2B**). Meanwhile, six point mutations (46, 65, 86, 88, 134 and 180 sites)

were distributed in the inhibitor binding pocket of SARS-CoV Mpro (**Figure 1F**). It suggested that inhibitors of Mpro for SARS which target these sites probably could not inhibit the activity of Mpro for SARS-CoV-2.

On the other hand, Spike-S1 has a conserved domain

which interacts with mammalian ACE2. Point mutation occurred in this domain may affect the interaction between Spike and ACE2, and then impact the capability of coronavirus entry into mammalian normal cells with ACE2. In 82 sequences of SARS-CoV-2 Spike-S1, there

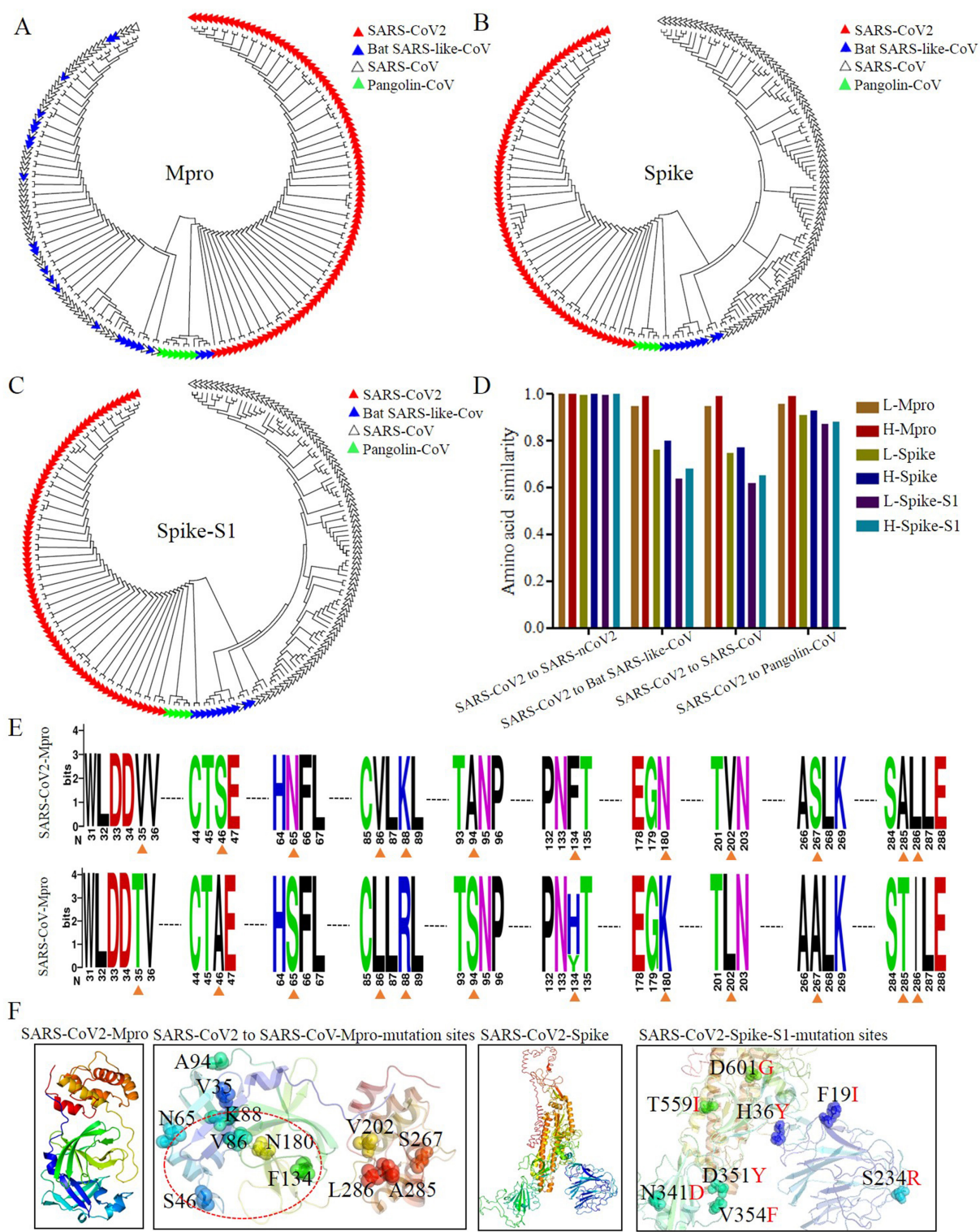


Figure 1. The evolution and conservation for Mpro and Spike of SARS-CoV-2. (A), (B) and (C) represent phylogenetic trees of Mpro, Spike, and Spike RBD. (D) Amino acid sequence similarity among SARS-CoV-2, bat SARS-like-CoV, and SARS-CoV. L- and H- represent the lowest and highest sequence similarity between sequences in two types of coronavirus, respectively. (E) Sequence logo of Mpro for SARS-CoV-2 and SARS-CoV. (F) Differential conserved amino acids of Mpro/Spike between SARS-CoV-2 and SARS-CoV.

were 11 sequences with amino acid point mutation (11/82, 13%) and up to three mutation sites for each mutation sequence (**Figure 1D**). Eight key mutation sites (F19I, H36Y, S234R, N341D, D351Y, V354F, T559I, and D601G) located in different regions (**Figure 1F**). F19I and T559I were only distributed in SARS-CoV-2 that were from Wuhan, China. N341D and D351Y were only presented in SARS-CoV-2 that were from Shenzhen, China. H36Y, S234R, V354F, and D601G only occurred in Spike-RDB domains of SARS-CoV-2 that were from Guangdong (China), Australia, France, and Germany, respectively. These results indicated SARS-CoV-2

probably underwent adaptive evolution in the human body.

It was well-known that protein sequence determines its structure, which in turn decides its biological function, such as pharmacological properties. According to sequence analysis, we tested the hypothesis that the Mpro structure of SARS-CoV-2 was very close to that of SARS-CoV. Then, Mpro of SARS-CoV-2 (PDB ID: 6LU7, <https://www.rcsb.org/structure/6LU7>) was utilized to compare with other Mpro proteins in the PDB database. We found that Mpro of SARS-CoV (PDB ID: 5c5o) was very close to Mpro of SARS-CoV-2 (RMSD = 0.41Å)

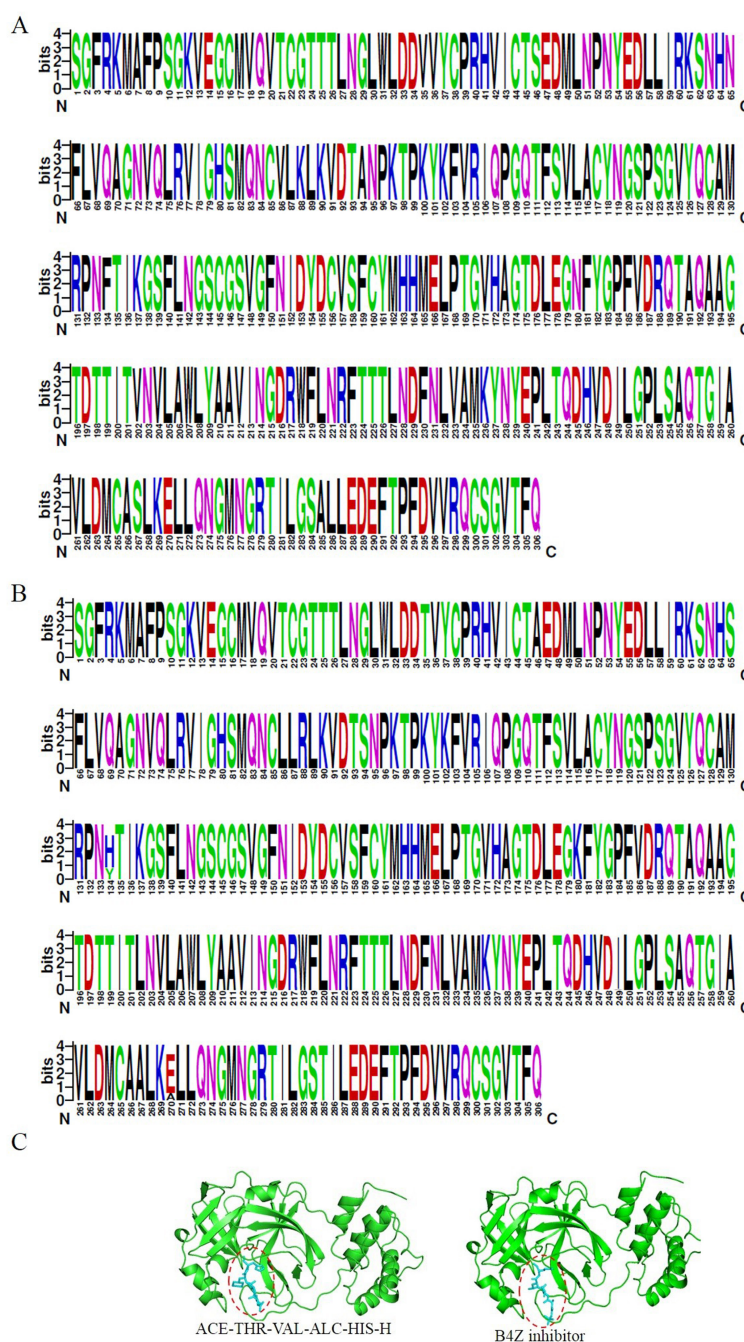


Figure 2. Sequence logos and inhibitors for Mpro in SARS-CoV-2 and SARS-CoV. (A) The sequence logo for Mpro in SARS-CoV-2; (B) The sequence logo for Mpro in SARS-CoV; (C) The complex for peptide and Mpro-SARS-CoV-2.

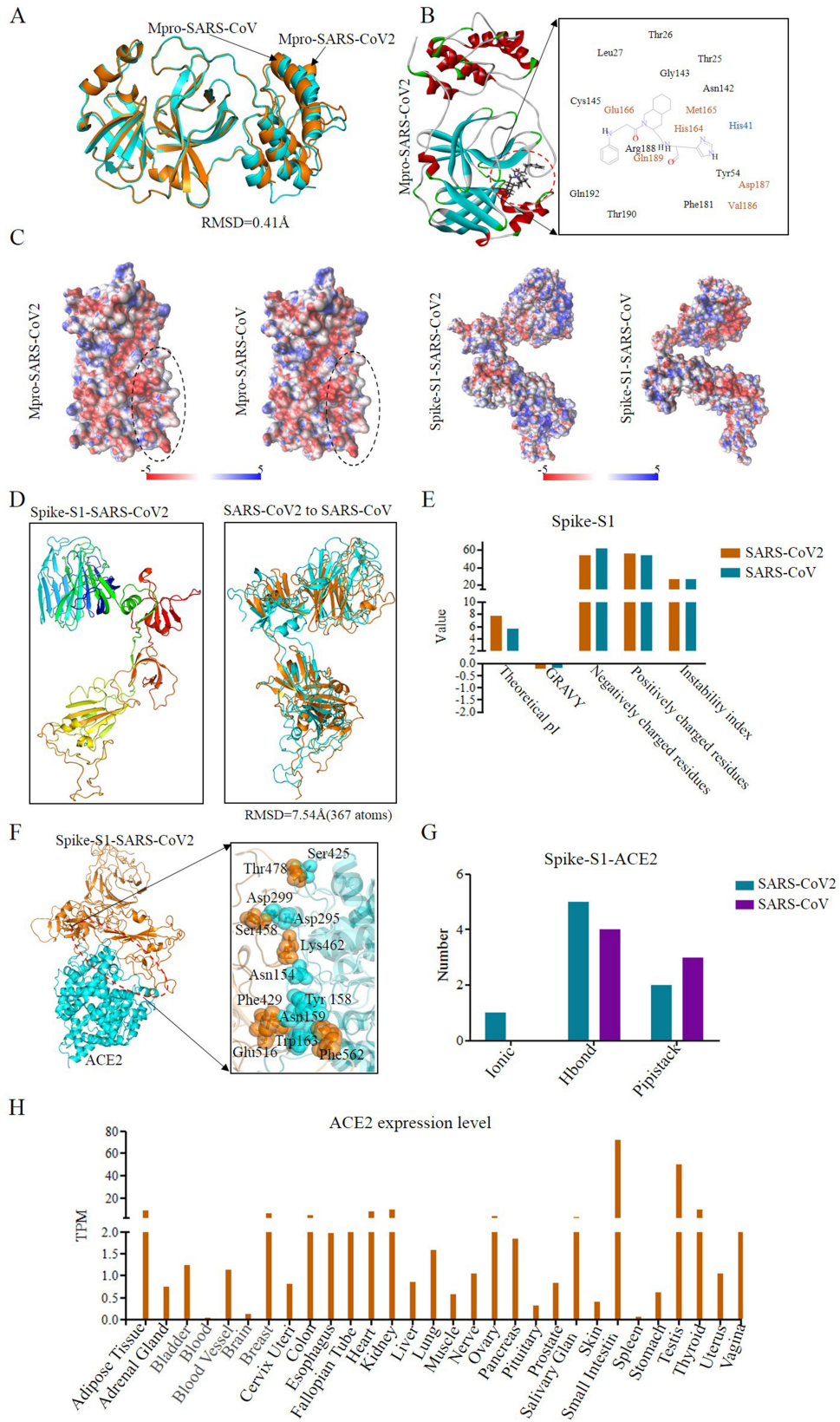


Figure 3. Structural pharmacology analysis of Mpro and Spike. (A) Structural alignment of SARS-CoV-2 Mpro and corresponding protein of SARS-CoV (PDB ID: 5c5o). (B) A ligand (SDJ) of SARS-CoV interacts with Mpro of SARS-CoV-2. Orange, cyan, and black represent h bond, Pipistack, and VDW. (C) Surface electrostatic potential values of Mpro and Spike-S1 for SARS-CoV-2 and SARS-CoV. (D) Structure comparison of SARS-CoV-2 Spike-S1 and SARS-CoV Spike-S1 (orange cartoon). (E) Physical and chemical parameters, including theoretical pI (isoelectric point), GRAVY (grand average of hydropathicity), negatively/ positively, and instability index of Spike-S1 for SARS-CoV-2 and SARS-CoV. (F) The complex of Spike-S1 for SARS-CoV-2 and ACE2, and the interaction sites between two proteins. (G) The interaction force type of Spike-S1 and ACE2 in SARS-CoV-2 and SARS-CoV. (H) ACE2 expression levels in different human organs.

(**Figure 3A**). This result indicated that the inhibitors of this protein probably also inhibit the activity of Mpro of SARS-CoV-2. After screening, a unique ligand for 5c5o, i.e. SDJ (phenyl- β -alanyl (S, R)-N-decalin type inhibitor: (2S)-3-(1H-imidazol-5-yl)-2-((3S, 4aR, 8aS)-2-(N-phenyl- β -alanyl) decahydroisoquinolin-3-yl] methyl} amino) propanal, was utilized to dock with Mpro of SARS-CoV-2 (**Figure 3B**). The interaction sites between the inhibitor and Mpro of SARS-CoV-2 were almost the same (**Figure 3B**, PDB ID: 5c5o) (9), and there is no conserved amino acid mutations for Mpro. Hence, this inhibitor of SARS-CoV Mpro could act as an effective inhibitor for Mpro of SARS-CoV-2. Furthermore, based on other proteins that were close to Mpro of SARS-CoV-2, we also found that peptide ACE-THR-VAL-ALC-HIS-H [Biologically Interesting Molecule Reference Dictionary (BIRD), ID: PRD_000815] and B4Z inhibitor (BIRD ID: PRD_000910) could probably inhibited SARS-CoV-2 infection (**Figure 2C**). Meanwhile, electrostatic potential values for the surface of Mpro proteins of SARS-CoV and SARS-CoV-2 were almost the same as well (**Figure 3C**). These results indicated that inhibitors (molecules and peptides, such as, Beclabuvir, Saquinavir, and Lopinavir) for Mpro of SARS-CoV probably could be utilized to inhibit the activity of Mpro in SARS-CoV-2, if these inhibitors do not target conserved mutation amino acids of Mpro SARS-CoV-2.

For drug design based on the interaction between Spike-S1 and ACE2, we obtained protein structures of Spike and Spike-S1 by utilizing the I-TASSER algorithm. Compared to SARS-CoV, five factors were different from Spike-S1 proteins of SARS-CoV-2: 1) surface electrostatic potential values (**Figure 3C**); 2) 3D spatial structure (RMSD = 7.54 Å, aligned based on 367 atoms) (**Figure 3D**, PDB ID: 5i08); 3) physical and chemical parameters (theoretical pI, GRAVY, negatively/ positively, and instability index) (**Figure 3E**); 4) interaction sites between Spike-S1 and ACE2 (**Figure 3F**, PDB ID: 5i08) (11); 5) interaction force type between Spike-S1 and ACE2 (**Figure 3G**, PDB ID: 5i08). These factors play important roles in designing inhibitors for ACE2. Therefore, some inhibitors of ACE2, effectively inhibit the entry of SARS-CoV into normal cells, might not perform well for SARS-CoV-2. However, peptides for SARS-CoV-2 could be designed based on predicted structure of the Spike-S1-ACE2 complex. For example, such as a predicted peptide, FRKSNLKPFERDISTEIYQAGSTPC, could interact with ACE2 with one ionic bond and four H bonds according to predicted Spike-S1-ACE2 complex (**Figure 3F**). Meanwhile, according to ACE2 expression level in human tissues (**Figure 3H**), we should pay more attention to the functional changes of intestinal tract, testis, liver, lung and kidney, during clinical diagnosis and treatment for SARS-CoV-2 patients.

Discussion

Our results showed that both Mpro and Spike of SARS-CoV-2 were homologous with those of bat SARS-

like-CoV. Although protein sequence and structure for SARS-CoV-2 Mpro were conserved, Spike of SARS-CoV-2 had many mutations and some mutant sites were distributed in the Spike-S1. Based on sequence and structural pharmacology analysis, we found that some Mpro inhibitors for SARS-CoV probably also inhibit activity of Mpro if they do not target conserved mutated sites for Mpro of SARS-CoV-2 (**Figure 1**). In addition, we found that SARS-CoV-2 probably underwent adaptive evolution when the virus spread from Wuhan patients to other non-Wuhan patients, which could be helpful for discovering potential drugs for the treatment of NCP cases. Furthermore, a potential peptide, that was predicted based on interactions between Spike-S1 and ACE2, could probably serve as a potential drug, and further study can be conducted for its function test and peptide modification.

Conflict of interest

The authors declare that they have no competing interests.

Acknowledgement

Authors acknowledge the contributions for Mpro structure of Zihe Rao group.

References

1. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med*. 2020 Feb 20;382(8):727-733.
2. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*. 2020 Feb 15;395(10223):497-506.
3. Guan W-j, Ni Z-y, Hu Y, Liang W-h, Ou C-q, He J-x, et al. Clinical characteristics of 2019 novel coronavirus infection in China. *medRxiv*. 2020:2020.02.06.20020974.
4. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. *JAMA*. 2020 Feb 7. [Epub ahead of print]
5. Jin YH, Cai L, Cheng ZS, Cheng H, Deng T, Fan YP, et al. A rapid advice guideline for the diagnosis and treatment of 2019 novel coronavirus (2019-nCoV) infected pneumonia (standard version). *Mil Med Res*. 2020;7(1):4.
6. Zhang Z, Li X, Zhang W, Shi ZL, Zheng Z, Wang T. Clinical features and treatment of 2019-nCoV pneumonia patients in Wuhan: report of a couple cases. *Virol Sin*. 2020 Feb 7. [Epub ahead of print]
7. Wu A, Peng Y, Huang B, Ding X, Wang X, Niu P, et al. Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in China. *Cell Host Microbe*. 2020 Mar 11;27(3):325-328.
8. Woo PC, Lau SK, Chu CM, Chan KH, Tsoi HW, Huang Y, et al. Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. *J Virol*. 2005 Jan; 79(2): 884-895.
9. Shimamoto Y, Hattori Y, Kobayashi K, Teruya K, Sanjoh A, Nakagawa A, et al. Fused-ring structure of decahydroisoquinolin as a novel scaffold for SARS 3CL protease inhibitors. *Bioorg Med Chem*. 2015 Feb 15; 23(4):876-90.
10. Wong, K. S. A 193-amino acid fragment of the SARS coronavirus S protein efficiently binds angiotensin-converting enzyme 2. *J Biol Chem*. 2004 Jan 30;279(5):3197-201.
11. Kirchdoerfer RN, Cottrell CA, Wang N, Pallesen J, Yassine HM,

- Turner HL, et al. Pre-fusion structure of a human coronavirus spike protein. *Nature*. 2016 Mar 3;531(7592):118-21.
12. Wang S, Guo F, Liu K, Wang H, Rao S, Yang P, et al. Endocytosis of the receptor-binding domain of SARS-CoV spike protein together with virus receptor ACE2. *Virus Res*. 2008 Sep;136(1-2):8-15.
13. Richardson P, Griffin I, Tucker C, Smith D, Oechsle O, Phelan A, et al. Baricitinib as potential treatment for 2019-nCoV acute respiratory disease. *Lancet*. 2020 Feb 15;395(10223):e30-e31.
14. Liu W, Morse JS, Lalonde T, Xu S. Learning from the past: possible urgent prevention and treatment options for severe acute respiratory infections caused by 2019-nCoV. *ChemBiochem*. 2020 Mar 2;21(5):730-738.
15. Wang M, Cao R, Zhang L, Yang X, Liu J, Xu M, et al. Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. *Cell Res*. 2020 Mar;30(3):269-271.
16. Helge-Friedrich T. Analysis for free: Comparing programs for sequence analysis. *Brief Bioinform*. 2004 Mar;5(1):82-7.
17. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool (BLAST). *J Mol Biol*. 1990 Oct 5;215(3):403-10.
18. Sussman JL, Lin D, Jiang J, Manning NO, Prilusky J, Ritter O, et al. Protein data bank (PDB): database of three-dimensional structural information of biological macromolecules. *Acta Crystallographica*. 2010;54(6-1):1078-84.
19. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. *Genome Res*. 2004 Jun; 14(6):1188-1190.
20. Lewis PO, Kumar S, Tamura K, Nei M. MEGA: molecular evolutionary genetics analysis, version 1.02. *Systematic Biology*. 1995;44(4).
21. Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc*. 2010 Apr;5(4):725-738.
22. Garg VK, Avashthi H, Tiwari A. MFPPI – multi FASTA ProtParam interface. *Bioinformation*. 2016;12(2): 4-77.
23. Dolinsky TJ, Nielsen JE, Andrew MJ, Baker NA. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res*. 2004 Jul 1; 32(Web Server issue):W665-W667.
24. Rohl CA, Strauss CE, Misura KM, Baker D. Protein structure prediction using Rosetta. *Method Enzymol*. 2004;383:66-93.
25. Gao YD, Huang JF. An extension strategy of Discovery Studio 2.0 for non-bonded interaction energy automatic calculation at the residue level. *Dongwuxue Yanjiu*. 2011 Jun;32(3):262-6. [Article in Chinese]
26. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The genotype-tissue expression (GTEx) project. *Nat Genet*. 2013 Jun;45(6):580-585.
27. Ordog R. PyDeT, a PyMOL plug-in for visualizing geometric concepts around proteins. *Bioinformation*. 2008;2(8):346-347.