Systemic *in silico* screening in drug discovery for Coronavirus Disease (COVID-19) with an online interactive web server

Chi Xu†, Zunhui Ke †, Chuandong Liu3,4†, Zhihao Wang5,6†, Denghui Liu1, Lei Zhang1, Jingning Wang9, Wenjun He†, Zhimeng Xu†, Yanqing Li5, Yanan Yang5, Zhaowei Huang1, Panjing Lv9, Xin Wang5, Dali Han3,4,7,8#, Yan Li9,10#, Nan Qiao1#, Bing Liu5,6,11#

1Laboratory of Health Intelligence, Huawei Technologies Co., Ltd
2Wuhan Children's Hospital, Tongji Medical College, Huazhong University of Science & Technology, Wuhan, China
3Key Laboratory of Genomic and Precision Medicine, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China.
4College of Future Technology, Sino-Danish College, University of Chinese Academy of Sciences, Beijing 100049, China.
5BioBank, The First Affiliated Hospital of Xi'an Jiaotong University, Shaanxi, 710061, China
6MRC Centre for Molecular Bacteriology and Infection, Imperial College London, SW7 2AZ, UK
7Department of Pediatrics, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China.
8China National Center for Bioinformation, Beijing 100101, China.
9Institute for Stem Cell and Regeneration, Chinese Academy of Sciences, Beijing 100101, China.
10Department of Pathogen Biology, School of Basic Medicine, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China.
11Instrument Analysis Centre, of Xi'an Jiaotong University, Shaanxi, China.

†These authors contributed equally
#Correspondence: bliu2018@xjtu.edu.cn; qiaonan3@huawei.com; yanli@hust.edu.cn; handl@big.ac.cn
Graphical abstract

The SARS-CoV-2 viral genome codes a cluster of proteins, each with a unique function in the event of host invasion or viral development. To provide a fast-track solution, we performed virtual screening using 8506 compounds from the DrugBank, targeting some of the druggable viral proteins and human ACE2 receptor. Our eighteen thousand docking results are fully accessible from our interactive web server - Shennong porject (https://shennongproject.com).
Abstract

The emergence of the new coronavirus (nCoV-19) has brought global impact on human health, whilst the interaction between the virus and the host is the foundation of the disease. The viral genome codes a cluster of proteins, each with a unique function in the event of host invasion or viral development. Under current adverse situation, we employ virtual screening tools in searching for drugs which has been already deposited in the DrugBank in attempt to accelerate the drug discovery process. This study provides an initial evaluation of current drug candidates from various reports using our systemic in silico drug screening based on structures of viral proteins and human ACE2 receptor. Besides, we build an interactive online platform for browsing these results with the visual display of small molecule docked on its potential target protein, without installing any specialized structural software. With continuous maintenance and incorporation of data from laboratory works, it may serve not only as the assessment tool for the new drug discovery but also an educational website to meet general interest from the public.
**Introduction**

The notorious coronaviruses, belong to the family *Coronaviridae* and subfamily *Coronavirinae*, are pathological significant to many mammals including human. Just after the millennium, two betacoronaviruses form this group of viruses also named Severe Acute Respiratory Syndrome coronavirus (SARS-CoV) and Middle East Respiratory Syndrome coronavirus (MERS-CoV) have swept the part of world and brought impacts on health and economy in 2003 and 2012, respectively (1). Recently, another member - Severe Acute Respiratory Syndrome coronavirus 2 (SARS-CoV-2) became an unavoidable topic for almost everyone around the globe and the disease it brings (COVID-19) was declared pandemic by the world health organization (WHO) has caused over 148 thousand cases and 5,400 fatalities in 149 countries and territories. The early cases of the diseases emerged as in Wu Han - a Chinese metropolitan with over 11 million population in December 2019, which were diagnosed as cryptogenic pneumonia in several hospitalized patients (2). Since then, it loses control and becomes a world panic during which most countries take stringent measures via tighten board control, people movement *etc.*

The origin of the virus remains undefined, although homology comparison shows that its genome has a 96.3% sequence similarity compared with BatCoV RaTG13 (a coronavirus with bat origin) and 79% compared with SARS-CoV (3, 4). The most common feature shared by coronaviruses is the single-strand, positive-sense RNA genomes which are 26-32 kilobases in length containing 6-12 open reading frames (ORFs) (5). The first ORF takes up to two third length of the whole genome of coronavirus and contains genetic codes for two polyproteins named ppla and pplab, both of which are autoproteolytically cleaved into 15 or 16 nonstructural proteins: nsp1 - nsp16
(nsp1 is absent in deltacoronavirus and gammacoronavirus). Meanwhile the remaining ORFs encode some accessory proteins, including four indispensable structural proteins: spike glycoprotein, small envelope protein, matrix protein and nucleocapsid protein (6). These proteins play different roles at various stages during the viral invasion and viral development, many of which are vital for the survival of the coronaviruses (7-9). The genome of SARS-CoV-2 is comprised of 29,891 nucleotides, which encode the 12 putative ORFs coding for about 28 structural and nonstructural proteins (NCBI Reference Sequence: NC_045512.2).

There are the four dispensable structural proteins coded by the viral genome - Membrane (M), envelope (E), nucleocapsid (N), and spike (S). M protein is a small membrane protein with three transmembrane domains and the most abundant of the four, whose presence is required to form the shape of the virion (10-12). The E protein is a small protein within the virion with the functions like assembling and releasing of the virus (13-15). The N protein only presents in the nucleocapsid handling RNA structure and functions (16, 17). For a successful infection, the virus needs to recognize the host cell via the interaction between its S protein and the host cellular receptors - Angiotensin-Converting Enzyme 2 (ACE2) receptors in human. The S protein has two subunits - S1 which contains a so-called receptor binding domain (RBD) allowing the virus to bind to the peptidase domain (PD) of ACE2 and the S2 subunit which helps the viral particle fuse with host membrane (18-21).

After entering the cell, the virus hijacks the host translational machinery and starts express its own proteins. A polyprotein is then translated via ORF1ab and subsequently cleaved into 16 different
nsp proteins, some of which are better characterized than others (22, 23). For example, nsp1 suppresses the host gene expression by inducing template-dependent endonucleolytic cleavage of host mRNAs and preventing the accumulation of IFN-beta, which may provide a susceptible condition for viral infection and replication in cellular (24-27). Papain-like protease (PLpro), also named nsp3, is the largest multi-domain proteins encoded by the virus. Among a dozen domains of nsp3, ubiquitin-like domain mediating multitudinous viral protein interactions with themselves or host proteins and papain-like domain responsible for releasing nsp1, nsp2 and nsp3 from the polyprotein become the a potential target for antiviral drug exploitation (28-37). Main protease (Mpro), being synonymous with 3C-like protease (3CLpro) or nsp5, is able to cleave the polyprotein at 11 sites, generating at least 10 indispensable nonstructural proteins (6, 38-40). And its importance in the viral development makes it one of the most popular drug targets. Nsp8, a RNA-dependent RNA polymerase (RdRp), is verified to be capable of de novo initiation of RNA which initiate the synthesis of complementary oligonucleotides of <6 residues in a reaction and has been proposed to operate as a primase with the cooperation of nsp7 (41-44). Similarly, nsp12 is the second RdRp of the virus which contains the canonical viral RdRp motifs in its C-terminal part and employs a primer-dependent RNA synthesis mechanism with assistance of primase nsp8 (45, 46). Nsp13 is the viral helicase which has both RNA and DNA duplex-unwinding activities considering natural nucleotides and deoxynucleotides as its substrates (47-49). Nsp16, activated by cofactor nsp10, functioning as 2’-O methyltransferase, exert pivotal roles in capping process, as well as C-terminal of nsp14 acting as N7-methyltransferase (N7Tase) (50-52). Besides, at the present of nsp10, N-terminal of nsp14 serves as exoribonuclease and cooperate with endoribonuclease (nsp15) to insure the accurate cleavage of coronavirus RNA genome in host
As the pandemic affecting our health and lifestyles, there is still no vaccine for the SARS-CoV-2. The priority remains finding drugs for treatment of the infected patients. Considering that the above-mentioned proteins and their importance in alone or synergistically during virus infection and replication, finding drugs to interdict their functions and interactions would stop the viral development and thus spreading.

Drug discovery is a very lengthy process while virtual screening is regarded as the fastest and most accurate method in the early stage of drug design (Fig. 1A). Many studies are based on *in silico* tools have virtually screened small molecule databases and published huge amount of information on new drug discoveries for the coronavirus disease (COVID-19)(59). However, these results are neither based on the approved drugs in the DrugBank nor very user-friendly to scientists outside its niche. Here, we carried out structure-based virtual screening using FDA approved drugs and drugs that currently undergoing the phase 3 clinical trials as the library and constructed an interactive online platform for quick browsing - Shennong (https://www.shennongproject.com:11443/#/home). The advantages of the platform include: searching drug name or protein target name, 3D display of drugs docked on their potential target proteins, a dedicated section for nature products and continuous maintenance. Shennong is a collaborative effort with more data to be incorporated in the pipeline, and possibly the prototype of its kind.
Results

SARS-CoV-2 protein sequence variations compared to SARS and homology modelling

Structure-based virtual screening requires the three-dimensional structure of its protein target and a function to estimate the likelihood of the ligand binding affinity to the protein. To use the best available structures for screening, we listed all 28 putative viral proteins encoded in its genome (Fig. 1B) and removed the small peptides (ORF6, ORF7, ORF10 and nsp11) which are less likely to be druggable. Then we further removed another 10 proteins from the list as there is no structure for either SARS-CoV-2 or SARS. Among the 16 proteins left, S protein and nsp5 of SARS-CoV-2 have structures deposited in the protein data bank (PDB) with PDB ID 6CS2 and 6LU7, respectively. Not surprisingly, two structures show little structural variation compared with their SARS counterparts. The remaining 14 viral proteins share high sequence identities with their SARS counterparts, ranging from 76.60% in nsp3 to 99.84% in nsp13, with nsp4 as the only exception (Fig. 1.C). The high sequence identities ensure the reliabilities of homologous structures prediction using SARS proteins as templates. Nsp4, whose template was using homologous structure of mouse hepatitis virus A59 (61.36% sequence identity to nsp4 of SARS-CoV-2), has no other close homology. Using SWISS-MODEL(60) and structures of SARS proteins and nsp4 of mouse hepatitis virus A59 as templates, we built 14 structural models and followed by molecular dynamics refinement and simulation for optimized protein structures (Fig. 1D).

Screening library and targets
Virtual screening is a technique largely based on its libraries of small molecules and the target sites. DrugBank has a collection of 9591 drug entries including 2037 FDA-approved small molecule drugs, 241 FDA-approved polypeptide drugs, 96 nutraceuticals and over 6000 experimental drugs (61). As repurposing current drugs is the fastest way to meet urgency of COVID-19, we built our library by selecting only FDA-approved drugs and drugs currently in clinical trials in DrugBank. Then we select a list of active sites from structures of the 16 viral proteins and ACE2 protein (PDB ID: 6CS2) as the ligand targets for screening (Table 1). As individual protein has different biological role and a successful drug should be able to block its function by directly acting on the active site or indirectly via conformational change of the structure. For example, drugs screened based on human ACE2 protein and viral S protein are designed to block the interaction between human cell and the virus while those for nsp5 are ought to have effect on preventing its protease activity.

**Docking results overview**

To avoid over interpretation of the results by ourselves, we uploaded the data to our web server for one’s own assessment. The complete set of the docking results (178, 626 in total) are available at our interactive server - [https://shennongproject.com:11443/#/home](https://shennongproject.com:11443/#/home). And we built two heatmaps for drugs have lowest binding energies and natural compounds (some of which do not require a doctor’s prescription), respectively (Fig. 3a and b).

In general, the binding energies are relatively high for the dockings at active sites we chose for nsp1, nsp3 and nsp7. No specific active sites for nsp1 and nsp3 were given during screening due the lack of characterization while the key residues (K7, H36 and N37) of nsp7 at its interface with nsp12 were selected in the screening. It is likely that these sites, either automatic generated or
specified, are not suitable as drug targets, at least not for the candidates in our library. And absences of hydrophobic residues at these sites are the likely explanation for this phenomenon. Meanwhile the binding energies for nsp5 (Mpro), nsp14 and nsp15 are generally low as the surface geometry and hydrophobicity of the active sites make them more druggable (discussed in detail later). Antiviral drugs like Saquinavir, Lopinarvir, Darunavir Nafamostat, Raltegravir Dolutegravir, Bictegravir, Tipranavir, Indinarvir and Montelukast are among the highest scored drugs in our screening (Fig. 3a). In the other hand, natural products have higher binding energy in general although still have the similar preference for nsp14 (Fig. 3a). Proscillaridin extracted plants of the genus *Scilla* and in *Drimia maritima*, which is used for treating congestive heart failure and cardiac arrhythmia, achieved comparable reading as the above antiviral drugs.

A group of chemotherapeutic drugs, including Tivantinib, Lifirafenib, Entrectinib, Nilotinib and Radotinib, should not be neglected either. These tyrosine kinase (or tyrosine kinase receptor) inhibitors are either approved or investigational to be used in therapy of certain hematopathy and metastatic cancers like acute myeloid leukemia (AML), Acute lymphocytic leukemia (ALL) and lung cancers. In our docking results, these drugs are ranked among the top with main protease and exonuclease of SARS-CoV-2, as well as other nonstructural and structural proteins, indicating that they are worthy for further investigations in treatment for coronaviruses.

**Drugs under clinical trials**

Our results coincide with many of current researches for drug development. our website offers detailed docking results for most of them. For example, Remdesivir is a nucleotide analog used for antiviral purpose. Although it was designed as a treatment for Ebola virus disease, it also been
found to show antiviral activity against other single stranded RNA viruses and used in treatment of COVID-19 (62-65). In our screening, it is predicted to bind to nsp14, nsp5 and nsp13 with low binding energy (-8.3 kcal/mol, -7.4 kcal/mol and -7.2 kcal/mol, respectively). Hydrogen bonds were predicted between Gly164, Gln270, Tyr274, Asp303 and the compound. Also, π–π stacking was found between Tyr265 and Remdesivir (Fig. 4a). The strong hydrogen bonding and hydrophobic interactions between the ligand and protein imply it may be a potential inhibitor.

Lopinavir, an anti-HIV drug in the category of protease inhibitor and is another popular drug that has been reported to have strong positive results in a few trials (66-71). It is one of the highest ranked drugs in our screening with a binding energy of -9.3 kcal/mol against nsp15. In our docking, Lopinavir occupies the exonuclease site, interacting with nsp15 via hydrogen bonding, hydrophobic interaction and π–π stacking (Fig. 4b). Thus, Lopinavir may be a potent nsp15 inhibitor based on our result.

Natural products in the screening

We picked two natural products of our interest (Quinine and Doconexent) from the 924 docking results from our screening (https://shennongproject.com:11443/#/naturalProducts). Quinine is a famous anti-malaria drug which was recently repurposed quinine as antiviral against dengue virus infection. It has a binding energy of -7.5 kcal/mol against nsp13, which is comparable to the some of the drugs under clinical trials (Fig. 4c). Its interaction with nsp13 includes hydrogen bonding, hydrophobic interaction and π–π stacking food supplement, making it a potential inhibitor for nsp13. While Doconexent is a mixture of fish oil and primrose oil and used as a high-docosahexaenoic acid (DHA) with minor anti-inflammatory effects. It is ranked at the bottom half against all active sites, likely due to the lack of π–π stacking. However, it has a low
binding energy with nsp14 at -6.7 kcal/mol (Fig. 4d). Although it is undoubtedly a less preferred ligand in our screening, the fact that purchasing DHA or fish oil does not require prescription makes it a potential mild viral inhibitor for self-protection.

**Drug perform well in our screening but not under clinical trial**

A few drugs, including Saquinavir, Beclabuvir, Bictegravir and Dolutegravir are not currently under investigation for treatment of COVID-19 to our knowledge. However, the antiviral mechanisms of these drugs, together with their performance in our screening, make them the drugs we recommended should be tested in treatment of COVID-19.

Among them, Beclabuvir is the only antiviral drug with the purpose for the treatment of HCV infection (72, 73), whilst the rest are drugs for HIV infection. With a low binding energy of -10.4 kcal/mol to nsp5, Beclabuvir is one of the drugs performed the best in the docking. With strong hydrogen bonding, hydrophobic interaction and $\pi$-$\pi$ stacking, it is likely a stronger inhibitor for exonuclease activity inhibitor of nsp15 (Fig. 5a). It is possibly a better nsp15 inhibitor than the popular Lopinavir, at least in our screening.

Saquinavir is an antiretroviral drug used in a cocktail for treating HIV patients (74) and has a binding energy of -9.9 kcal/mol to nsp5 in our screening (Fig. 5b). And Bictegravir and Dolutegravir are integrase inhibitors used in combination with other drug for treatment of HIV infection. They are structurally related as the former is a derivation from the latter (75). And their binding energies to nsp5 are also very similar (-9.5 kcal/mol for Bictegravir and -8.9 kcal/mol) (Fig. 5c and d). Interestingly, all three drugs are in the category of protease inhibitors and has low binding energies against the nsp5 – the main protease of the virus. These underlying similarities makes them worthy of repurposing for potential COVID-19 treatment.
Shennong web server and results reporting

To give users the familiar search engine style experience, we adopted a user-friendly homepage and a graphic interface for viewing the docking results. The web server supports searches by either drug name or protein target name, with additional features like updates for drugs under clinical trials and a tab dedicated for natural compounds. This online platform may not only assist fast and cost-efficient drug discovery but also serve as an educational website for general public.

Discussion

To provide a fast track solution, we performed virtual screening using drugs from the DrugBank, targeting some of the viral proteins and human ACE2 receptor. Our results coincide with some of the most popular drugs currently under clinical trials as well as provide some potential new candidates. The drugs on the top of our list are relate anti-HIV drugs, anti-HCV drugs, influenza virus antagonists, chemotherapeutic drugs and asthma drugs.

Anti-HIV drug are popularized across our docking list and can be divided into two groups: enzyme inhibitors, generally on the top and dideoxynucleoside (or nucleoside) analogs, generally at the bottom of our list. Nucleoside reverse transcriptase inhibitors (NRTIs), including Emtricitabine and Tenofovir, may not work well in coronaviruses attributed to that coronavirus is positive-sense single-stranded RNA virus which lacks nucleoside reverse transcriptase, and this is also reflected in our docking as most of them ranked at the bottom with low binding affinity. Among the enzyme inhibitors of HIV in our docking results, Lopinavir, Dolutegravir and Raltegravir exhibit strong binding affinity with multiple target sites, especially at the catalytic sites of main protease and exonuclease, suggesting their great potential of clinical drugs in therapies of
COID-19. For example, Saquinavir, acting on HIV protease cleavage site, is a highly specific inhibitor of HIV-1 and HIV-2 proteases. Interestingly, it shows a strong affinity with main protease of SARS-CoV-2, which is coincident to the recent results of other researchers. It is also worth noting that that S protein, RdRp (nsp12 and nsp8), exo nuclease (nsp14), 2’-O methyltransferase (nsp16), helicase (nsp13) and nsp10 of SARS-CoV-2 are potential targets of saquinavir. The binding energy of nsp13, nsp14 and nsp16 with saquinavir even surpass that of nsp5, suggesting that saquinavir might be a multi-target inhibitor of SARS-CoV-2. Not surprisingly, other enzyme inhibitors of HIV such as Ritonavir, Tipranavir, Elvitegravir, Nelfinavir, Darunavir and Fosamprenavir have relatively low binding affinity with the chosen targets in our docking. Six anti-HCV drugs including five RdRp (NS5B of HCV) inhibitors, including Bictegravir, Filibuvir, Ribavirin-monophosphate, Sofosbuvir and one protease (NS3/4B) inhibitor - Bictegravir are also our best performed drugs. It is worth noting that Bictegravir has an impressively strong affinity with Mpro (binding energy -10.4 kcal/mol), nsp13 (binding energy -9.8 kcal/mol), nsp14 (binding energy -8.8 kcal/mol) and nsp15 (binding energy -8.3 kcal/mol), making it the one of best performed drugs in our docking. And the comprehensive score of filibuvir does not fall far behind that of Bictegravir and even exceeds in some docking sites. Therefore, anti-HCV drugs should be tested for battling with SARS-CoV-2. Last but not the least, Tivantinib, Lifirafenib, Entrectinib, Nilotinib and Radotinib, the chemotherapeutic drugs also for cancer treatments and Montelukast and Zafirlukast which are used in the therapy of asthma are also on the top of our list.

At the beginning of COVID-19 pandemic, two drugs used for influenza virus, Oseltamivir and Arbidol, are widely used in the treatments. However, there is no further evidence to show that Oseltamivir has obvious clinical effect so far. Both Arbidol and Oseltamivir are thought be
interacting with mainly binds to surface hemagglutinin (HA) of H2 strain of influenza viruses to block infections(76). However, no protein having such functions found in SARS-CoV-2 so far. Coincidentally, our docking result also display the low binding energies of oseltamivir with different targeted proteins of SARS-CoV-2.

Another interesting finding in our results is the performance of the natural compounds. Although most of them are at the bottom of the league and one should not over interpret the results, that fact that many of them could be find in large quantity without prescriptions make them potentially the best household compounds, especially when half of the world is on self-isolation.

There are still limitations in our study. For example, Remdesivir in the previous studies acting as RdRp inhibitors had a promising efficiency in interdicting the infection of MERS-CoV (65, 77, 78). Whereas, the binding affinity of Remdesivir with RdRp (binding energy -6.3 kcal/mol) is lower than that with endonuclease (binding energy -8.3 kcal/mol), likely to the difficulty to choose the right key residues in this interaction.

Last but not the least, our web server – Shennong, offers a new way to browse drug-protein docking results. It supports searches by either drug name or protein target name, with additional features like updates for drugs under clinical trials and a tab dedicated for natural compounds. This online platform may not only assist fast and cost-efficient drug discovery but also serve as an educational website for general public.

1. Hui DS. Epidemic and Emerging Coronaviruses (Severe Acute Respiratory Syndrome and Middle East Respiratory Syndrome). Clin Chest Med. 2017;38(1):71-86.
2. 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;382(8):727-33.
3. Paraskevis D, Kostaki EG, Magiorkinis G, Panayiotakopoulos G, Sourvinos G, Tsiodras S. Full-genome evolutionary analysis of the novel corona virus (2019-nCoV) rejects the hypothesis of emergence as a result of a recent recombination event. Infect Genet Evol. 2020;79:104212.
4. Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet. 2020;395(10224):565-74.
5. Luk HKH, Li X, Fung J, Lau SKP, Woo PCY. Molecular epidemiology, evolution and phylogeny of SARS coronavirus. Infect Genet Evol. 2019;71:21-30.
6. Ziebuhr J. Molecular biology of severe acute respiratory syndrome coronavirus. Curr Opin Microbiol. 2004;7(4):412-9.
7. Weiss SR, Leibowitz JL. Coronavirus pathogenesis. Adv Virus Res. 2011;81:85-164.
8. Brian DA, Baric RS. Coronavirus genome structure and replication. Curr Top Microbiol Immunol. 2005;287:1-30.
9. Narayanan K, Huang C, Makino S. SARS coronavirus accessory proteins. Virus Res. 2008;133(1):113-21.
10. Arndt AL, Larson BJ, Hogue BG. A conserved domain in the coronavirus membrane protein tail is important for virus assembly. J Virol. 2010;84(21):11418-28.
11. Neuman BW, Kiss G, Kunding AH, Bhella D, Baksh MF, Connelly S, et al. A structural analysis of M protein in coronavirus assembly and morphology. J Struct Biol. 2011;174(1):11-22.
12. Siu KL, Chan CP, Kok KH, Chiu-Yat Woo P, Jin DY. Suppression of innate antiviral response by severe acute respiratory syndrome coronavirus M protein is mediated through the first transmembrane domain. Cell Mol Immunol. 2014;11(2):141-9.
13. Schoeman D, Fielding BC. Coronavirus envelope protein: current knowledge. Virol J. 2019;16(1):69.
14. Liu DX, Yuan Q, Liao Y. Coronavirus envelope protein: a small membrane protein with multiple functions. Cell Mol Life Sci. 2007;64(16):2043-8.
15. Ruch TR, Machamer CE. The coronavirus E protein: assembly and beyond. Viruses. 2012;4(3):363-82.
16. McBride R, van Zyl M, Fielding BC. The coronavirus nucleocapsid is a multifunctional protein. Viruses. 2014;6(8):2991-3018.
17. Chang CK, Hou MH, Chang CF, Hsiao CD, Huang TH. The SARS coronavirus nucleocapsid protein--forms and functions. Antiviral Res. 2014;103:39-50.
18. Yan R, Zhang Y, Li Y, Xia L, Guo Y, Zhou Q. Structural basis for the recognition of the SARS-CoV-2 by full-length human ACE2. Science. 2020.
19. Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. Cell. 2020.
20. Hoffmann M, Kleine-Weber H, Schroeder S, Kruger N, Herrler T, Erichsen S, et al. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell. 2020.
21. Dimitrov DS. The secret life of ACE2 as a receptor for the SARS virus. Cell. 2003;115(6):652-3.
22. Subissi L, Imbert I, Ferron F, Collet A, Coutard B, Decroly E, et al. SARS-CoV ORF1b-encoded nonstructural proteins 12-16: replicative enzymes as antiviral targets. Antiviral Res. 2014;101:122-30.
23. Graham RL, Sparks JS, Eckerle LD, Sims AC, Denison MR. SARS coronavirus replicate proteins in pathogenesis. Virus Res. 2008;133(1):88-100.
24. Kamitani W, Narayanan K, Huang C, Lokugamage K, Ikegami T, Ito N, et al. Severe acute respiratory syndrome coronavirus nsp1 protein suppresses host gene expression by promoting host mRNA degradation. Proc Natl Acad Sci U S A. 2006;103(34):12885-90.
25. Narayanan K, Huang C, Lokugamage K, Kamitani W, Ikegami T, Tseng CT, et al. Severe acute respiratory syndrome coronavirus nsp1 suppresses host gene expression, including that of type I interferon, in infected cells. J Virol. 2008;82(9):4471-9.
26. Tohya Y, Narayanan K, Kamitani W, Huang C, Lokugamage K, Makino S. Suppression of host gene expression by nsp1 proteins of group 2 bat coronaviruses. J Virol. 2009;83(10):5282-8.
27. Huang C, Lokugamage KG, Rozovics JM, Narayanan K, Semler BL, Makino S. SARS coronavirus nsp1 protein induces template-dependent endonucleolytic cleavage of mRNAs: viral mRNAs are resistant to nsp1-induced RNA cleavage. PLoS Pathog. 2011;7(12):e1002433.
28. Lei J, Kusov Y, Hilgenfeld R. Nsp3 of coronaviruses: Structures and functions of a large multi-domain protein. Antiviral Res. 2018;149:58-74.
29. Chen X, Yang X, Zheng Y, Yang Y, Xing Y, Chen Z. SARS coronavirus papain-like protease inhibits the type I interferon signaling pathway through interaction with the STING-TRAF3-TBK1 complex. Protein Cell. 2014;5(5):369-81.
30. Baez-Santos YM, Mielech AM, Deng X, Baker S, Mesecar AD. Catalytic function and substrate specificity of the papain-like protease domain of nsp3 from the Middle East respiratory syndrome coronavirus. J Virol. 2014;88(21):12511-27.
31. Mielech AM, Deng X, Chen Y, Kindler E, Wheeler DL, Mesecar AD, et al. Murine coronavirus ubiquitin-like domain is important for papain-like protease stability and viral pathogenesis. J Virol. 2015;89(9):4907-17.
32. Zhang W, Bailey-Elkin BA, Knaap RCM, Khare B, Dalebout TJ, Johnson GG, et al. Potent and selective inhibition of pathogenic viruses by engineered ubiquitin variants. PLoS Pathog. 2017;13(5):e1006372.
33. Baez-Santos YM, Barraza SJ, Wilson MW, Agius MP, Mielech AM, Davis NM, et al. X-ray structural and biological evaluation of a series of potent and highly selective inhibitors of human coronavirus papain-like proteases. J Med Chem. 2014;57(6):2393-412.
34. Chou CY, Chien CH, Han YS, Prebanda MT, Hsieh HP, Turk B, et al. Thiopurine analogues inhibit papain-like protease of severe acute respiratory syndrome coronavirus. Biochem Pharmacol. 2008;75(8):1601-9.
35. Ghosh AK, Takayama J, Aubin Y, Ratia K, Chaudhuri R, Baez Y, et al. Structure-based design, synthesis, and biological evaluation of a series of novel and reversible inhibitors for the severe acute respiratory syndrome-coronavirus papain-like protease. J Med Chem. 2009;52(16):5228-40.
36. Ratia K, Pegan S, Takayama J, Sleeman K, Coughlin M, Baliji S, et al. A noncovalent class of papain-like protease/deubiquitinase inhibitors blocks SARS virus replication. Proc Natl Acad Sci U S A. 2008;105(42):16119-24.
37. Zheng D, Chen G, Guo B, Cheng G, Tang H. PLP2, a potent deubiquitinase from murine hepatitis virus, strongly inhibits cellular type I interferon production. Cell Res. 2008;18(11):1105-13.
38. Yin J, Niu C, Cherney MM, Zhang J, Huitema C, Eltis LD, et al. A mechanistic view of enzyme inhibition and peptide hydrolysis in the active site of the SARS-CoV 3C-like peptidase. J Mol Biol. 2007;371(4):1060-74.
39. Chou KC, Wei DQ, Zhong WZ. Binding mechanism of coronavirus main protease with ligands and its implication to drug design against SARS. Biochem Biophys Res Commun. 2003;308(1):148-51.
40. Muramatsu T, Kim YT, Nishii W, Terada T, Shirouzu M, Yokoyama S. Autoprocessing mechanism of severe acute respiratory syndrome coronavirus 3C-like protease (SARS-CoV 3CLpro) from its polyproteins. FEBS J. 2013;280(9):2002-13.
41. te Velthuis AJ, van den Worm SH, Snijder EJ. The SARS-coronavirus nsp7+nsp8 complex is a unique multimeric RNA polymerase capable of both de novo initiation and primer extension. Nucleic Acids Res. 2012;40(4):1737-47.
42. Xiao Y, Ma Q, Restle T, Shang W, Svergun DI, Ponnusamy R, et al. Nonstructural proteins 7 and 8 of feline coronavirus form a 2:1 heterotrimer that exhibits primer-independent RNA polymerase activity. J Virol. 2012;86(8):4444-54.
43. Zhai Y, Sun F, Li X, Pang H, Xu X, Bartlam M, et al. Insights into SARS-CoV transcription and replication from the structure of the nsp7-nsp8 hexadecamer. Nat Struct Mol Biol. 2005;12(11):980-6.
44. Imbert I, Guillemot JC, Bouthis JM, Bussetta C, Coutard B, Egloff MP, et al. A second, non-canonical RNA-dependent RNA polymerase in SARS coronavirus. EMBO J. 2006;25(20):4933-42.
45. te Velthuis AJ, Arnold JJ, Cameron CE, van den Worm SH, Snijder EJ. The RNA polymerase activity of SARS-coronavirus nsp12 is primer dependent. Nucleic Acids Res. 2010;38(1):203-14.
46. Kirchdoerfer RN, Ward AB. Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. Nat Commun. 2019;10(1):2342.
47. Seybert A, Hegyi A, Siddell SG, Ziebuhr J. The human coronavirus 229E superfamily 1 helicase has RNA and DNA duplex-unwinding activities with 5’-to-3’ polarity. RNA. 2000;6(7):1056-68.
48. Ivanov KA, Thiel V, Dobbe JC, van der Meer Y, Snijder EJ, Ziebuhr J. Multiple enzymatic activities associated with severe acute respiratory syndrome coronavirus helicase. J Virol. 2004;78(11):5619-32.
49. Lee NR, Kwon HM, Park K, Oh S, Jeong YJ, Kim DE. Cooperative translocation enhances the unwinding of duplex DNA by SARS coronavirus helicase nsP13. Nucleic Acids Res. 2010;38(21):7626-36.
50. Decroly E, Imbert I, Coutard B, Bouvet M, Selisko B, Alvarez K, et al. Coronavirus nonstructural protein 16 is a cap-0 binding enzyme possessing (nucleoside-2’O)-methyltransferase activity. J Virol. 2008;82(16):8071-84.
51. Chen Y, Cai H, Pan J, Xiang N, Tien P, Ahola T, et al. Functional screen reveals SARS coronavirus nonstructural protein nsp14 as a novel cap N7 methyltransferase. Proc Natl Acad Sci U S A. 2009;106(9):3484-9.
52. Wang Y, Sun Y, Wu A, Xu S, Pan R, Zeng C, et al. Coronavirus nsp10/nsp16 Methyltransferase Can Be Targeted by nsp10-Derived Peptide In Vitro and In Vivo To Reduce Replication and Pathogenesis. J Virol 2015;89(16):8416-27.
53. Ogando NS, Ferron F, Decroly E, Canard B, Posthuma CC, Snijder EJ. The Curious Case of the Nidovirus Exoribonuclease: Its Role in RNA Synthesis and Replication Fidelity. Front Microbiol. 2019;10:1813.
54. Bouvet M, Lugari A, Posthuma CC, Zevenhoven JC, Bernard S, Betzi S, et al. Coronavirus Nsp10, a critical co-factor for activation of multiple replicative enzymes. J Biol Chem. 2014;289(37):25783-96.
55. Ministry E, Hertzig T, Gorbalenya AE, Campanacci V, Cambillau C, Canard B, et al. Discovery of an RNA virus 3’->5’ exoribonuclease that is critically involved in coronavirus RNA synthesis. Proc Natl Acad Sci U S A. 2006;103(13):5108-13.
56. Kang H, Bhardwaj K, Li Y, Palaninathan S, Sacchettini J, Guarino L, et al. Biochemical and genetic analyses of murine hepatitis virus Nsp15 endoribonuclease. J Virol. 2007;81(24):13587-97.
57. Bhardwaj K, Sun J, Holzenburg A, Guarino LA, Kao CC. RNA recognition and cleavage by the SARS coronavirus endoribonuclease. J Mol Biol. 2006;361(2):243-56.
58. Zhang L, Li L, Yan L, Ming Z, Jia Z, Lou Z, et al. Structural and Biochemical Characterization of Endoribonuclease Nsp15 Encoded by Middle East Respiratory Syndrome Coronavirus. J Virol. 2018;92(22).
59. Ton AT, Gentile F, Hsing M, Ban F, Cherkasov A. Rapid Identification of Potential Inhibitors of SARS-CoV-2 Main Protease by Deep Docking of 1.3 Billion Compounds. Mol Inform. 2020.
60. Waterhouse A, Bertoni M, Bienert S, Studer G, Taurello G, Gumienny R, et al. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 2018;46(W1):W296-W303.
61. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res. 2018;46(D1):D1074-D82.
62. Tchesnokov EP, Feng JY, Porter DP, Gotte M. Mechanism of Inhibition of Ebola Virus RNA-Dependent RNA Polymerase by Remdesivir. Viruses. 2019;11(4).
63. Al-Tawfiq JA, Al-Homoud AH, Memish ZA. Remdesivir as a possible therapeutic option for the COVID-19. Travel Med Infect Dis. 2020:101615.
64. Agostini ML, Andres EL, Sims AC, Graham RL, Sheahan TP, Lu X, et al. Coronavirus Susceptibility to the Antiviral Remdesivir (GS-5734) Is Mediated by the Viral Polymerase and the Proofreading Exoribonuclease. mBio. 2018;9(2).
65. de Wit E, Feldmann F, Cronin J, Jordan R, Okumura A, Thomas T, et al. Prophylactic and therapeutic remdesivir (GS-5734) treatment in the rhesus macaque model of MERS-CoV infection. Proc Natl Acad Sci U S A. 2020.
66. Kim JY. Letter to the Editor: Case of the Index Patient Who Caused Tertiary Transmission of Coronavirus Disease 2019 in Korea: the Application of Lopinavir/Ritonavir for the Treatment of COVID-19 Pneumonia Monitored by Quantitative RT-PCR. J Korean Med Sci. 2020;35(7):e88.
67. Khot WY, Nadkar MY. The 2019 Novel Coronavirus Outbreak - A Global Threat. J Assoc Physicians India. 2020;68(3):67-71.
68. Zeng YM, Xu XL, He XQ, Tang SQ, Li Y, Huang YQ, et al. Comparative effectiveness and safety of ribavirin plus interferon-alpha, lopinavir/ritonavir plus interferon-alpha and ribavirin plus lopinavir/ritonavir plus interferon-alpha in patients with mild to moderate novel coronavirus pneumonia. Chin Med J (Engl). 2020.
69. Wang Z, Chen X, Lu Y, Chen F, Zhang W. Clinical characteristics and therapeutic procedure for four cases with 2019 novel coronavirus pneumonia receiving combined Chinese and Western medicine treatment. Biosci Trends. 2020;14(1):64-8.
70. Martinez MA. Compounds with therapeutic potential against novel respiratory 2019 coronavirus. Antimicrob Agents Chemother. 2020.
71. Rebello KM, Andrade-Neto VV, Zuma AA, Motta MCM, Gomes CRB, de Souza MVN, et al. Lopinavir, an HIV-1 peptidase inhibitor, induces alteration on the lipid metabolism of Leishmania amazonensis promastigotes. Parasitology. 2018;145(10):1304-10.
72. Tatum H, Thuluvath PJ, Lawitz E, Martorell C, DeMico M, Cohen S, et al. A randomized, placebo-controlled study of the NS5B inhibitor beclabuvir with peginterferon/ribavirin for HCV genotype 1. J Viral Hepat. 2015;22(8):658-64.
Methods

Compound libraries

We prepared a large-scale library consisting of 8,506 small molecular compounds from DrugBank. It covers all FDA-approved drugs, compounds in midst of clinical trials and molecules under experimental investigations. The SDF files were downloaded for each compound from DrugBank, whereas the SMILES files were downloaded for compounds without 3D SDF files, for example Saquinavir, Lopinavir, Ritonavir and Carfilzomlib. We converted the SMILES files to 3D SDF files for the four drugs using python rdkit library.

SARS-CoV-2 genome annotation

Reference genome of SARS-CoV-2 was downloaded from NCBI with accession number: NC_045512.2. But due to the lack of genome annotation, the protein sequence of SARS-CoV-2 cannot be obtained directly. Considering the high similarity between SARS-CoV and
SARS-CoV-2, we aligned the protein sequence of SARS-CoV to SARS-CoV-2 genome and selected the best match region as the corresponding protein sequence for SARS-CoV-2. Using this method, we obtained all the 28-protein sequence of SARS-CoV-2, including 16 non-structural proteins (nsp1-16), 4 structural proteins, Spike(S), Membrane (M), Nucleocapsid(N) and Envelope(E) 8 putative accessory proteins.

**Homology modeling of SARS-CoV-2 proteins**

Homology modelling was performed by SWISS-MODEL. PDB entries 2GDT, 6VXS, 3VCB, 6NUR, 6NUS, 1UW7, 2G9T, 6NUS, 6JYT, 5C8S, 2OZK, 3R24, 2GIB and 1SSK were used as templates to model the structures for nsp1, nsp3, nsp4, nsp7, nsp8, nsp9, nsp10, nsp12, nsp13, nsp14, nsp15, nsp16, N and E, respectively. Structures of nsp5 and S protein were extracted from PDB entries 6CS2 and 6LU7, respectively.

**Virtual docking**

*Format conversion*

The structure to be used in docking was first examined and any ligand, metal ion or other substances presenting in the structure will be removed. The PDB format was then converted to PDBQT format to meet the requirement of AutoDock Vina (79). To prepare ligand file for docking, chemical files of FDA approved, and investigational drugs were downloaded from DrugBank and then converted into PDBQT format file by OpenBabel or AutoDock Tools in Huawei Cloud EI Health Platform.

*Docking parameters*

Following our selection criteria (Table 1), amino acids of interests were highlighted, and the corresponding coordinates and size of binding box were obtained using AutoDock Tools.
Large-scale docking between protein receptors and chemicals

Protein receptors and chemical ligands were docked using over 10 thousand of CPU nodes in parallel using Huawei Cloud El Health Platform. The values of binding energy of the first model in docking PDBQT output files were used to represent and compare the binding strength for each receptor-chemical pair.

Drug-likeness analysis

We calculated five drug-likeness indexes for each compound – the ratio of sp³ hybridized carbons over the total carbon count of the molecule (Fraction Csp3) for saturation, the molecular weight for size, TPSA for polarity, XLOGP for lipophilicity, and the number of rotatable bonds for flexibility using python rdkit library. We set corresponding thresholds for each drug-likeness index to evaluate whether a compound could be drug-like, Fraction Csp3 >= 0.25, 150 <= MW <=500, 20 <= TPSA <= 130, 0.7 <= XLOGP3 <= 6, Rotatable bond num. <= 9 (60).

Shennong web server

The Bootstrap framework (www.getbootstrap.com/) and custom JavaScript were used to construct Shennong server. The Bootstrap framework provides a familiar look and provide usability on devices such as tablets and phones with different screen sizes and resolutions.

Figure captions

Fig. 1 Structure-based in silico screening and homology modelling: a) A schematic description of the drug discovery process; b) The annotation of SARS-CoV-2 genome; c) Structure availability of SARS-CoV-2 and the sequence identity compared with SARS; d) Structures obtained from PDB (PDB ID for 6CS2 nsp5 and 6LU7for S protein) and homology models built for SARS-CoV-2 using their SARS and mouse hepatitis virus A59 counterparts . PDB entries 2GDT, 6VXS, 3VCB, 6NUR, 6NUS, 1UW7, 2G9T, 6NUS, 6JYT, 5C8S, 2OZK, 3R24, 2GIB and 1SSK were used as templates to model the structures for nsp1, nsp3, nsp4, nsp7, nsp8, nsp9, nsp10, nsp12, nsp13, nsp14, nsp15, nsp16, N and E, respectively.
Table 1. **The active sites used in ligand screening:** The drug target sites, and the expected biological effects are listed for each protein; maximized space search and automatic docking was performed if no active site was given.

**Fig. 2 Overall result heatmap of binding energy for the predicted drugs:** a) The listed drugs have been reported to used in clinical trials; b) Common natural compounds. The predicted energy rank from the most antagonistic pair to the most synergistic pair is colored from blue to red.

**Fig. 3 Low-energy binding conformations of ligand and protein complexes generated by AutoDock VINA:** a) Antiviral drug Remdesivir docked in the active pocket of SARS-CoV-2 nsp14 at its interface with nsp10; b) antiretroviral drug Lopinavir docked in nsp15; c) Quinine from the Cinchona calisaya extract is docked on nsp13; d) Deconexent from fish oil docked on nsp14 at its interface with nsp10.

**Fig. 4 Best performing drugs in our docking but not currently in clinical trial to our knowledge:** a) Antiviral drug Remdesivir docked on the RNase site of nsp15; b) anti-HIV drug Beclabuvir docked at the protease site of nsp5; c) Anti-HIV drug Bictegravir docked at the protease site of nsp5; d) Antiretroviral drug Dolutegravir docked on nsp14 at the protease site of nsp5.

**Fig. 5 Shengnog web server:** Two search engines support enquiries by drug name or by target protein name and offers detailed docking results including rankings and graphic interfaces.
| Protein   | Target sites                              | Expected biological effect                                      |
|-----------|-------------------------------------------|-----------------------------------------------------------------|
| ACE2_1    | H34                                       | Prevent ACE2- S protein interaction                              |
| ACE2_2    | K353                                      | Prevent ACE2- S protein interaction                              |
| S         | F456                                      | Prevent ACE2- S protein interaction                              |
| nsp5      | L27, H41, H164                            | Block main protease activity                                    |
| nsp4      | Unspecified                               | Automatic docking by VINA                                       |
| nsp1      | Unspecified                               | Automatic docking by VINA                                       |
| nsp3      | Unspecified                               | Automatic docking by VINA                                       |
| nsp7      | K7, H36, N37                              | Prevent nsp7 forming complex with nsp12                         |
| nsp8_1    | K58                                       | Prevent nsp8 binding to RNA                                     |
| nsp8_2    | D50, D52                                  | Block its catalytic motif                                       |
| nsp9      | Unspecified                               | Automatic docking by VINA                                       |
| nsp10_1   | Ala1, Asn3, Glu6, Phe16, Phe19, Val21, Asn40, Lys43, Leu45, Thr58, Ser72, Lys93, Tyr96, His80, Cys90 | Block interaction of nsp10 and nsp14 |
| nsp10_2   |                                           |                                                                  |
| nsp10_3   |                                           |                                                                  |
| nsp12     | K332                                      | Prevent nsp12 forming complex with nsp8                         |
| nsp13     | Unspecified                               | Automatic docking by VINA                                       |
| nsp14_1   | D90, E92, E191, D273, H268                | Block Exonuclease activity                                      |
| nsp14_2   | C378, F367                                |                                                                  |
| nsp15     | K289, H234, H249, Y342                    | Block Exonuclease activity                                      |
| nsp16     | L100, N101, D130, M131                    | Block SAM binding pocket                                        |
| N         | Unspecified                               | Automatic docking by VINA                                       |
| E         | Unspecified                               | Automatic docking by VINA                                       |
Figure 3

a) Remdesivir

b) Lopinavir

c) Quinine (Cinchona calisaya extract)

d) Doconexent (DHA)
Shennong Based on Drugbank

An interactive *in silico* drug discovery database for SARS-CoV-2

- Search by drug name: e.g. Remdesivir
- Search by protein name: e.g. Spike

Primary citation: to be updated