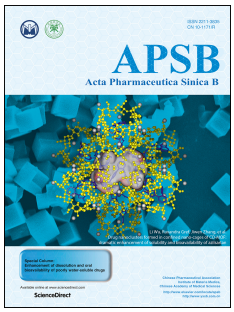
Journal Pre-proof



Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods

Canrong Wu, Yang Liu, Yueying Yang, Peng Zhang, Wu Zhong, Yali Wang, Qiqi Wang, Yang Xu, Mingxue Li, Xingzhou Li, Mengzhu Zheng, Lixia Chen, Hua Li

PII: S2211-3835(20)30299-9

DOI: https://doi.org/10.1016/j.apsb.2020.02.008

Reference: APSB 788

To appear in: *Acta Pharmaceutica Sinica B*

Received Date: 12 February 2020

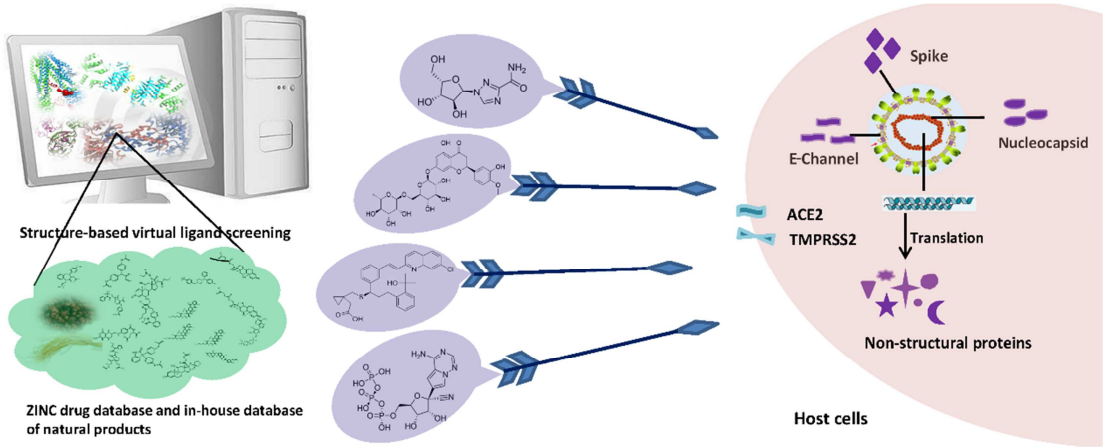
Revised Date: 17 February 2020

Accepted Date: 18 February 2020

Please cite this article as: Wu C, Liu Y, Yang Y, Zhang P, Zhong W, Wang Y, Wang Q, Xu Y, Li M, Li X, Zheng M, Chen L, Li H, Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods, *Acta Pharmaceutica Sinica B*, https://doi.org/10.1016/j.apsb.2020.02.008.

Graphical abstract

Twenty structures including 19 SARS-CoV-2 targets and 1 human target were built by homology modeling. Library of ZINC drug database, natural products, 78 anti-viral drugs were screened against these targets plus human ACE2. This study provides drug repositioning candidates and targets for further *in vitro* and *in vivo* studies of SARS-CoV-2.



Original article

Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods

Canrong Wu[a,†], Yang Liu[b,†], Yueying Yang[b,†], Peng Zhang[b], Wu Zhong[c], Yali Wang[b], Qiqi Wang[b], Yang Xu[b], Mingxue Li[b], Xingzhou Li[c,*], Mengzhu Zheng[a,*], Lixia Chen[b,*], Hua Li[a,b,*]

[a]*Hubei Key Laboratory of Natural Medicinal Chemistry and Resource Evaluation, School of Pharmacy, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, China*

[b]*Wuya College of Innovation, Key Laboratory of Structure-Based Drug Design & Discovery, Ministry of Education, Shenyang Pharmaceutical University, Shenyang 110016, China*

[c]*National Engineering Research Center for the Emergency Drug, Beijing Institute of Pharmacology and Toxicology, Beijing 100850, China*

Received 12 February 2020; Received in revised form 17 February 2020; Accepted 18 February 2020

[*]Corresponding authors.

E-mail addresses: li_hua@hust.edu.cn (Hua Li), syzyclx@163.com (Lixia Chen), mengzhu_zheng@hust.edu.cn (Mengzhu Zheng), xingzhouli@aliyun.com (Xingzhou Li).

[†]These authors made equal contributions to this work.

Abstract SARS-CoV-2 has caused tens of thousands of infections and more than one thousand deaths. There are currently no registered therapies for treating coronavirus infections. Because of time consuming process of new drug development, drug repositioning may be the only solution to the epidemic of sudden infectious diseases. We systematically analyzed all the proteins encoded by SARS-CoV-2 genes, compared them with proteins from other coronaviruses, predicted their structures, and built 19 structures that could be done by homology modeling. By performing target-based virtual ligand screening, a total of 21 targets (including two human targets) were screened against compound libraries including ZINC drug database and our own database of natural products. Structure and screening results of important targets such as 3-chymotrypsin-like protease (3CLpro), Spike, RNA-dependent RNA polymerase (RdRp), and papain like protease (PLpro) were discussed in detail. In addition, a database of 78 commonly used anti-viral drugs including those currently on the market and undergoing clinical trials for SARS-CoV-2 was constructed. Possible targets of these compounds and potential drugs acting on a certain target were predicted. This study will provide new lead compounds and targets for further *in vitro* and *in vivo* studies of SARS-CoV-2, new insights for those drugs currently ongoing clinical studies, and also possible new strategies for drug repositioning to treat SARS-CoV-2 infections.

KEY WORDS SARS-CoV-2; Drug repurposing; Molecular docking; Remdesivir; Homology modeling

Abbreviations:

3CLpro, 3-chymotrypsin-like protease; E, envelope; M, membrane protein; N, nucleocapsid protein; Nsp, non-structure protein; ORF, open reading frame; PDB, protein data bank; RdRp, RNA-dependence RNA polymerase; S, Spike; SUD, SARS unique domain; UB, ubiquitin-like domain

1. Introduction

Coronaviruses (CoVs) have caused a major outbreak of human fatal pneumonia since the beginning of the 21st century. Severe acute respiratory syndrome coronavirus (SARS-CoV) broke out and spread to five continents in 2003 with a lethal rate of 10%[1,2]. The Middle East Respiratory Syndrome Coronavirus (MERS-CoV) broke out in the Arabian Peninsula in 2012 with a fatality rate of 35%[3,4]. Both SARS-CoV and MERS-CoV are zoonotic viruses, and their hosts are bat/civet and dromedary, respectively[5,6]. To date, no specific therapeutic drug or vaccine has been approved for the treatment of human coronavirus. Therefore, CoVs are considered to be a kind of viruses, of which the outbreak poses a huge threat to humans. Because Wuhan Viral Pneumonia cases were discovered at the end of 2019, the coronavirus was named as 2019 novel coronavirus or "2019-nCoV" by the World Health Organization (WHO) on January 12, 2020[7,8]. Since 2019-nCoV is highly homologous with SARS-CoV, it is considered a close relative of SARS-CoV. The International Virus Classification Commission (ICTV) classified 2019-nCoV as Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) on February 11, 2020. At the same time, WHO named the disease caused by 2019-nCoV as COVID-19. Common symptoms of a person infected with coronavirus include respiratory symptoms, fever, cough, shortness of breath, and dyspnea. In more severe cases, infection can cause pneumonia, severe acute respiratory syndrome, kidney failure, and even death. There is currently no specific medicine or treatment for diseases caused by SARS-CoV-2[9].

CoVs are enveloped viruses with a positive RNA genome, belonging to the Coronaviridae family of the order Nidovirales, which are divided into four genera (α, β, γ, and δ). The SARS-CoV-2 belongs to the β genus. CoVs contain at least four structural proteins: Spike (S) protein, envelope (E) protein, membrane (M) protein, and nucleocapsid (N) protein[10]. Among them, Spike promotes host attachment and virus–cell membrane fusion during virus infection. Therefore, Spike determines to some extent the host range.

Potential anti-coronavirus therapies can be divided into two categories depending on the target, one is acting on the human immune system or human cells, and the other is on coronavirus itself. In terms of the human immune system, the innate immune system response plays an important role in controlling the replication and infection of coronavirus, and interferon is expected to enhance the immune response[11]. Blocking the signal pathways of human cells required for virus replication may show a certain anti-viral effect. In addition, viruses often bind to receptor proteins on the surface of cells in order to entering human cells, for example, the SARS virus binds to the angiotensin-converting enzyme 2 (ACE2) receptor[12-14] and the MERS binds to the DPP4 receptor[15,16]. The therapies acting on the coronavirus itself include preventing the synthesis of viral RNA through acting on the genetic material of the virus, inhibiting virus replication through acting on critical enzymes of virus, and blocking the virus binding to human cell receptors or inhibiting the virus's self-assembly process through acting on some structural proteins.

In the fight against coronavirus, scientists have come up with three strategies for developing new drugs[17]. **The first strategy** is to test existing broad-spectrum anti-virals[18]. Interferons, ribavirin, and cyclophilin inhibitors used to treat coronavirus pneumonia fall into this category. The advantages of these therapies are that their metabolic characteristics, dosages used, potential efficacy and side effects are clear as they have been approved for treating viral infections. But the disadvantage is that these therapies are too

"broad-spectrum" and cannot kill coronaviruses in a targeted manner, and their side effects should not be underestimated. **The second strategy** is to use existing molecular databases to screen for molecules that may have therapeutic effect on coronavirus[19,20]. High-throughput screening makes this strategy possible, and new functions of many drug molecules can be found through this strategy, for example, the discovery of anti-HIV infection drug lopinavir/ritonavir. **The third strategy** is directly based on the genomic information and pathological characteristics of different coronaviruses to develop new targeted drugs from scratch. Theoretically, the drugs found through these therapies would exhibit better anti-coronavirus effects, but the research procedure of new drug might cost several years, or even more than 10 years[11].

For the development of medicines treating SARS-CoV-2, the fastest way is to find potential molecules from the marketed drugs. Once the efficacy is determined, it can be approved by the Green Channel or approved by the hospital ethics committee for rapid clinical treatment of patients. Herein, bioinformatics analysis on the proteins encoded by the novel coronavirus genes was systematically conducted, and the proteins of SARS-CoV-2 were compared with other coronaviruses, such as SARS-CoV and MERS-CoV. We conducted homology modeling to build all possible protein structures, including viral papain like protease (PLpro), main protease (3CLpro, also named 3-chymotrypsin-like protease), RNA-dependent RNA polymerase (RdRp), helicase, Spike, etc. Further, we used these proteins and human relative proteins [human ACE2 and type-II transmembrane serine protease (TMPRSS2) enzymes] as targets to screen ZINC U. S Food and Drug Administration (FDA)-approved drug database (ZINC drug database, ZDD), our own database of traditional Chinese medicine and natural products (including reported common anti-viral components from traditional Chinese medicine), and the database of commonly used anti-viral drugs (78 compounds) by virtual ligand screening method. This study predicts a variety of compounds that may inhibit novel coronaviruses and provides scientists with information on compounds that may be effective. Subsequent validation of anti-viral effects *in vitro* and *in vivo* will provide useful information for clinical treatment of novel coronavirus pneumonia.

2. Methods and materials

2.1. Homology genome blast and genomes information

The complete genome of Wuhan-Hu-1 (NC_045512.2) was downloaded from NCBI nucleotide database. The nucleotide sequences were aligned with whole database using BLASTn to search for homology viral genomes (Alogorithm parameters, max target sequences: 1000, expect threshold: 10). Accession numbers of 23 sequences in GenBank are listed as follows: Wuhan-Hu-1 (NC_045512.2), BetaCoV/YN2018B (MK211376.1), Human/CoV HKU1 isolate SI17244 (MH940245.1), Human/CoV HKU1 genotype B (DQ415911.1), Rodent/RtTn-CoV (KY370043.1), DUCK/CoV (KM454473.1), Feline/CoV UU5 (FJ938056.1), Bat/CoV HKU9-4 (EF065516.1), MERS NL13845 (MG021451.1), 2019-nCoV/USA-IL1 (MN988713.1), 2019-nCoV WHU02 (MN988669.1), 2019-nCoV/USA-CA2 (MN994468.1), 2019-nCoV_HKU-SZ-005b (MN975262.1), 2019-nCoV/USA-WA1/2020 (MN985325.1), 2019-nCoV/USA-AZ1/2020(MN997409.1), 2019-nCoV/USA-CA1 (MN994467.1),

2019-nCoV_HKU-SZ-002a (MN938384.1), Bat SL CoV ZC45 (MG772933.1), Bat SL-CoVZXC21 (MG772934.1), SARS CoV ZS-C (AY395003.1), SARS CoV ZS-B (AY394996.1), SARS CoV SZ16 (AY304488.1), and Bat CoV isolate RaTG13 (MN996532.1)

2.2. Nucleotide and amino acid sequence alignment and analysis

Nucleotide sequence editing was conducted using Bioedit and DNAMAN, and sequence alignment was conducted using DNAMAN and Clustalw. The evolutionary history was inferred using the Neighbor-Joining method in MEGA 7 software package. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test was determined by 500 replicates. Protein sequence management and analysis were carried out by using snap gene view, and sequence alignment was performed using Clustalw method in Jalview software. The second structure of protein sequences were predicted by Jpred4.

The homology model prediction was carried out through searching in PDB1018 database included in Fold and Function Assignment System (ffas.godziklab.org)[21]. Prediction of transmembrane helices in proteins was carried out in TMHMM Server v. 2.0 online (http://www.cbs.dtu.dk/services/TMHMM/). 3D structure structures are aligned by pymol structure alignment tool.

2.3. Compounds database

Approved drug database was from the subset of ZINC database, ZDD (ZINC drug database) containing 2924 compounds[22]. Natural products database was constructed by ourselves, containing 1066 chemicals separated from traditional Chinese herbals in own laboratory and naturally occurring potential anti-viral components and derivatives. Anti-viral compounds library contains 78 known anti-viral drugs and reported anti-viral compounds through literature search.

2.4. Homology modeling and molecular docking

Corresponding homology models predicted by Fold and Function Assignment System server for each target protein were downloaded from Protein Data Bank (www.rcsb.org). Alignment of two protein sequences and subsequent homology modeling were performed by bioinformatics module of ICM 3.7.3 modeling software on an Intel i7 4960 processor (MolSoft LLC, San Diego, CA, USA). For the structure-based virtual screening, ligands were continuously resiliently made to dock with the target that was represented in potential energy maps by ICM 3.7.3 software, to identify possible drug candidates. 3D compounds of each database were scored according to the Internal Coordinate Mechanics (Internal Coordinate Mechanics, ICM)[23]. Based on Monte Carlo method, stochastic global optimization procedure and pseudo-Brownian positional/torsional steps, the position of intrinsic molecular was optimized. By visually inspecting, compounds outside the active site, as well as those weakly fitting to the active site were eliminated. Compounds with scores less than –30 or mfScores less than –100 (generally represents strong interactions) have priority to be selected. Protein–protein docking was followed the manual of ICM software.

3. Results

3.1. Analysis, structure prediction and homology modeling of SARS-CoV-2 encoded proteins

We obtained the SARS-CoV-2 genome from Gene Bank. The genome sequence of Wuhan-Hu-1 was aligned with whole database using BLASTn to search for homology viral genomes. After phylogenetic analysis and sequence alignment of 23 coronaviruses from various species. We found three coronaviruses from bat (96%, 88% and 88% for Bat-Cov RaTG13, bat-SL-CoVZXC12 and bat-SL-CoVZC45, respectively) have the highest genome sequence identity to SARS-CoV-2 (Fig. 1A). Moreover, as shown in Fig. 1B, Bat-Cov RaTG13 exhibited the closest linkage with SARS-CoV-2. These phylogenetic evidences suggest that SARS-CoV-2 may be evolved from bat CoVs, especially RaTG13. Among all coronaviruses from human, SARS-CoV (80%) exhibited the highest genome sequence identity to SARS-CoV-2. And MERS/isolate NL13845 also has 50% identity with SARS-CoV-2. SARS-CoV is the most clearly studied one among all these viruses according to previous literatures. Structure and function of its most genome encoded proteins have been elucidated in recent years. In this study, because of high genome identity between SARS-CoV-2 and SARS-CoV, the structure and function prediction of SARS-CoV-2 genome encoded protein were mainly based on those researches on homology protein in SARS-CoV. SARS-CoV-2 genome has 10 open reading frames (Fig. 2A). ORF1ab encodes replicase polyprotein 1ab. After cleaved by two proteases, replicase proteins showed multifunction involved in transcription and replication of viral RNAs. ORF2-10 encodes viral structural proteins such as S, M, N, and E proteins, and other auxiliary proteins. The S, M, E proteins are involved in the formation of the viral coat, and the N protein is involved in the packaging of the RNA genome (Fig. 2C). By aligning with the amino acid sequence of SARS PP1ab and analyzing the characteristics of restriction cleavage sites recognized by 3CLpro and PLpro, we speculated 14 specific proteolytic sites of 3CLpro and PLpro in SARS-CoV-2 PP1ab (Fig. 2B). PLpro cleaves three sites at 181–182, 818–819, and 2763–2764 at the N-terminus and 3CLpro cuts at the other 11 sites at the C-terminus, and forming 15 non-structural proteins. Among them, Nsp3 contains multiple domains, including a segment of SARS unique domain and a deubiquitination and proteolytic enzyme PLpro. Nsp5 is 3CLpro, Nsp12 is an RdRp, and Nsp13 is helicase. As a new coronavirus, structure biology study about these proteins still at early stage. Until now, only one crystal structure of 3CLpro has been deposited in PDB (pdb code: 6LU7).

In order to acquire more three-dimensional structure information of proteins about these new coronaviruses for subsequent drug screening, we aligned all protein sequences from SARS-CoV-2 with all sequences in PDB1018 database in Fold & Function Assignment System (Supporting Information Figs. S19–S34). Fortunately, most of these proteins have found their high homology proteins that have three-dimensional structure, with homology between 72%–99% (Supporting Information Table S1). Those PDB codes were labeled below the corresponding sequences in Fig. 2B. Unsurprisingly, all these proteins with the highest homology are from SARS. Nevertheless, there are some proteins still have not high homologous in the database. After the prediction of transmembrane helices in these proteins carried out in TMHMM Server, as expected, we found that all these proteins are transmembrane proteins except for Nsp2 (Supporting Information Figs. S35–S41). So far, we have found as much structural information of this viral proteins as

possible, which provides the basis for subsequent homology modeling and drug screening. Before homology modeling, all these proteins sequences were aligned with model sequences derived from SARS-CoV and predicted secondary structure as the same time (Fig. 3A and Table S1). Interestingly, as shown in Fig. 3A, important anti-virus drug target protein like 3CLpro, PLpro, and RdRp are highly conserved between those two human coronaviruses, especially in functional region. As shown in Table S1, all these potential drug target proteins have been homologously modeled, and all generated protein models were provided as the PDB files in Supporting Information. All other coordinates of target-screening hit complexes can be provided upon request.

In order to verify the accuracy of homologous modeling, we aligned the computational structure of the SARS-CoV-2 3CLpro that modeled from the SARS-CoV 3CLpro with its crystal structure of SARS-CoV-2 3CLpro (6LU7) just solved and released during this manuscript was being prepared. The computational model of SARS-CoV-2 3CLpro showed a Cα RMSD of 0.471 Å on the overall structure compared to the SARS-CoV-2 3CLpro structure (Fig. 3B). What's more, the Cα RMSD between their substrate binding regions is only 0.126 Å, showing very subtle difference.

ACE2 molecule was known as a human entry receptor for Spike which facilitates its cross-species transmission. Sequence alignment results show that the homology of the Spike-receptor binding domain (RBD) sequence between SARS-CoV-2 and SARS-CoV is 76% (Fig. 3C). Recent researches speculated that SARS-CoV-2 could also bind to ACE2, and this was verified by computational docking and ELISA measurement[24,25]. Moreover, homology of the Spike-RBD sequence between SARS-CoV-2 and Bat-CoV RaTG13 is as high as 95%. Despite the high homology of RaTG13 and SARS-CoV-2 in Spike sequence, our analysis found that four among the five most important amino acids (L465, L495, Y502, D510, and H514) that bind to ACE2[12] in Bat-CoV RaTG13 differ from SARS-CoV-2 (Fig. 3C). And there is no related research literature about whether Bat-CoV RaTG13 can infect human yet. We also performed homology modeling on the Bat-CoV RaTG13 Spike RBD (Supporting Information Fig. S1). Three Spike RBD structures have been docked with human ACE2. Among them, for the conformations which most resemble the crystal structure of SARS RBD–ACE2 complex[26], the binding free energy between SARS-CoV-2 Spike RBD and human ACE2 was –33.72 kcal/mol (Supporting Information Fig. S2), that between SARS-CoV Spike RBD and ACE2 was –49.22 kcal/mol (Supporting Information Fig. S3), and that between Bat-CoV RaTG13 Spike RBD and ACE2 was –31.06 kcal/mol (Supporting Information Fig. S4).

3.2. Virtual ligand screening of potential drug targets of SARS-CoV-2

The therapies that act on the coronavirus can be divided into several categories based on the specific pathways: (1) some acting on enzymes or functional proteins that are critical to virus, preventing the virus RNA synthesis and replication; (2) some acting on structural proteins of virus, blocking virus from binding to human cell receptors, or inhibiting the virus's self-assembly process; (3) some producing virulence factor to restore host's innate immunity; (4) some acting on host's specific receptors or enzymes, preventing virus from entering into host's cells. The related target proteins include Nsp1, Nsp3 (Nsp3b, Nsp3c, PLpro, and Nsp3e),

Nsp7_Nsp8 complex, Nsp9–Nsp10, and Nsp14–Nsp16, 3CLpro, E-channel (E protein), ORF7a, Spike, ACE2, C-terminal RNA binding domain (CRBD), N-terminal RNA binding domain (NRBD), helicase, RdRp, and TMPRSSS2.

Based on homology models of the above 18 viral proteins (19 models) and 2 human targets, we resorted to structure-based virtual ligand screening method using ICM 3.7.3 modeling software (MolSoft LLC) to screen potential small-molecule compounds from a ZINC Drug Database (2924 compounds) and a small in-house database of traditional Chinese medicine and natural products (including reported common anti-viral components from traditional Chinese medicine) and derivatives (1066 compounds). Compounds with lower calculated binding energies (being expressed with scores and mfScores) are considered to have higher binding affinities with the target protein.

3.2.1. Targets preventing the virus RNA synthesis and replication

As significant functional proteins of coronavirus, Nsps are involved in RNA transcription, translation, protein synthesis, processing and modification, virus replication and infection of the host. Among them, 3CLpro, PLpro, RdRp and helicase are the most important targets for the development of small-molecule inhibitors due to their clear biological functions and vital enzyme active site.

3.2.1.1. Papain-like proteinase (PLpro) PLpro is responsible for the cleavages of N-terminus of the replicase poly-protein to release Nsp1, Nsp2 and Nsp3, which is essential for correcting virus replication[27]. PLpro was also confirmed to be significant to antagonize the host's innate immunity[28-30]. As an indispensable enzyme in the process of coronavirus replication and infection of the host, PLpro has been a popular target for coronavirus inhibitors. It is very valuable for targeting PLpro to treat coronavirus infections, but no inhibitor has been approved by the FDA for marketing.

The screening results (Table 1 and supporting excel file PLpro.xlsx) showed that a series of anti-virus drugs (ribavirin, valganciclovir and thymidine), anti-bacterial drugs (chloramphenicol, cefamandole and tigecycline), muscle relaxant drug (chlorphenesin carbamate), anti-tussive drug (levodropropizine) may have high binding affinity to PLpro. The natural products (Table 2 and Supporting excel file PLpro_NP.xlsx), such as platycodin D from *Platycodon grandiflorus*, baicalin from *Scutellaria baicalensis*, sugetriol-3,9-diacetate from *Cyperus rotundus*, phaitanthrin D and 2,2-di(3-indolyl)-3-indolone from *Isatis indigotica*, catechin compounds ((−)-epigallocatechin gallate and 2-(3,4-dihydroxyphenyl)-2-[[2-(3,4-dihydroxyphenyl)-3,4-dihydro-5,7-dihydroxy-2*H*-1-benzopyran-3-yl]oxy]-3,4-dihydro-2*H*-1-benzopyran-3,4,5,7-tetrol) exhibited high binding affinity to PLpro protein, suggesting the potential utility of these compounds in the treatment of SARS-CoV-2.

Anti-viral drug ribavirin was predicted to bind to PLpro with low binding energy (Scores=–38.58). From generated docking model, ribavirin was bound in the active site of the enzyme as reported SARS-PLpro inhibitors (PDB code 3e9s). Hydrogen bonds were predicted between Gly164, Gln270, Tyr274, Asp303 and the compound. Also, π–π stacking was found between Tyr265 and triazolering in the compound (Fig. 4A and

B). The strong hydrogen bonding and hydrophobic interaction between ribavirin with the enzyme imply it may be a potent PLpro inhibitor.

3.2.1.2. 3C-like main protease (3CLpro) The 3CLpro, also known as Nsp5, is first automatically cleaved from poly-proteins to produce mature enzymes, and then further cleaves downstream Nsps at 11 sites to release Nsp4–Nsp16[31]. 3CLpro directly mediates the maturation of Nsps, which is essential in the life cycle of the virus. The detailed investigation on the structure and catalytic mechanism of 3CLpro makes 3CLpro an attractive target for anti-coronavirus drug development. Inhibitors targeting at SARS-CoV 3CLpro mainly include peptide inhibitors and small-molecule inhibitors [32].

3CLpro monomer has three domains, domain I (residues 8–101), domain II (residues 102–184) and domain III (residues 201–303), and a long loop (residues 185–200) connects domains II and III. The active site of 3CLpro is located in the gap between domains I and II, and has a CysHis catalytic dyad (Cys145 and His41)[33]. As shown in Table 3 and Supporting excel file 3CLpro.xlsx), anti-bacterial drugs (lymecycline, demeclocycline, doxycycline and oxytetracycline), anti-hypertensive drugs (nicardipine and telmisartan), and conivaptan treating hyponatremia show highest binding affinity to 3CLpro. Several natural compounds and derivatives with anti-virus and anti-inflammatory effects also exhibited high binding affinity to 3CLpro (Table 4 and Supporting excel file 3CLpro_NP.xlsx), including a series of andrographolide derivatives (chrysin-7-*O*-*β*-glucuronide from *Scutellaria baicalensis*, betulonal from *Cassine xylocarpa*, 2*β*-hydroxy-3,4-*seco*-friedelolactone-27-oic acid, isodecortinol and cerevisterol from *Viola diffusa*, hesperidin and neohesperidin from *Citrus aurantium*, kouitchenside I and deacetylcentapicrin from the plants of *Swertia* genus. The above results suggest that these small-molecule compounds might be the potential 3CLpro inhibitors and could probably be used for treating SARS-CoV-2.

It's worth mentioning, anti-asthmatic drug montelukast also showed low binding energy to 3CLpro. As shown in Fig. 5A, montelukast was well fitted into the active pocket of 3CLpro, in which lots of hydrophobic amino acids, just like Thr24, Leu27, His41, Phe140, Cys145, His163, Met165, Pro168 and His172 compose a relatively hydrophobic environment to contain the compound and stabilize its conformation. Hydrogen bonding was predicted between Asn142 and the carbonyl group of the compound (Fig. 5B).

3.2.1.3. RNA-dependent RNA polymerase (RdRp) Nsp12, a conserved protein in coronavirus, is an RNA-dependent RNA polymerase (RdRp) and the vital enzyme of coronavirus replication/transcription complex. The RdRp domain of polymerase is located at the C-terminus and has a conserved Ser-Asp-Asp motif[34]. Nsp8 can *de novo* synthesize up to 6 nucleotides in length, which can be used as a primer for Nsp12-RdRp RNA synthesis. Further, the Nsp7_Nsp8 complex increases the binding of Nsp12 to RNA and enhances the RdRps enzyme activity of Nsp12[35]. In the research of SARS-CoV and MERS-CoV inhibitors, Nsp12-RdRp has been used as a very important drug target. In principle, targeted inhibition of Nsp12-RdRp could not cause significant toxicity and side effects on host cells, but no specific inhibitors have been found until now[36].

Virtual screening results of RdRp demonstrated some drugs might be potential inhibitors with mfScores lower than –110, such as several anti-fungal drug itraconazole, anti-bacterial drug novobiocin, gallstone-dissolving drug chenodeoxycholic acid, anti-allergic drug cortisone, anti-tumor drug idarubicin, hepatoprotective drug silybin, muscle relaxant drug pancuronium bromide, and chronic enteritis, anti-coagulant drug dabigatran etexilate (Table 5 and Supporting excel file RdRp.xlsx). The natural products and derivatives with anti-virus, anti-inflammation and anti-tumor effects exhibited high binding affinity to RdRp, such as betulonal from *Cassine xylocarpa*, gnidicin and gniditrin from *Gnidia lamprantha*, 2*β*,30*β*-dihydroxy-3,4-*seco*-friedelolactone-27-lactone from *Viola diffusa*, 14-deoxy-11,12-didehydroandrographolide from *Andrographis paniculata*, 1,7-dihydroxy-3-methoxyxanthone from *Swerti apseudochinensis*, theaflavin 3,3′-di-*O*-gallate from *Camellia sinensis*, and andrographolide derivative (*R*)-((1*R*,5a*S*,6*R*,9a*S*)-1,5a-dimethyl-7-methylene-3-oxo-6-((*E*)-2-(2-oxo-2,5-dihydrofuran-3-yl)ethenyl)decahydro-1*H*-benzo[*c*]azepin-1-yl)methyl 2-amino-3-phenylpropanoate (Table 6 and Supporting excel file RdRp_NP.xlsx).

3.2.1.4 Helicase Helicase (Nsp13), a multi-functional protein, include N-terminal metal binding domain (MBD) and helicase domain (Hel). N-terminal structure contains 26 cysteine residues to form a Zn^{2+} binding domain and helicase domain with a conserved motif at the C-terminus. Nsp13 can unravel double-stranded (ds) DNA and RNA along the 5′–3′ direction in an NTP-dependent manner[37]. Importantly, it has been reported that the SARS-Nsp13 sequence is conserved and indispensable, and is a necessary component for the replication of coronavirus. Therefore, it has been identified as a target for anti-viral drug discovery, but there are few reports about Nsp13 inhibitors[38,39].

Based on structure modeling of helicase protein, anti-bacterial drugs (lymecycline, cefsulodine and rolitetracycline), anti-fungal drug itraconazole, anti-human immunodeficiency virus-1 (HIV-1) drug saquinavir, anti-coagulant drug dabigatran, and diuretic drug canrenoic acid were predicted to be helicase inhibitors with high mfScores through virtual ligand screening. The natural products, such as many flavanoids from different sources (*α*-glucosyl hesperidin, hesperidin, rutin, quercetagetin 6-*O*-*β*-D-glucopyranoside and homovitexin), xanthones (3,5-dimethoxy-1-[(6-*O*-*β*-D-xylopyranosyl-*β*-D-glucopyranosyl)oxy]-9*H*-xanthen-9-one, kouitchenside H, kouitchenside A, 8,2-dihydroxy-3,4,5-trimethoxy-1-[(6-*O*-*β*-D-xylopyranosyl-*β*-D-glucopyranosyl)oxy]-9*H*-xanthen-9-one, kouitchenside D, 1-hydroxy-2,6-dimethoxy-8-[(6-*O*-*β*-D-xylopyranosyl-*β*-D-glucopyranosyl)oxy]-9*H*-xanthen-9-one and triptexanthoside D from *Swertia* genus, phyllaemblicin B and phyllaemblinol from *Phyllanthus emblica* showed high binding affinity to this target.

Besides the above targets, some non-structural proteins, including Nsp3b, Nsp3e, Nsp7_Nsp8 complex, Nsp9, Nsp10, Nsp14, Nsp15, and Nsp16, also play an important role in the virus RNA synthesis and

replication, suggesting these proteins may be useful targets for the anti-viral drug discovery. The virtual screening results showed many anti-bacterial, anti-viral, or anti-inflammatory drugs from ZINC drug database and our in-house natural products/derivatives database displayed potential good affinity to these targets, and the detailed information of virtual screening results is shown in Supporting excel files (for ZDD screening results, file names as target.xlsx; for natural products screening results, file names as target_NP.xlsx).

3.2.2. Targets inhibiting virus structural proteins

Spike is the main structural protein of coronavirus and assembles into a special corolla structure on the surface of the virus as a trimer. Spike is a main protein that interacts with the host by binding to host cell receptors to mediate virus invasion and determine viral tissue or host tropism[40]. Spike is cleaved into S1 and S2 by the host cell protease like TMPRSS2, etc. The main function of S1 is to bind with host cell surface receptors, and the S2 subunit mediates virus–cell and cell–cell membrane fusion. Spike structural integrity and cleavage activation play a key role in virus invasion and virulence[41]. Therapeutic strategies to block coronavirus from entering host cells by targeting Spike proteins or specific receptors on the host surface are valuable for the development of anti-viral drugs.

On the basis of virtual screening results of small-molecule compounds against Spike protein, some drugs showed high binding affinity (mfScore<–150 or score<–35), such as anti-hypertensive drugs (rescinnamine, iloprost and prazosin), anti-fungal drugs (posaconazole and itraconazole), anti-bacterial drug (sulfasalazine, azlocillin, penicillin and cefsulodin) and anti-coagulant drug dabigatran etexilate. Some natural flavanoids, licoflavonol from *Glycyrrhiza uralensis*, cosmosiinfrom *Scutellaria baicalensis*, neohesperidin from *Citrus aurantium*, mangostinfrom *Garcinia mangostana*, kouitchenside D from *Swertia kouitchensis*, excoecariatoxin from *Excoecaria agallocha*, phyllaemblicin G7 from *Phyllanthus emblica*, and piceatannol from *Vitis vinifera*, exhibited high binding affinity. The detailed virtual screening results are shown in Supporting excel files.

However, most of above compounds were not predicted to bind with the binding interface of the Spike–ACE2 complex. The only compound that could target the binding interface between Spike and ACE2 was hesperidin, as shown in Fig. 6A. Hesperidin was predicted to lie on the middle shallow pit of the surface of RBD of Spike, where the dihydroflavone part of the compound went parallel with the β-6 sheet of RBD. And the sugar part was inserted into the shallow pit in the direction away from ACE2, where a few hydrophobic amino acids, including Tyr436, Try440, Leu442, Phe443, Phe476, Try475, Try481 and Tyr49 form a relatively hydrophobic shallow pocket to contain the compound (Fig. 6B). Hydrogen bonding was predicted between Tyr440 and the compound. By superimposing the ACE2–RBD complex to the hesperidin–RBD complex, a distinct overlap of hesperidin with the interface of ACE2 could be observed (Fig. 6C), suggesting hesperidin may disrupt the interaction of ACE2 with RBD.

Except for Spike protein, E protein (E-channel) possesses important biological functions for the structural integrity of coronavirus and host virulence. NRBD and CRBD of coronavirus N protein are needed for N proteins in host cells to bind with coronavirus RNA efficiently. Therefore, E protein or N protein (NRBD and CRBD domains) can be used as targets for the discovery of anti-viral drugs. Through virtual

screening, many anti-bacterial, anti-viral, anti-tumor, anti-asthmatic, and anti-inflammatory drugs, etc. from ZINC database and our in-house natural products/derivatives database were found to display relatively good affinity to these targets. And the detailed results of virtual screening are given in Supplementary excel files.

3.2.3 Targets inhibiting virulence factor

There are three coronavirus virulence factors Nsp1, Nsp3c and ORF7a related to interfering host's innate immunity and assisting coronavirus immune escape. Nsp1 interacts with host 40S ribosomal subunit that induces specifically host mRNA degradation[42] and also inhibits type-I interferon production[43]. Nsp3c has ability to bind with host's ADP-ribose to help coronavirus resist host innate immunity[44]. Bone marrow matrix antigen 2 (BST-2) can inhibit the release of newly-assembled coronavirus from host cells. SARS-CoV ORF7a directly binds to BST-2 and inhibits its activity by blocking the glycosylation of BST-2[45]. These evidences suggest that Nsp1, Nsp3c and ORF7a may be potential targets for anti-viral drug discovery.

The detailed screening results of Nsp1, Nsp3c, and ORF7a showed that a series of clinical drugs and natural products with anti-bacterial and anti-inflammatory effects exhibited relatively high binding affinity to these three target proteins, such as piperacillin, cefpiramide, streptomycin, lymecycline, tetracycline, platycodin D from *Platycodon grandifloras*, wogonoside from *Scutellaria baicalensis*, vitexin from *Vitex negundo*, andrographolide derivatives, and xanthones from *Swertia* genus. The detailed results of virtual screening are shown in Supporting excel files.

3.2.4. Targets blocking hose specific receptor or enzymes

The host ACE2 has been proved by many studies to be the specific receptor for the Spike RBD of SARS-CoV[12]. The latest research shows that the host receptor of SARS-CoV-2 is consistent with SARS-CoV, exhibiting that the Spike RBD sequence of SARS-CoV-2 is similar to SARS-CoV RBD and there are important interactions between several key amino acid residues of RBD receptor-binding motif and ACE2[24]. Based on the current research progress, ACE2 is considered as a host target for the treatment of coronavirus infection to block SARS-CoV-2 from entering host cells.

Based on the virtual screening results of ACE2 protein, anti-diabetes drug troglitazone, anti-hypertensive drug losartan, analgesia drug ergotamine, anti-bacterial drug cefmenoxime, and hepatoprotective drug silybin, etc., were predicted to bind with ACE2 with low energy. The natural products, such as phyllaemblicin G7 from *Phyllanthus emblica*, xanthones from the plants of *Swertia*genus, neohesperidin and hesperidin from *Citrus aurantium*, exhibited potentially high binding affinity to ACE2 protein. However, none of above ACE2 binding compounds was predicted to target the ACE2–RBD interface.

In addition, TMPRSS2 was known to cut the Spike to trigger the infection of SARS-CoV and MERS-CoV. Studies have shown that inhibiting the enzyme activity of TMPRSS2 can prevent some coronaviruses from entering host cells[46]. As a possible target for anti-viral drug discovery, the virtual screen results (shown in Supporting excel files) predicted many anti-bacterial drugs (pivampicillin, hetacillin, cefoperazone and clindamycin) and anti-virus natural compounds (phyllaemblicin G7, neoandrographolide, kouitchenside I), etc. to be potential TMPRSS2 inhibitors.

3.3 Virtual screening and target identification of common anti-viral drugs

In order to further verify the screening results of ZINC drug database and utilize the current resource of anti-viral drugs immediately, we constructed a database of 78 anti-viral drugs for deep calculation, including compounds already on the market and currently undergoing clinical trials to treat SARS-CoV-2 infections. The compounds were docked with 19 constructed targets of SARS-CoV-2 and 2 human targets to predict their possible targets, and also to search possible binding partners of certain important targets. Special attentions were paid to the drugs currently in clinical trials. All results of calculated docking scores were listed in the Supporting excel file (anti-virals_vs_targets_scores), significant scores (score<–32, mfScore<–110) were highlighted in brighten yellow.

Remdesivir (GS-5734), a nucleoside analogue, is an RdRp inhibitor. It can inhibit virus by inhibiting synthesis of viral nucleic acid and has not yet been approved for marketing in any country. On January 31 of this year, *the New England Journal of Medicine* reported the diagnosis and treatment of the first SARS-CoV-2 patient in the United States[47]. Remdesivir showed some potential in the treatment of the first patient with novel coronavirus infection. On Vero E6 cells, the EC_{50} of remdesivir for SARS-CoV-2 is 0.77 μmol/L, and the selection index SI is greater than 129[48]. Remdesivir is currently undergoing a randomized, double-blind, controlled phase III clinical study in China. Our molecular docking results show that the potential targets of remdesivir is Nsp3b (score=–36.5), RdRp (mfScores=–112.8), E-channel (mfScore=–125.1), and TMPRSS2 (score=–36.23, mfScores=–109.4).

For RdRp, from generated docking model, remdesivir could bind to the RNA-binding channel of the SARS-CoV-2 RdRp, and the binding mode and site were highly similar to that of coxsackievirus B3 (CVB3) RdRp inhibitor GPC-N114[49] .The binding pocket of remdesivir was in the bottom of the RNA template channel, which position was for the acceptor template nucleotide (Fig. 7A and B). The compound was well fitted with the shape of the pocket, where it formed three hydrogen bonds with Asn497, Arg569 and Asp684. In addition, hydrophobic interactions with Leu576, Ala685 and Tyr687 may further direct the favorite conformation of remdesivir (Fig. 7C).

Interestingly, remdesivir was predicted to bind with the target TMPRSS2 with low binding energy for both score and mfScore. As shown in Fig. 8A, remdesivir was bound in a relatively positively-charged allosteric pocket which is far away from the enzyme active center. Asn84 and Arg405 formed two hydrogen bonds with the phosphate groups of the compound. Weak hydrophobic interaction between the pyrrolotriazine ring of remdesivir with Tyr131 and Try401, and side chains of some polar amino acids may further stabilize the compound conformation (Fig. 8B).

Lopinavir and ritonavir have been marketed in China and are mainly used to treat HIV-1 infection in adults and children over 2 years of age. *In vitro* studies have shown that lopinavir and ritonavir can inhibit the replication of MERS-CoV and SARS-CoV to exert anti-viral effects. The drug is listed as a recommended drug in the Anti-viral Drugs section of *the New Coronavirus Infected Pneumonia Diagnosis and Treatment Program*. The molecular docking results showed that ritonavir's possible target is Nsp3c or E-channel with the

mfScores of –152.100 and –277.769, respectively. Lopinavir's possible target is Nsp3b, Nsp3c, helicase, NRBD or E-channel with the mfScores of –158.050, –189.140, –114.018, –171.127, and –221.785, respectively.

Arbidol is a broad-spectrum anti-viral drug, mainly for the treatment of upper respiratory tract infections caused by influenza A and B viruses, etc. In recent years, many studies have proven its effectiveness against both SARS-CoV and MERS-CoV. Arbidol hydrochloride can block virus replication by inhibiting the fusion of the lipid membrane of the virus with the host cells. Compared with the untreated control group, arbidol can effectively inhibit coronavirus up to 60 times at a concentration of 10–30 μmol/L, and significantly inhibit the virus's pathological effects on cells. The docking results of arbidol with the possible drug targets of the new coronavirus showed that it may interact with Nsp7_Nsp8 complex, Nsp14, Nsp15, E-channel, or Spike with the mfScores of –136.087, –118.253, –118.253, –117.879, and –145.125, respectively.

Darunavir is an HIV-1 protease inhibitor that selectively inhibits the cleavage of HIV-encoded Gag-Pol polyprotein in virally infected cells, thereby preventing the formation of mature infectious virus particles. At a concentration of 300 μmol/L, darunavir can significantly inhibit virus replication, and the inhibition efficiency is 280 times compared with the untreated group. Our docking results showed that the possible targets of darunavir are Nsp3c, PLpro, E-channel or Spike proteins with the mfScores of –126.149, –110.759, –157.184, and –111.865, respectively.

Favipiravir, a broad-spectrum anti-viral drug, is used to treat flu. The Shenzhen Health Commission has now initiated clinical studies on the use of favipiravir to treat SARS-CoV-2 infections. While, the scores of favipiravir docking with the targets in our virtual screening are relatively low.

Chloroquine phosphate has been used in the treatment of malaria since the 1940s and later in rheumatoid arthritis. Chloroquine has been reported in some earlier studies to have direct anti-viral effects, such as inhibiting flavivirus, retrovirus (such as HIV), and many coronaviruses. Recent study showed that with EC_{50} of 1.13 μmol/L and SI greater than 88, chloroquine can effectively inhibit SARS-CoV-2 in the cell level. Its efficiency in the human body for SARS-CoV-2 infection has not yet been clinically proven. Chloroquine phosphate was turned into chloroquine in the body to play therapeutic effect. Our docking results showed that the possible target of chloroquine is Nsp3b or E-channel with the docking mfScores of –130.355 and –107.889, respectively.

We also performed the longitudinal analysis on the drugs against 21 targets, and the results showed that only tenofovir, disoproxil, fumarate, and beclabuvir may bind to Nsp1. Fewer compounds were predicted to act on some targets, such as Nsp1, Nsp3e, Nsp9, Nsp10, Nsp16, NRBD, CRBD, ORF7a, and TMPRSS2, showing that these selected anti-viral drugs are unlikely to acting on the above targets of the new coronavirus, which provides a meaningful reference for our future research.

For other targets, such as Nsp3b, Nsp3c, Nsp7_Nsp8 complex, Nsp14, Nsp15, PLpro, 3CLpro, RdRp, helicase, E-channel, Spike and ACE2, more anti-viral drugs were predicted to bind with them, especially E-channel, RdRp, 3CLpro and PLpro, indicating that these targets are more likely to be useful for the discovery of SARS-CoV-2 therapeutic drugs from known anti-virals, and should be the focus of our

subsequent research.

4. Discussion

The ongoing SARS-CoV-2 epidemic makes us painfully realize that our current options for treating life-threatening zoonotic coronavirus infections are very limited. Although the outbreaks of SARS in 2003 and MERS-CoV in 2012 triggered extensive research efforts, there are currently no drugs that can treat any zoonotic coronavirus. The transient nature of this epidemic is one of the major reasons why no prototype coronavirus inhibitor has progressed to the (early) preclinical stage to date. As with the SARS virus 17 years ago and the current SARS-CoV-2, emerging coronaviruses in the future may continue to pose a threat to global public health. Therefore, finding broad-spectrum inhibitors that may reduce the effects of human coronavirus infection remains a challenging research focus. Given the time-consuming nature of anti-viral drug development and registration, existing treatments for other diseases may be the only fastest treatment option for emerging infectious diseases. For most of these drugs that have been prepared, the medication has sufficient experience and dosage, and their safety and ADME situation are well known.

In this study, based on the results of bioinformatics analysis, 20 homology structures of 18 SARS-CoV-2 proteins and 1 human protein were built, plus human ACE2 structure, totally 21 targets were setup for high throughput virtual ligand screening. The special double scoring system ICM-Pro soft allowed us to evaluate the docking results more accurately. The score of ICM software is calculated by the overall empirical function of predicted physical interaction, whiles the mfScore (mean force Score) is computed by the knowledge-based potential functions derived from statistics of ligand–receptor complex in PDB[50,51].

From protein–protein docking results, we can deduce that SARS-CoV-2 Spike protein have strong binding affinity to human ACE2, although weaker than that of SARS-CoV Spike protein. Unexpectedly, Bat-CoV RaTG13 Spike protein only has slightly weaker binding affinity with ACE2 compared to SARS-CoV-2. It is speculated that the virus may be capable of infecting humans directly, or it is close. If SARS-CoV-2 evolved from Bat-CoV RaTG13, there may be no more than one or two host evolutions involved. But there are no published articles that directly related to intermediate hosts yet. According to the most recent research paper on the clinical characteristics of SARS-CoV-2 infection[52], SARS-CoV-2 can spread rapidly by human-to-human transmission, this was consistent with observed protein–protein docking results of SARS-CoV-2 Spike RBD and human ACE2, for both of ours and previous prediction[25]. However, the fast growing incidents of SARS-CoV-2 pneumonia and related deaths implied the weaker binding of SARS-CoV-2 Spike with human ACE2 compared to SARS Spike may not fully explained the current situation of this epidemic. There must have some other receptors of SARS-CoV-2 or infection facilitating mechanisms in the human host, or SARS-CoV-2 Spike-RBD may take other lower energy binding conformation different from current SARS-CoV Spike-RBD conformation found in the crystal structure, probably by induce fitting. These need to be further investigated.

To search potential coronavirus therapeutic drugs as soon as possible, we first screened potential compounds from a ZINC drug database (2924 compounds) and a small in-house database of natural products

(about 1066 compounds). A series of clinical drugs and natural products with anti-viral, anti-bacterial and anti-inflammatory effects exhibited high binding affinity to different target proteins, indicating their potential utility for treating SARS-CoV-2.

We have identified a number of compounds that might have anti-viral activity from the approved drugs library, such as anti-virus drugs (ribavirin, valganciclovir and thymidine), anti-bacterial drugs (cefpiramide, sulfasalazine, phenethicillin, lymecycline, demeclocycline, doxycycline, oxytetracycline and tigecycline), anti-asthmatic drugs (montelukast, fenoterol and reproterol), and hepatoprotective drug silybin. The original pharmacological actions of these drugs could be helpful for the therapy of viral infection pneumonia. The natural products, such as flavanoids like neohesperidin, hesperidin, baicalin, kaempferol 3-*O*-rutinoside and rutin from different sources, andrographolide, neoandrographolide and 14-deoxy-11,12-didehydroandrographolide from *Andrographis paniculata*, and a series of xanthones from the plants of *Swertia* genus, with anti-virus, anti-bacteria and anti-inflammation activity could effectively interact with these targets of SARS-CoV-2. Therefore, the herbal medicines containing these compounds as major components might be meaningful for the treatment of SARS-CoV-2 infections.

For ACE2 target, although several compounds could bind with ACE2 through virtual screening in our studies, no compound was found to bind with the contact surface of ACE2–Spike complex, suggesting that these compounds are only the inhibitors of ACE2 enzyme activities, rather than inhibitors of ACE2 driven virus infections. Just like what described in recently published research[53], most of selected compounds are also unable to bind with the contact surface of ACE2–Spike complex. Actually, these potential ACE2 inhibitors may not be suitable to use as drugs for treating SARS-CoV-2 infection because the poor prognosis would be induced by the inhibition of ACE2 enzyme activities, for ACE2 was considered as a protective factor of lung injury[54].

For those targets which are difficult to find direct inhibitors, or non-druggable targets, just like Nsp1, Nsp3b, Nsp3c, and E-channel, etc., currently popular PROTAC technology may be a good strategy to degrade these proteins and then inhibit the virus. The potential binding compounds found in this study for these targets might be a good start point.

For Spike protein, we found only one compound, natural hesperidin was targeting the binding between Spike RBD and human ACE2. However, not like the ACE2 binding compounds, non-interface binding compound may still meaningful applications, considering that the fusion of CoVs membrane with host cell membrane need the big conformational change of remained Spike part after RBD removal[55]. Any small molecule bound to Spike at this time may interfere the re-folding of Spike therefore inhibits the viral infection process. Furthermore, small molecule that can target any part of Spike protein may be a good start point to design PROTAC based therapy.

Also, we dock existing anti-viral drugs with our targets, analyze the possible targets of each anti-viral drug horizontally, and analyze the drugs that may interact with 21 targets vertically. We analyzed 21 targets based on the docking results and found that Nsp3b, Nsp3c, Nsp7_Nsp8 complex, Nsp14, Nsp15, PLpro, 3CLpro, RdRp, helicase, E-channel, Spike and ACE2 are more likely to be therapeutic targets of anti-viral

drugs. The three targets Nsp3b, Nsp3c, and E-channel are screened more anti-viral drugs. This may be due to the model problem because of flexible small protein (Nsp3b and Nsp3c) or partial model (E-channel). Whether the screened anti-viral drugs really work on these targets needs further verification. We also do not recommend the application of new coronavirus pneumonia to compounds for which no target has been predicted.

The triphosphate nucleotide product of remdesivir, remdesivir-TP, competes with RdRp for substrate ATP, so it can interfere with viral RNA synthesis. Our docking results show that remdesivir-TP binds to SARS-CoV-2 RdRp, with a score of –112.8, and the docking results are consistent with its original anti-viral mechanism, so we think remdesivir may be good in treating SARS-CoV-2 pneumonia. In addition, remdesivir also predicted to bind with the human TMPRSS2, a protein facilitating the virus infection, this is a new discovery and provides ideas for subsequent research.

Chloroquine phosphate has shown better anti-SARS-CoV-2 effects in recent studies, but this drug has no clear target of action. In our docking results, chloroquine phosphate is predicted to possibly combine with Nsp3b and E-channel. But we need to do further experiments to verify this conclusion.

In response to the recently reported anti-AIDS drugs lopinavir and ritonavir tablets, which have a poor effect on the treatment of novel coronavirus pneumonia and have toxic side effects, we analyzed it in conjunction with the docking results. The molecular docking results show that ritonavir's possible target is Nsp3c or E-channel. Lopinavir's possible target is Nsp3b, Nsp3c, helicase, NRBD or E-channel. Some of these targets (such as Nsp3b, Nsp3c, E-channel) may be false positives due to the model inaccuracy for small flexible protein or partial model. For both lopinavir and ritonavir, we did not observe possible binding to major targets like 3CLpro, PLpro, RdRp, and so on. This docking result implies lopinavir and ritonavir tablets may not be suitable for treatment of SARS-CoV-2 infections.

The results of the entire article are based on computer virtual screening. We have not conduct further *in vivo* and *in vitro* anti-viral experiments yet, because we want to share our results with scientists in anti-SARS-CoV-2 research as soon as possible. Our subsequent research will try to solve the three-dimensional structures of all 24 proteins of SARS-CoV-2 and their drug complexes, providing more target information for drug intervention and long-term drug design, perform *in vivo* and *in vitro* evaluations for candidate drugs obtained in this study, and prepare for clinical trial applications.

Acknowledgments

We acknowledge support from National Mega-project for Innovative Drugs (grant number 2019ZX09721001-004-007, China), National Natural Science Foundation of China (NSFC, grant numbers U1803122, 81773637, 81773594, and U1703111).

Author contributions

Hua Li, Lixia Chen, Mengzhu Zheng and Xingzhou Li designed the research, performed virtual screening and revised the manuscript. Canrong Wu, Yang Liu, and Yueying Yang performed the bioinformatics analysis,

analyzed the virtual screening experiment data and drafted the manuscript. Peng Zhang, Wu Zhong, Yali Wang, Qiqi Wang, Yang Xu, and Mingxue Li checked the structures and made the Excel forms. All authors have read and approved the final manuscript.

Conflict of interests

The authors claim that the researchers in this study have no conflict of interest.

5. References

1. Cheng VC, Lau SK, Woo PC, Yuen KY. Severe acute respiratory syndrome coronavirus as an agent of emerging and reemerging infection. *Clin Microbiol Rev* 2007; **20**: 660-94.

2. Lee N, Hui D, Wu A, Chan P, Cameron P, Joynt GM, et al. A major outbreak of severe acute respiratory syndrome in Hong Kong. *N Engl J Med* 2003; **348**: 1986-94.

3. Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus AD, Fouchier RA. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J Med* 2012; **367**: 1814-20.

4. de Groot RJ, Baker SC, Baric RS, Brown CS, Drosten C, Enjuanes L, et al. Middle East respiratory syndrome coronavirus (MERS-CoV): announcement of the Coronavirus Study Group. *J Virol* 2013; **87**: 7790-2.

5. Lau SK, Woo PC, Li KS, Huang Y, Tsoi HW, Wong BH, et al. Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. *Proc Natl Acad Sci U S A* 2005; **102**: 14040-5.

6. Reusken CB, Haagmans BL, Müller MA, Gutierrez C, Godeke GJ, Meyer B, et al. Middle East respiratory syndrome coronavirus neutralising serum antibodies in dromedary camels: a comparative serological study. *Lancet Infect Dis* 2013; **13**: 859-66.

7. 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. Available from: https://doi.org/10.1056/nejmoa2001017.

8. Chen Y, Liu Q, Guo D. Emerging coronaviruses: genome structure, replication, and pathogenesis. *J Med Virol* 2020. Available from: https://doi.org/10.1002/jmv.25681.

9. 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. *The Lancet* 2020; **395**: 497-506.

10. Bosch BJ, van der Zee R, de Haan CA, Rottier PJ. The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. *J Virol* 2003; **77**: 8801-11.

11. Omrani AS, Saad MM, Baig K, Bahloul A, Abdul-Matin M, Alaidaroos AY, et al. Ribavirin and interferon alfa-2a for severe Middle East respiratory syndrome coronavirus infection: a retrospective cohort study. *Lancet Infect Dis* 2014; **14**: 1090-5.

12. Ge XY, Li JL, Yang XL, Chmura AA, Zhu G, Epstein JH, et al. Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature* 2013; **503**: 535-8.

13. Han DP, Penn-Nicholson A, Cho MW. Identification of critical determinants on ACE2 for SARS-CoV

entry and development of a potent entry inhibitor. *Virology* 2006; **350**: 15-25.

14. Li W, Moore MJ, Vasilieva N, Sui J, Wong SK, Berne MA, et al. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* 2003; **426**: 450-4.

15. Zhao J, Li K, Wohlford-Lenane C, Agnihothram SS, Fett C, Zhao J, et al. Rapid generation of a mouse model for Middle East respiratory syndrome. *Proc Natl Acad Sci U S A* 2014; **111**: 4970-5

16. Agrawal AS, Garron T, Tao X, Peng BH, Wakamiya M, Chan TS, et al. Generation of a transgenic mouse model of Middle East respiratory syndrome coronavirus infection and disease. *J Virol* 2015; **89**: 3659-70.

17. Zumla A, Chan JF, Azhar EI, Hui DS, Yuen KY. Coronaviruses—drug discovery and therapeutic options. *Nat Rev Drug Discov* 2016; **15**: 327-47.

18. Chan JF, Chan KH, Kao RY, To KK, Zheng BJ, Li CP, et al. Broad-spectrum antivirals for the emerging Middle East respiratory syndrome coronavirus. *J Infect* 2013; **67**: 606-16.

19. de Wilde AH, Jochmans D, Posthuma CC, Zevenhoven-Dobbe JC, van Nieuwkoop S, Bestebroer TM, et al. Screening of an FDA-Approved Compound Library Identifies Four Small-Molecule Inhibitors of Middle East Respiratory syndrome coronavirus replication in cell culture. *Antimicrob Agents Chemother* 2014; **14**: 4875-84.

20. Dyall J, Coleman CM, Hart BJ, Venkataraman T, Holbrook MR, Kindrachuk J, et al. Repurposing of clinically developed drugs for treatment of Middle East respiratory syndrome coronavirus infection. *Antimicrob Agents Chemother* 2014; **58**: 4885-93

21. Jaroszewski L, Rychlewski L, Li Z, Li W, Godzik A. FFAS03: a server for profile—profile sequence alignments. *Nucleic Acids Res* 2005; **33**(Web Server issue): W284-288.

22. Irwin JJ, Sterling T, Mysinger MM, Bolstad ES, Coleman RG. ZINC: a free tool to discover chemistry for biology. *J Chem Inf Model* 2012; **52**: 1757-68.

23. Abagyan R, Totrov M, Kuznetsov D. ICM—A new method for protein modeling and design_ applications to docking and structure prediction from the distorted native conformation. *J Comput Chem* 1994; **15**: 488-506

24. Wan Y, Shang J, Graham R, Baric RS, Li F. Receptor recognition by novel coronavirus from Wuhan: an analysis based on decade-long structural studies of SARS. *J Virol* 2020. Available from: https://doi.org/10.1128/jvi.00127-20.

25. Xu X, Chen P, Wang J, Feng J, Zhou H, Li X, et al. Evolution of the novel coronavirus from the ongoing Wuhan outbreak and modeling of its spike protein for risk of human transmission. *Sci China Life Sci* 2020. Available from: https://doi.org/10.1007/s11427-020-1637-5.

26. Li F, Li W, Farzan M, Harrison SC. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* 2005; **309**: 1864-8.

27. Harcourt BH, Jukneliene D, Kanjanahaluethai A, Bechill J, Severson KM, Smith CM, et al. Identification of severe acute respiratory syndrome coronavirus replicase products and characterization of papain-like protease activity. *J Virol* 2004; **78**: 13600-12.

28. 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**: 369-81.

29. Yuan L, Chen Z, Song S, Wang S, Tian C, Xing G, et al. p53 degradation by a coronavirus papain-like protease suppresses type I interferon signaling. *J Biol Chem* 2015; **290**: 3172-82.

30. Li SW, Wang CY, Jou YJ, Huang SH, Hsiao LH, Wan L, et al. SARS coronavirus papain-like protease inhibits the TLR7 signaling pathway through removing Lys63-linked polyubiquitination of TRAF3 and TRAF6. *Int J Mol Sci* 2016; **17**: 678.

31. Yang H, Xie W, Xue X, Yang K, Ma J, Liang W, et al. Design of wide-spectrum inhibitors targeting coronavirus main proteases. *PLoS Biol* 2005; **3**: e324.

32. Pillaiyar T, Manickam M, Namasivayam V, Hayashi Y, Jung SH. An overview of severe acute respiratory syndrome-coronavirus (SARS-CoV) 3CL protease inhibitors: peptidomimetics and small molecule chemotherapy. *J Med Chem* 2016; **59**: 6595-628.

33. Yang H, Yang M, Ding Y, Liu Y, Lou Z, Zhou Z, et al. The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *Proc Natl Acad Sci U S A* 2003; **100**: 13190-5.

34. 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.

35. Imbert I, Guillemot JC, Bourhis JM, Bussetta C, Coutard B, Egloff MP, et al. A second, non-canonical RNA-dependent RNA polymerase in SARS coronavirus. *EMBO J* 2006; **25**: 4933-42

36. Chu CK, Gadthula S, Chen X, Choo H, Olgen S, Barnard DL, et al. Antiviral activity of nucleoside analogues against SARS-coronavirus (SARS-coV). *Antivir Chem Chemother* 2006; **17**: 285-9

37. Ivanov KA, Ziebuhr J. Human coronavirus 229E nonstructural protein 13: characterization of duplex-unwinding, nucleoside triphosphatase, and RNA 5′-triphosphatase activities. *J Virol* 2004; **78**: 7833-8.

38. Shum KT, Tanner JA. Differential inhibitory activities and stabilisation of DNA aptamers against the SARS coronavirus helicase. *Chembiochem* 2008; **9**: 3037-45.

39. Jang KJ, Lee NR, Yeo WS, Jeong YJ, Kim DE. Isolation of inhibitory RNA aptamers against severe acute respiratory syndrome (SARS) coronavirus NTPase/Helicase. *Biochem Biophys Res Commun* 2008; **366**: 738-44.

40. Millet JK, Whittaker GR. Host cell proteases: critical determinants of coronavirus tropism and pathogenesis. *Virus Res* 2015; **202**: 120-34.

41. Xia S, Liu Q, Wang Q, Sun Z, Su S, Du L, et al. Middle East respiratory syndrome coronavirus (MERS-CoV) entry inhibitors targeting spike protein. *Virus Res* 2014; **194**: 200-10.

42. 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**: 12885-90.

43. Narayanan K, Huang C, Lokugamage K, Kamitani W, Ikegami T, Tseng CTK, 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**: 4471-9.

44. Forni D, Cagliani R, Mozzi A, Pozzoli U, Al-Daghri N, Clerici M, et al. Extensive positive selection drives the evolution of nonstructural proteins in lineage C betacoronaviruses. *J Virol* 2016; **90**: 3627-39.

45. Taylor JK, Coleman CM, Postel S, Sisk JM, Bernbaum JG, Venkataraman T, et al. Severe acute respiratory syndrome coronavirus ORF7a inhibits bone marrow stromal antigen 2 virion tethering through a novel mechanism of glycosylation interference. *J Virol* 2015; **89**: 11820-33.

46. Glowacka I, Bertram S, Muller MA, Allen P, Soilleux E, Pfefferle S, et al. Evidence that TMPRSS2 activates the severe acute respiratory syndrome coronavirus spike protein for membrane fusion and reduces viral control by the humoral immune response. *J Virol* 2011; **85**: 4122-34.

47. Holshue ML, De Bolt C, Lindquist S, Lofy KH, Wiesman J, Bruce H, et al. First case of 2019 novel coronavirus in the United States. *N Engl J Med* 2020. Available from: https://doi.org/10.1056/nejmoa2001191.

48. 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. Available from: https://doi.org/10.1038/s41422-020-0282-0.

49. van der Linden L, Vives-Adrián L, Selisko B, Ferrer-Orta C, Liu X, Lanke K, et al. The RNA template channel of the RNA dependent RNA polymerase as a target for development of antiviral therapy of multiple genera within a virus family. *PLoS Pathog* 2015; **11**: e1004733.

50. Muegge I, Martin YC. A general and fast scoring function for proteinligandinteractions: a simplified potential approach. *J Med Chem* 1999; **42**: 791-804.

51. Neves MA, Totrov M, Abagyan R. Docking and scoring with ICM: the bench marking results and strategies for improvement. *J Comput Aided Mol Des* 2012; **26**: 675-86.

52. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, et al. Clinical characteristics of 2019 novel coronavirus infection in China. *medRxiv* 2020. Available from: https://doi.org/10.1101/2020.02.06.20020974.

53. Chen H, Du Q. Potential natural compounds for preventing 2019-nCoV infection. *Preprints* 2020; 2020010358.

54. Ye R, Liu Z. ACE2 exhibits protective effects against LPS-induced acute lung injury in mice by inhibiting the LPS-TLR4 pathway. *Exp Mol Pathol*. 2019; **113**:104350.

55. Heald-Sargent T, Gallagher I. Ready. Set, Fuse! The coronavirus spike protein and acquisition of fusion competence. *Viruses* 2012; **4**: 557-80.

Figure captions

Figure 1 Phylogenetic analysis and sequence alignment of 23 coronaviruses from different hosts. (A) Homology tree of 23 nucleotide sequences alignment. The homology of each branch is marked in red. (B) The evolutionary history was inferred using the Neighbor-Joining method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) is shown above the branches.

Figure 2 Overall genome and protein analysis of SARS-CoV-2. (A) Genome of SARS-CoV-2. (B) The large replicase polyprotein PP1ab. The purple rectangle represents the replicase polyprotein PP1ab which contains 7096 amino acids. The position of red triangle means cleavage site of PLpro, and position of yellow triangle means the site of 3CLpro, the numbers below the purple rectangle represent the residues that bound the individual proteins or domains. Orange rectangle below the individual proteins or domains means exactly homology 3D structure were detected. (C) Four structure proteins are represented in the cartoon pattern of coronavirus.

Figure 3 Sequence alignment of the proteins from SARS-CoV-2 with other CoVs. (A) Sequence alignment of PLpro, 3CLpro, RdRp among SARS-CoV, SARS-CoV-2 and MERS CoV. (B) The computational structure of SARS-CoV-2 3CLpro (shown as pink cartoon) modeled from SARS 3CLpro (3WA0) aligned with crystal structure of SARS-CoV-2 3CLpro (6LU7) (shown in blue cartoon). The computational model of the of SARS-CoV-2 3CLpro showed a Cα RMSD of 0.471 Å on the overall structure compared to the SARS-CoV-2 3CLpro structure, and that showed a Cα RMSD of 0.126 Å on the substrate binding region. (C) Sequence alignment of spike protein receptor binding domain among SARS-CoV, bat-CoVRaTG13 and SARS-CoV-2. "Δ" represents residues previously identified as importantly interact with human ACE2 from SARS. Black arrow marked amino acids that are different between SARS-CoV-2 and bat-CoVRaTG13.

Figure 4 Low-energy binding conformations of ribavirin bound to PLpro generated by molecular docking. (A) Superimposing of 2019-nCoV PLpro (yellow) with the crystal structure of SARS PLpro (orange) (PDB code 3e9s). (B) Detailed view of rivavirin binding in the active site of the enzyme (Supporting PDB file SARS_CoV-2_PLpro_homo_Ribavirin.pdb).

Figure 5 Low-energy binding conformations of Montelukast bound to 3CLpro generated by molecular docking. (A) Montelukast was fitted well in the active pocket of SARS-CoV-2

3CLpro, 3CLpro was shown as electrostatic surface model. (B) Detailed view of montelukastin binding in the active pocket of 3CLpro (Supporting PDB file SARS_CoV-2_3CLpro_homo_Montelukast.pdb).

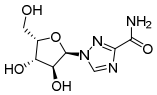
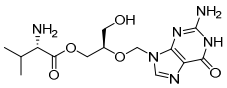
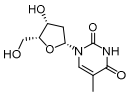
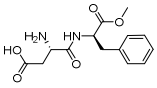
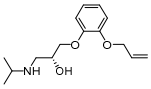
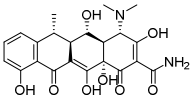
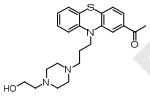
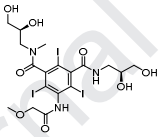
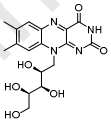
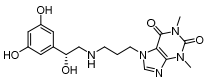
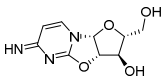
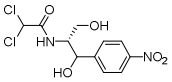
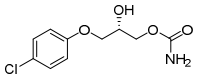
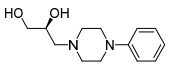
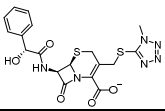
Figure 6 Low-energy binding conformation of hesperidin bound to spike RBD generated by molecular docking. (A) Hesperidinwas fitted into the shallow pocket in the surface of SARS-CoV-2 Spike RBD. (B) Detailed view of hesperidinbinding in the pocket of Spike RBD. (C) Hesperidin blocks the interface of ACE2 and Spike RBD binding (Supporting PDB file SARS_CoV-2 _Spike_RBD_homo_Hesperidin.pdb).

Figure 7 Low-energy binding conformation of remdesivir bound to SARS-CoV-2 RdRp generated by molecular docking. (A) Remdesivir was fitted in the bottom of the RNA template channel (top view). (B) Remdesivir was fitted in the bottom of the RNA template channel (bottom view). (C) Detailed view of remdesivir binding with RdRp (Supporting PDB file SARS_CoV-2_RdRp_homo_Remdesivir.pdb).

Figure 8 Low-energy binding conformation of remdesivir bound to human TMPRSS2 generated by molecular docking. (A) Remdesivir was fitted in an allosteric pocket far away from the enzyme active center (bottom). (B) Detailed view of remdesivir binding with TMPRSS2 (Supporting PDB file HS_TMPRSS2_homo_Remdesivir.pdb).

Tables

Table 1 Potential PLpro inhibitors from ZINC drug database (compounds in the table are ranked according to their docking scores).

No.	Drug name	Structure	Pharmacological function
1	Ribavirin		Anti-virus
2	Valganciclovir		Anti-virus
3	*β*-Thymidine		Anti-virus
4	Aspartame		Non-carbohydrate sweetener
5	Oxprenolol		Anti-hypertensive, anti-angina, anti-arrhythmic effects
6	Doxycycline		Anti-bacterial effect
7	Acetophenazine		Anti-psychotic effect
8	Iopromide		Low osmolar, non-ionic contrast agent
9	Riboflavin		Treatment of vitamin B2 deficiency
10	Reproterol		Treatment of bronchial asthma
11	2,2′-Cyclocytidine		Anti-tumor
12	Chloramphenicol		Anti-bacterial effect
13	Chlorphenesin carbamate		Muscle relaxant effect
14	Levodropropizine		Anti-tussive effect
15	Cefamandole		Anti-bacterial effect

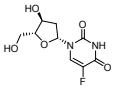
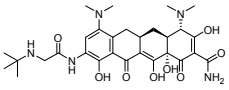
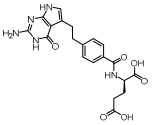
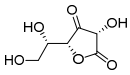
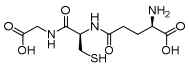
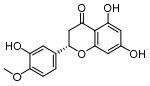
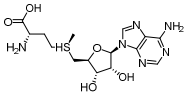
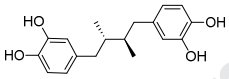
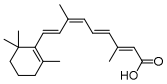
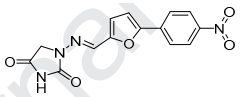
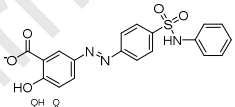
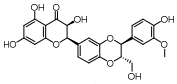
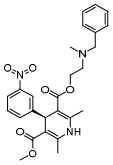
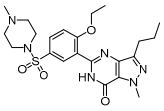
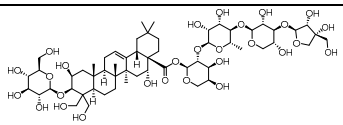
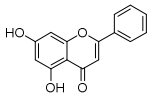
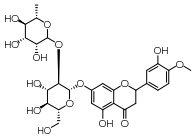
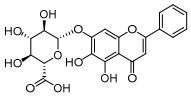
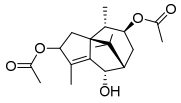
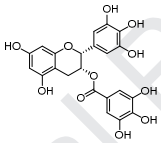
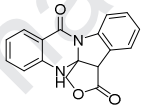
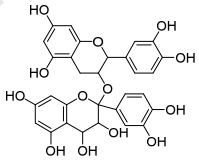
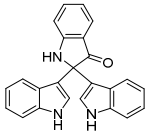
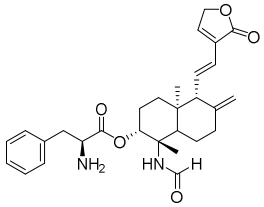
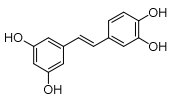
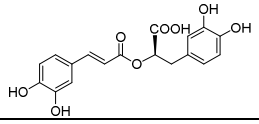
16	Floxuridine		Anti-tumor
17	Tigecycline		Anti-bacterial
18	Pemetrexed		Anti-tumor
19	L(+)-Ascorbic acid		Anti-scorbutic
20	Glutathione		Treatment of hepatic disease
21	Hesperetin		Anti-oxidation, anti-virus, anti-bacteria
22	Ademetionine		Cholagogue
23	Masoprocol		Treatment of actinic keratosis
24	Isotretinoin		Anti-tumor
25	Dantrolene		Muscle relaxant
26	Sulfasalazine		Anti-bacterial effect
27	Silybin		Hepatoprotective effect
28	Nicardipine		Anti-hypertensive effect
29	Sildenafil		Treatment of erectile dysfunction

Table 2 Potential PLpro inhibitors from in-house natural product database.

No.	Compound name	Structure	Pharmacological function	Source
1	Platycodin D		Anti-tumor, anti-inflammatory effect	*Platycodon grandiflorus*
2	Chrysin		Anti-virus, anti-inflammatory effect	*Scutellaria baicalensis*
3	Neohesperidin		Anti-tumor, anti-allergic effect	*Citrus aurantium* L.
4	Baicalin		Anti-tumor, anti-inflammatory, anti-bacterial, anti-virus effect	*Scutellaria baicalensis*
5	Sugetriol-3,9-diacetate		Anti-HBV, Anti-HSV-1	*Cyperus rotundus*
6	(−)-Epigallocatechin gallate		Anti-oxidation, anti-tumor, treatment of depression	*Camellia sinensis*
7	Phaitanthrin D		Anti-virus	*Isatis indigotica* Fort.
8	2-(3,4-Dihydroxyphenyl)-2-[[2-(3,4-dihydroxyphenyl)-3,4-dihydro-5,7-dihydroxy-2*H*-1-benzopyran-3-yl]oxy]-3,4-dihydro-2*H*-1-benzopyran-3,4,5,7-tetrol		Anti-oxidant, anti-inflammatory, anti-tumor	*Vitis vinifera*
9	2,2-Di(3-indolyl)-3-indolone		Anti-virus	*Isatis indigotica* Fort.
10	(*S*)-(1*S*,2*R*,4a*S*,5*R*,8a*S*)-1-Formamido-1,4a-dimethyl-6-methylene-5-((*E*)-2-(2-oxo-2,5-dihydrofuran-3-yl)ethenyl)decahydronaphthalen-2-yl-2-amino-3-phenylpropanoate		Anti-virus, anti-inflammatory effect	Andrographolide derivatives
11	Piceatannol		Anti-tumor, anti-virus effect	*Vitis vinifera*
12	Rosmarinic acid		Anti-virus, Anti-oxidant	*Salvia verticillata* L.

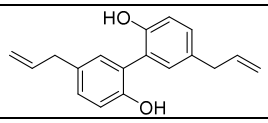
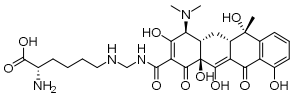
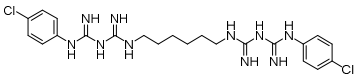
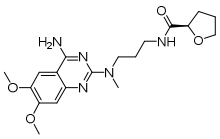
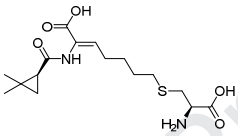
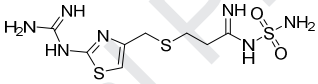
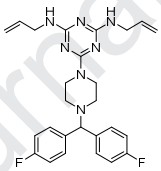
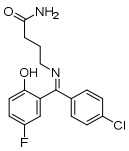
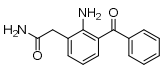
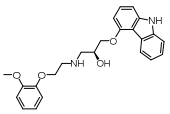
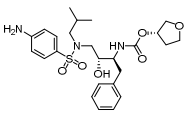
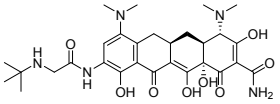
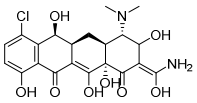
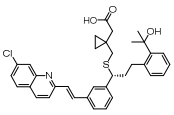
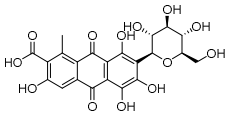
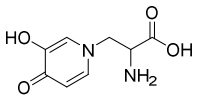
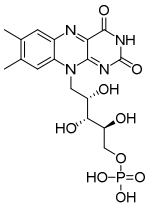
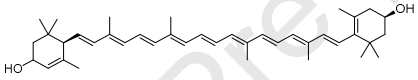
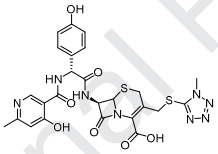
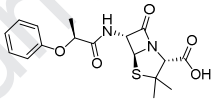
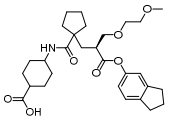
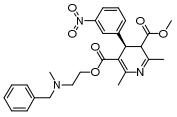
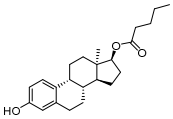
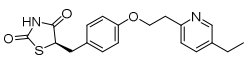
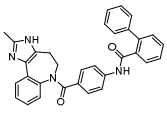
13	Magnolol		Anti-tumor, Anti-microbial effect	*Magnolia officinalis*

Table 3 Potential 3CLpro inhibitors from ZINC drug database.

No.	Drug name	Structure	Pharmacological functions
1	Lymecycline		Anti-bacterial effect
2	Chlorhexidine		Anti-bacterial effect
3	Alfuzosin		Anti-hypertensive agent, benign prostatic hyperplasia
4	Cilastatin		Renal peptidase inhibition
5	Famotidine		Anti-ulcerative activity
6	Almitrine		Treatment of cognitive and chronic sensory nerve impairment
7	Progabide		Anti-epileptic effect
8	Nepafenac		Treatment of pain and inflammation associated with cataract surgery
9	Carvedilol		Vasodilator effect
10	Amprenavir		HIV-1 protease inhibition
11	Tigecycline		Anti-bacterial effect

12	Demeclocycline		Anti-bacterial effect
13	Montelukast		Anti-allergic, anti-asthmatic effects
14	Carminic acid		Food additive
15	Mimosine		Depilatory effect
16	Flavin mononucleotide		Electron transference in biological oxidation
17	Lutein		Vision protection, Anti-oxidation
18	Cefpiramide		Anti-bacterial effect
19	Phenethicillin		Anti-bacterial effect
20	Candoxatril		Anti-hypertensive effect
21	Nicardipine		Anti-hypertensive effect
22	Estradiol valerate		Treatment of estrogen deficiency
23	Pioglitazone		Anti-diabetic effect
24	Conivaptan		Treatment of hyponatremia

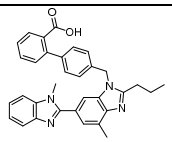
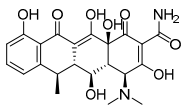
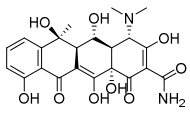
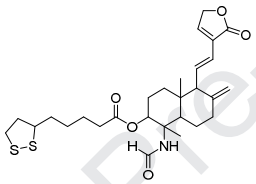
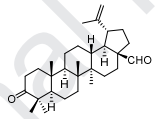
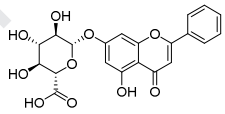
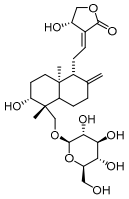
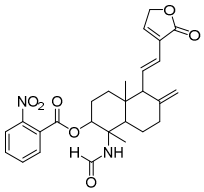
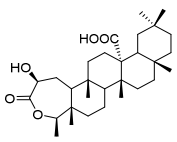
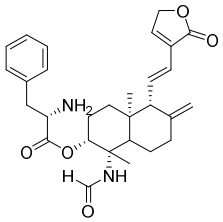
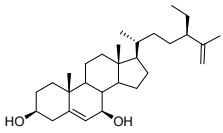
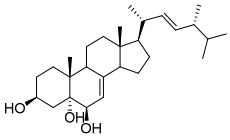
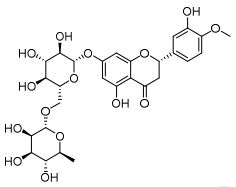
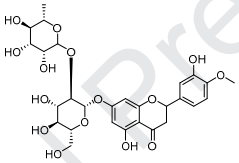
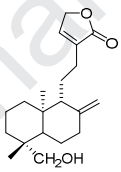
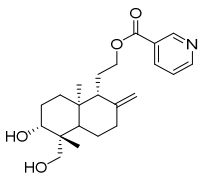
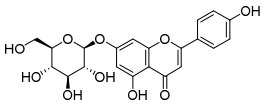
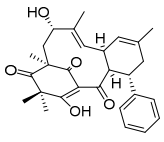
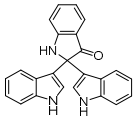
25	Telmisartan		Anti-hypertensive effect
26	Doxycycline		Anti-bacterial effect
27	Oxytetracycline		Anti-bacterial effect

Table 4 Potential 3CLpro inhibitors from in-house natural product database.

No.	Drug name	Structure	Pharmacological functions	Source
1	(1*S*,2*R*,4a*S*,5*R*,8a*S*)-1-Formamido-1, 4a-dimethyl-6-methylene-5-((*E*)-2-(2 -oxo-2,5-dihydrofuran-3-yl)ethenyl) decahydronaphthalen-2-yl 5-((*R*)-1,2-dithiolan-3-yl) pentanoate		Anti-virus, anti-inflammatory effect	Andrographolide derivatives
2	Betulonal		Anti-HIV-1	*Cassine xylocarpa*
3	Chrysin-7-*O*-*β*-glucuronide		Anti-virus, anti-inflammatory effect	*Scutellaria baicalensis*
4	Andrographiside		Anti-virus, anti-inflammatory effect	*Andrographis paniculata*
5	(1*S*,2*R*,4a*S*,5*R*,8a*S*)-1-Formamido-1, 4a-dimethyl-6-methylene-5-((*E*)-2-(2 -oxo-2,5-dihydrofuran-3-yl)ethenyl) decahydronaphthalen-2-yl 2-nitrobenzoate		Anti-virus, anti-inflammatory effect	Andrographolide derivatives
6	2*β*-Hydroxy-3, 4-seco-friedelolactone-27-oic acid		Anti-virus	*Viola diffusa*

7	(*S*)-(1*S*,2*R*,4a*S*,5*R*,8a*S*)-1-Formamido-1,4a-dimethyl-6-methylene-5-((*E*)-2-(2-oxo-2,5-dihydrofuran-3-yl)ethenyl)decahydronaphthalen-2-yl-2-amino-3-phenylpropanoate		Anti-virus, anti-inflammatory effect	Andrographolide derivatives
8	Isodecortinol		Anti-virus	*Viola diffusa*
9	Cerevisterol		Anti-virus	*Viola diffusa*
10	Hesperidin		Anti-inflammatory, anti-oxidant effect	*Citrus aurantium* L.
11	Neohesperidin		Anti-tumor, anti-allergic effect	*Citrus aurantium* L.
12	Andrograpanin		Anti-virus, anti-inflammatory effect	*Andrographis paniculata*
13	2-((1*R*,5*R*,6*R*,8a*S*)-6-Hydroxy-5-(hydroxymethyl)-5,8a-dimethyl-2-methylenedecahydronaphthalen-1-yl)ethyl benzoate		Anti-virus, anti-inflammatory effect	Andrographolide derivatives
14	Cosmosiin		Anti-inflammatory, Anti-oxidant, anti-HIV effect	*Scutellaria baicalensis*
15	Cleistocaltone A		Anti-virus	*Cleistocalyx operculatus*
16	2,2-Di(3-indolyl)-3-indolone		Anti-virus	*Isatis indigotica* Fort.

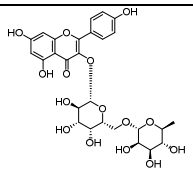
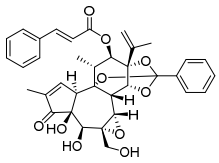
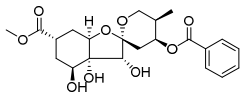
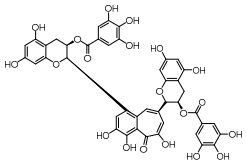
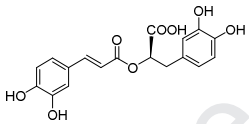
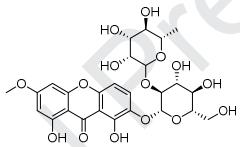
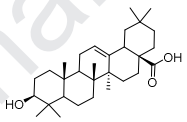
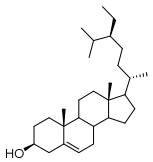
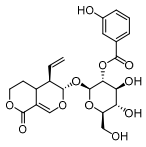
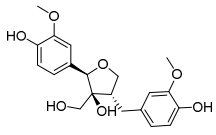
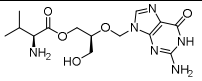
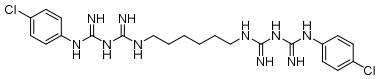
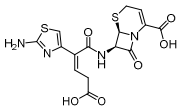
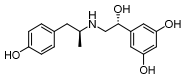
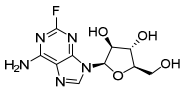
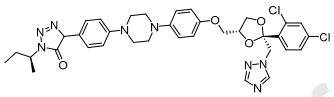
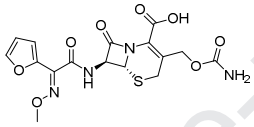
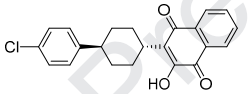
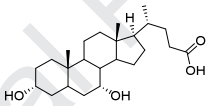
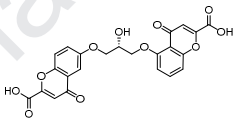
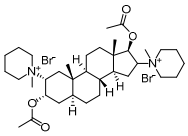
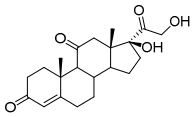
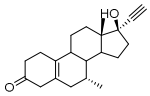
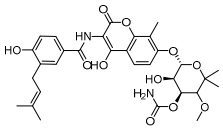
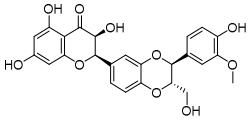
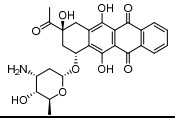
17	Biorobin		Anti-virus	*Ficus benjamina*
18	Gnidicin		Anti-tumor	*Gnidia lamprantha*
19	Phyllaemblinol		Anti-virus	*Phyllanthus emblica*
20	Theaflavin 3,3′-di-*O*-gallate		Anti-oxidant effect, anti-tumor, anti-virus	*Camellia sinensis*
21	Rosmarinic acid		Anti-viral, Anti-oxidant	*Salvia verticillata* L.
22	Kouitchenside I		Anti-virus, anti-inflammatory effect	*Swertia kouitchensis*
23	Oleanolic acid		Anti-virus, anti-inflammatory effect	*Swertia kouitchensis*
24	Stigmast-5-*en*-3-ol		Anti-virus, anti-inflammatory effect	*Swertia binchuanensis*
25	Deacetylcentapicrin		Anti-virus, anti-inflammatory effect	*Swertia macrosperma*
26	Berchemol		Anti-virus, anti-inflammatory effect	*Swertiabi maculata*

Table 5 Potential RdRp inhibitors from ZINC drug database.

No.	Drug name	Structure	Pharmacological function
1	Valganciclovir		Anti-virus
2	Chlorhexidine		Anti-bacterial effect
3	Ceftibuten		Anti-bacterial effect
4	Fenoterol		Treatment of bronchial asthma
5	Fludarabine		Anti-tumor
6	Itraconazole		Anti-fungal effect
7	Cefuroxime		Anti-bacterial effect
8	Atovaquone		Anti-malaria
9	Chenodeoxycholic acid		Dissolving gallstones
10	Cromolyn		Treatment of allergic asthma
11	Pancuronium bromide		Muscle relaxant effect
12	Cortisone		Anti-allergic effect
13	Tibolone		Contraceptive effect
14	Novobiocin		Anti-bacterial effect
15	Silybin		Hepatoprotective effect
16	Idarubicin		Anti-tumor

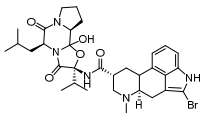
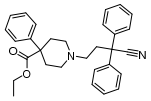
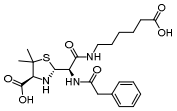
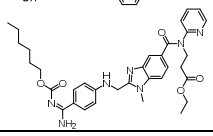
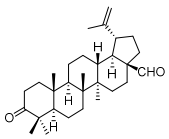
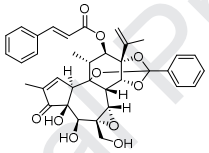
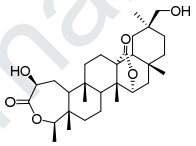
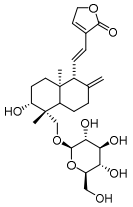
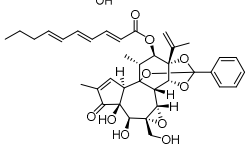
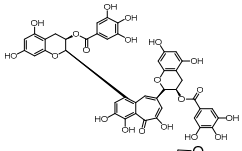
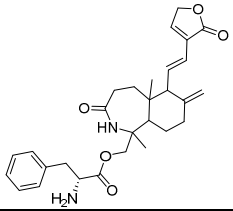
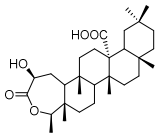
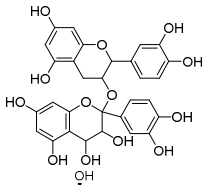
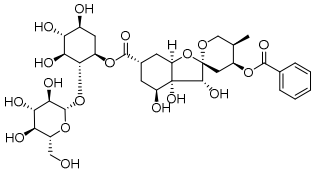
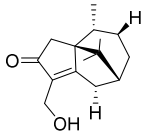
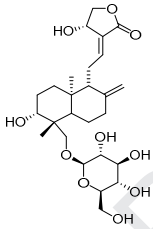
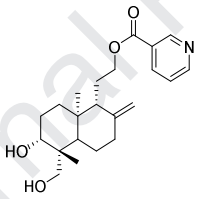
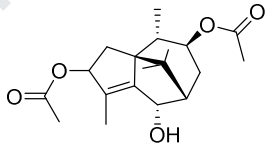
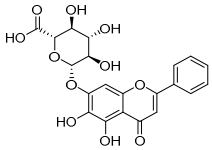
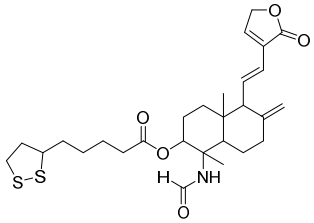
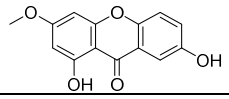
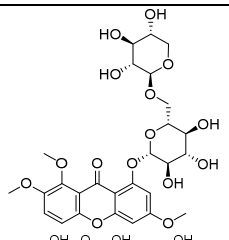
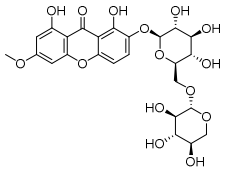
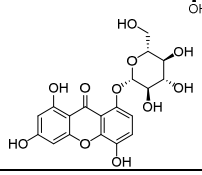
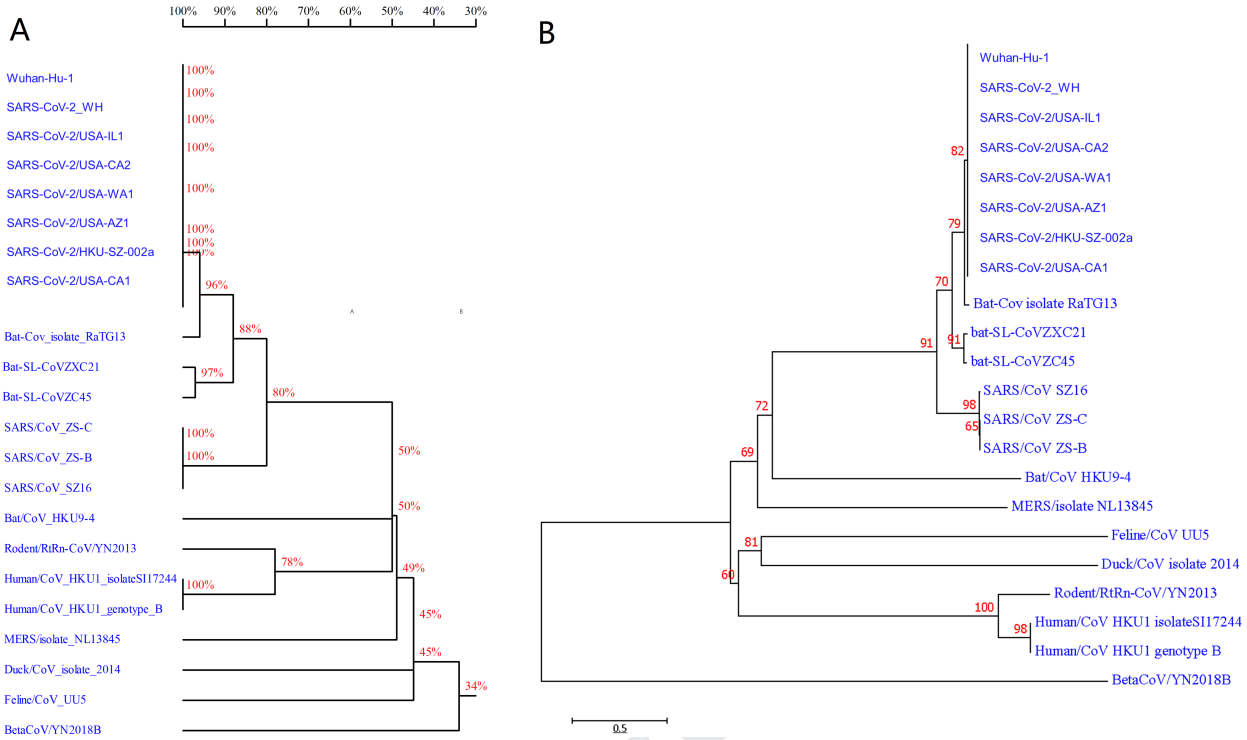
17	Bromocriptine		Treatment of amenorrhea
18	Diphenoxylate		Treatment of diarrhea and chronic enteritis
19	Benzylpenicilloyl G		Skin-testing reagent to detect penicillin allergy
20	Dabigatran etexilate		Anti-coagulant effect

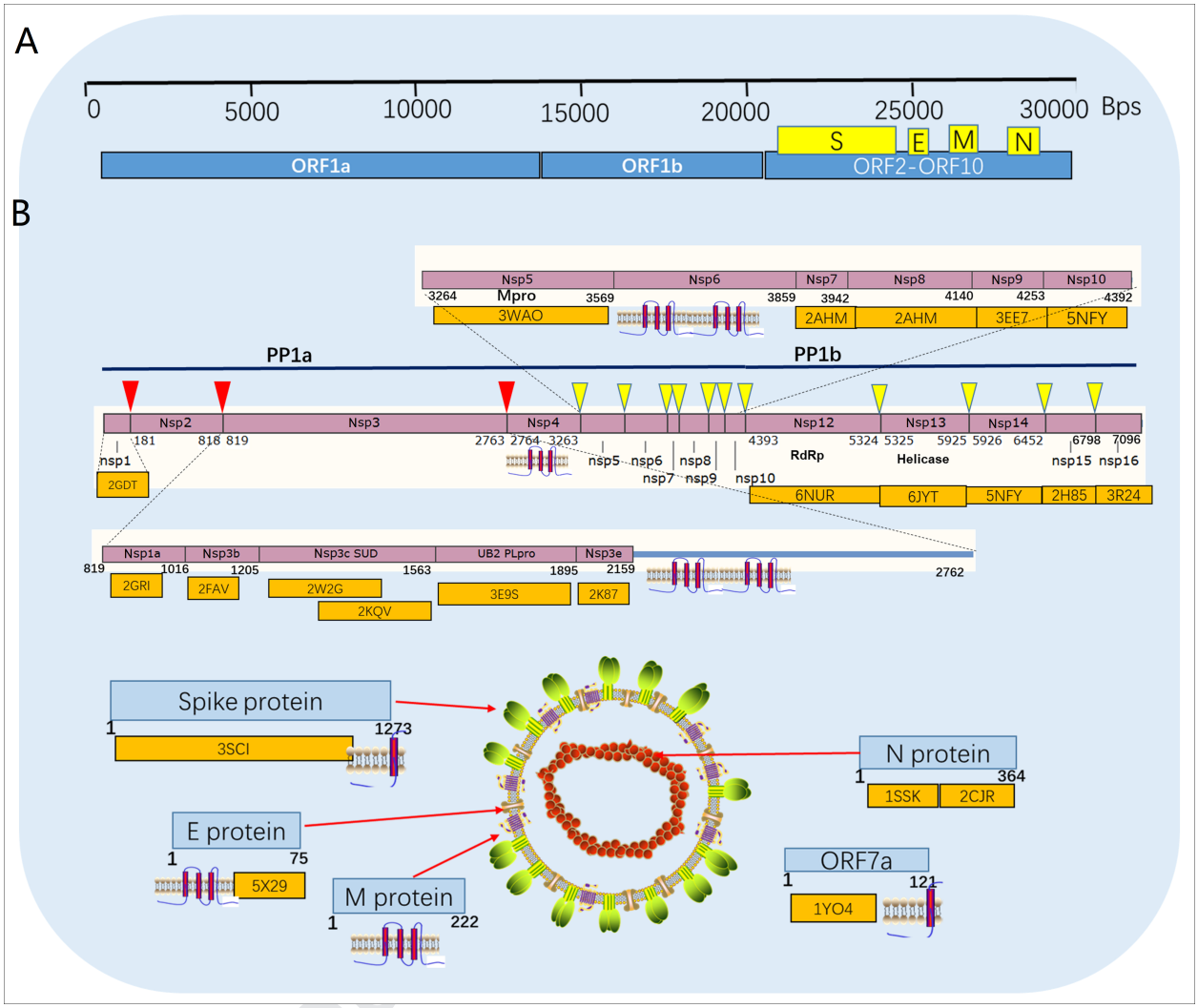
Table 6 Potential RdRp inhibitors from in-house natural product database.

No.	Drug name	Structure	Pharmacological function	Source
1	Betulonal		Anti-HIV-1	*Cassine xylocarpa*
2	Gnidicin		Anti-tumor	*Gnidia lamprantha*
3	2*β*,30*β*-Dihydroxy-3,4-seco-friedelolactone-27-lactone		Anti-virus	*Viola diffusa*
4	14-Deoxy-11,12-didehydroandrographolide		Anti-virus, anti-inflammatory effect	*Andrographis paniculata*
5	Gniditrin		Anti-tumor	*Gnidia lamprantha*
6	Theaflavin 3,3′-di-*O*-gallate		Anti-oxidant, anti-tumor, anti-virus	*Camellia sinensis*
7	(*R*)-((1*R*,5a*S*,6*R*,9a*S*)-1,5a-Dimethyl-7-methylene-3-oxo-6-((*E*)-2-(2-oxo-2,5-dihydrofuran-3-yl)ethenyl)decahydro-1*H*-benzo[*c*]azepin-1-yl)methyl 2-amino-3-phenylpropanoate		Anti-virus, anti-inflammatory effect	Andrographolide derivatives

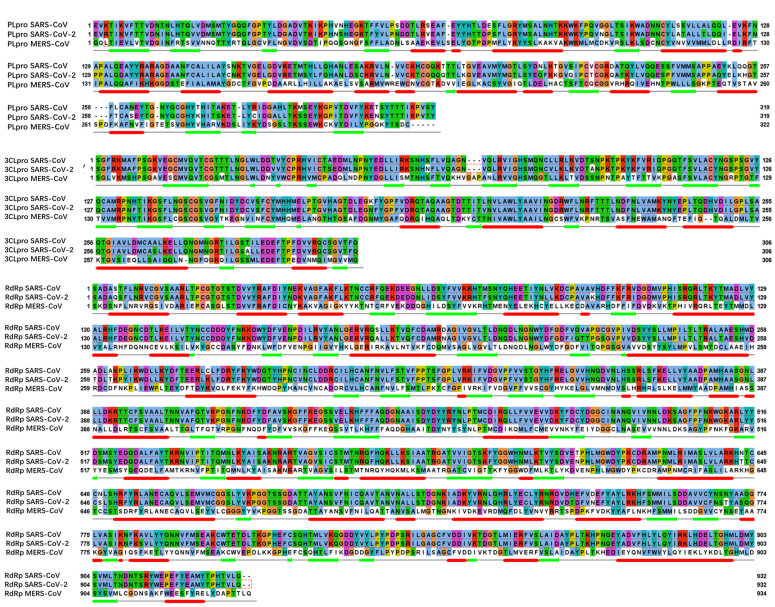
8	2*β*-Hydroxy-3, 4-seco-friedelolactone-27-oic acid		Anti-virus	*Viola diffusa*
9	2-(3,4-Dihydroxyphenyl)-2-[[2-(3,4-dihydroxyphenyl)-3,4-dihydro-5,7-dihydroxy-2*H*-1-benzopyran-3-yl]oxy]-3,4-dihydro-2*H*-1-benzopyran-3,4,5,7-tetrol		Anti-oxidant, anti-inflammatory, anti-tumor	*Vitis vinifera*
10	Phyllaemblicin B		Anti-virus	*Phyllanthus emblica*
11	14-hydroxycyperotundone		Anti-HBV	*Cyperus rotundus*
12	Andrographiside		Anti-virus, anti-inflammatory effect	*Androgra phispaniculata*
13	2-((1*R*,5*R*,6*R*,8a*S*)-6-Hydroxy-5-(hydroxymethyl)-5,8a-dimethyl-2-methylenedecahydronaphthalen-1-yl)ethyl benzoate		Anti-virus, anti-inflammatory effect	Andrographolide derivatives
14	Sugetriol-3,9-diacetate		Anti-HBV	*Cyperus rotundus*
15	Baicalin		Anti-tumor, anti-inflammatory, anti-bacterial, anti-virus effect	*Scutellaria baicalensis*
16	(1*S*,2*R*,4a*S*,5*R*,8a*S*)-1-Formamido-1,4a-dimethyl-6-methylene-5-((*E*)-2-(2-oxo-2,5-dihydrofuran-3-yl)ethenyl)decahydronaphthalen-2-yl 5-((*R*)-1,2-dithiolan-3-yl) pentanoate		Anti-virus, anti-inflammatory effect	Andrographolide derivatives
17	1,7-Dihydroxy-3-methoxyxanthone		Anti-virus, anti-inflammatory effect	*Swertia pseudochinensis*

18	1,2,6-Trimethoxy-8-[(6-*O*-*β*-D-xylopyranosyl-*β*-D-glucopyranosyl)oxy]-9*H*-xanthen-9-one		Anti-virus, anti-inflammatory effect	*Swertia mussotii*
19	1,8-Dihydroxy-6-methoxy-2-[(6-*O*-*β*-D-xylopyranosyl-*β*-D-glucopyranosyl)oxy]-9*H*-xanthen-9-one		Anti-virus, anti-inflammatory effect	*Swertia kouitchensis*
20	8-(*β*-D-Glucopyranosyloxy)-1,3,5-trihydroxy-9*H*-xanthen-9-one		Anti-virus, anti-inflammatory effect	*Swertia mussotii*

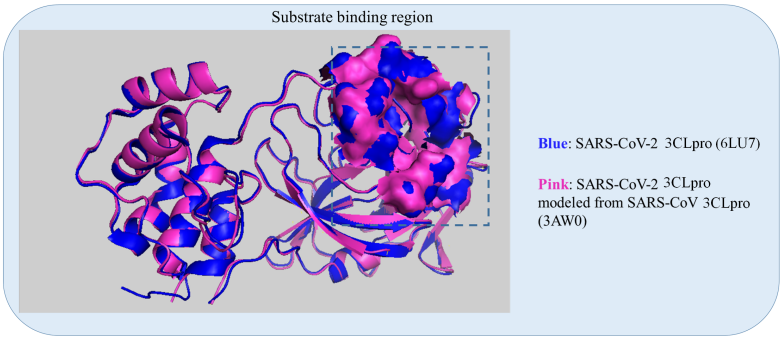



A
0
5000
10000
15000
20000
25000
30000
Bps
S
E
M
N
ORF1a
ORF1b
ORF2-ORF10
B
Nsp5
Nsp6
Nsp7
Nsp8
Nsp9
Nsp10
3264
Mpro
3569
3859
3942
4140
4253
4392
3WAO
2AHM
2AHM
3EE7
5NFY
PP1a
PP1b
Nsp2
Nsp3
Nsp4
Nsp12
Nsp13
Nsp14
181
818 819
2763 2764 3263
4393
5324 5325
5925 5926
6452
6798
7096
nsp1
nsp5
nsp6
nsp7
nsp8
nsp9
nsp10
RdRp
Helicase
nsp15
nsp16
2GDT
6NUR
6JYT
5NFY
2H85
3R24
Nsp1a
Nsp3b
Nsp3c SUD
UB2 PLpro
Nsp3e
819
1016
1205
1563
1895
2159
2762
2GRI
2FAV
2W2G
2KQV
3E9S
2K87
Spike protein
1
1273
3SCI
N protein
1
364
1SSK
2CJR
E protein
1
75
5X29
M protein
1
222
ORF7a
1
121
1YO4

A


PLpro SARS-CoV
PLpro SARS-CoV-2
PLpro MERS-CoV
3CLpro SARS-CoV
3CLpro SARS-CoV-2
3CLpro MERS-CoV
RdRp SARS-CoV
RdRp SARS-CoV-2
RdRp MERS-CoV

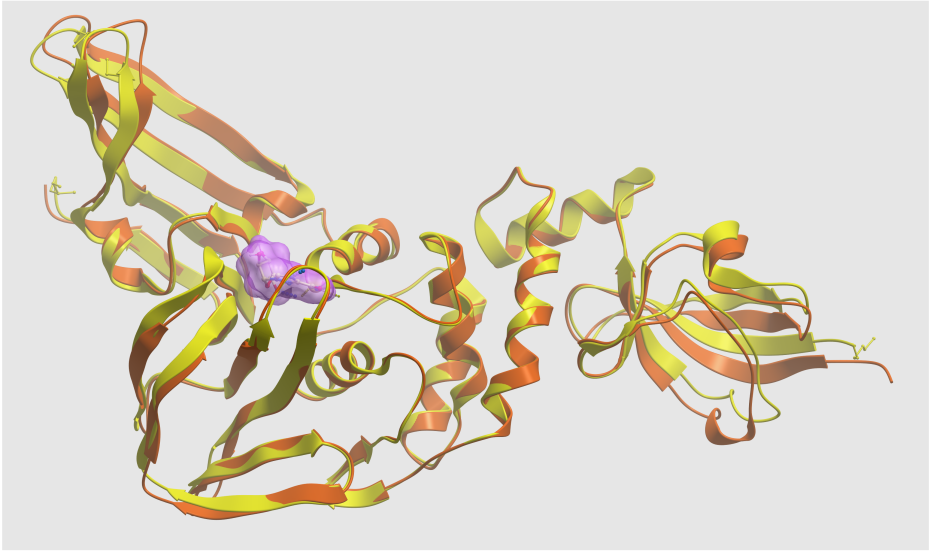
B


Substrate binding region
Blue: SARS-CoV-2 3CLpro (6LU7)
Pink: SARS-CoV-2 3CLpro modeled from SARS-CoV 3CLpro (3AW0)

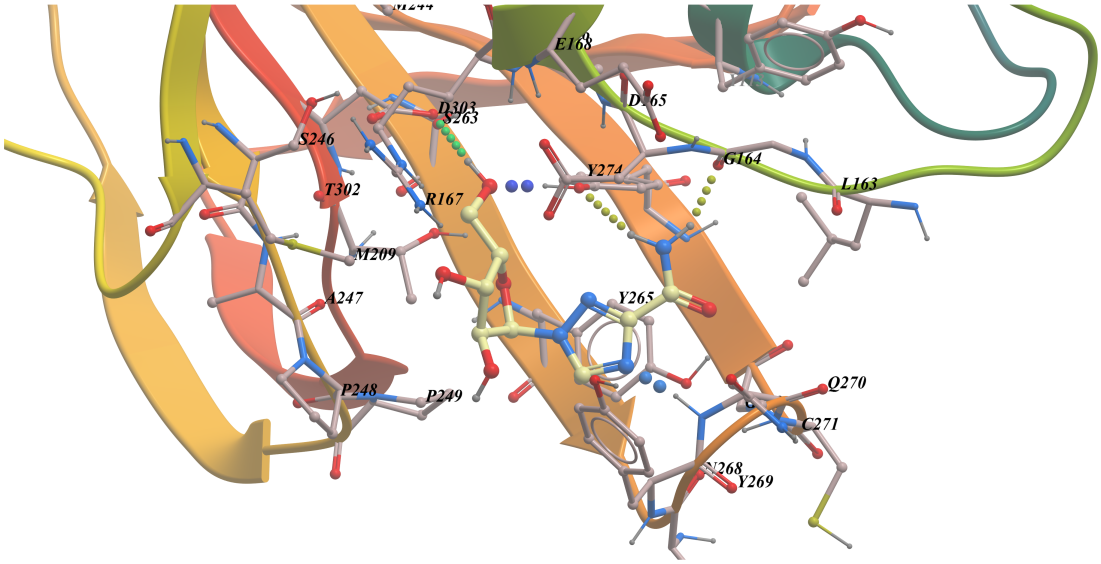
C


Spike RBD SARS-CoV
Spike RBD bat-CoVRaTG13
Spike RBD SARS-CoV-2

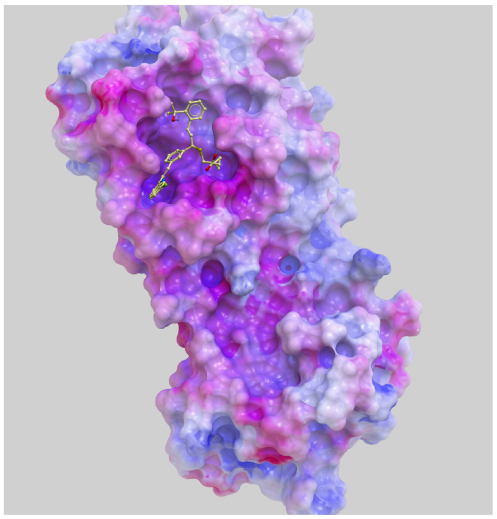
A



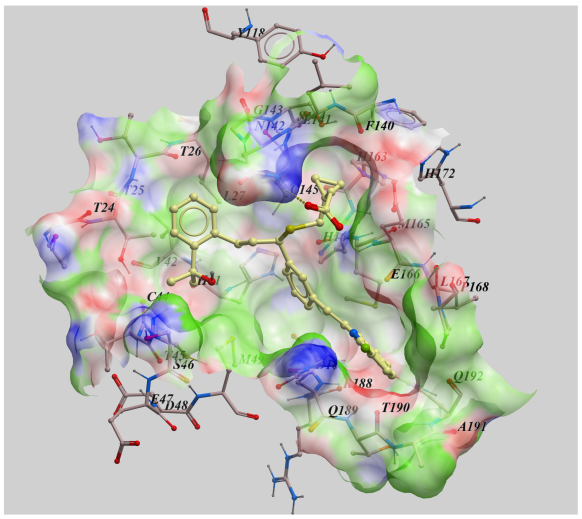
B


E168
D165
S263
S246
G164
Y274
T302
R167
L163
M209
Y265
A247
P248
P249
Q270
C271
Y269

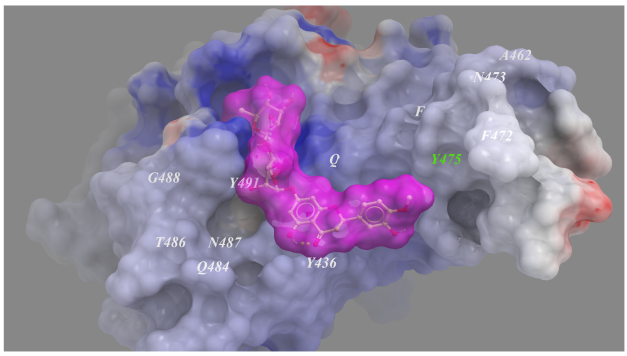
A



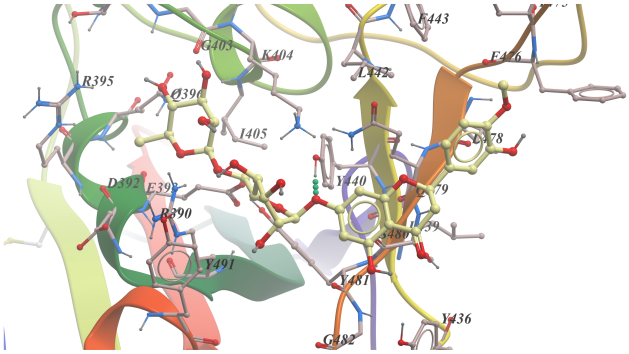
B



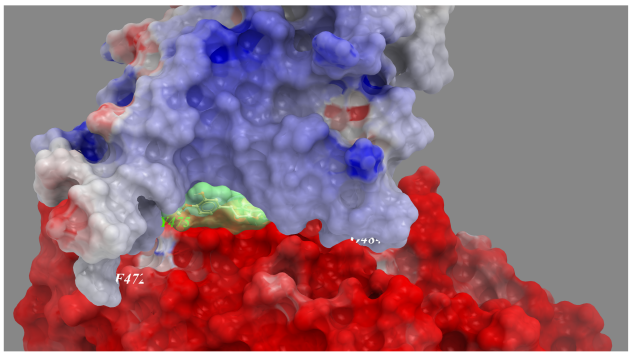
A

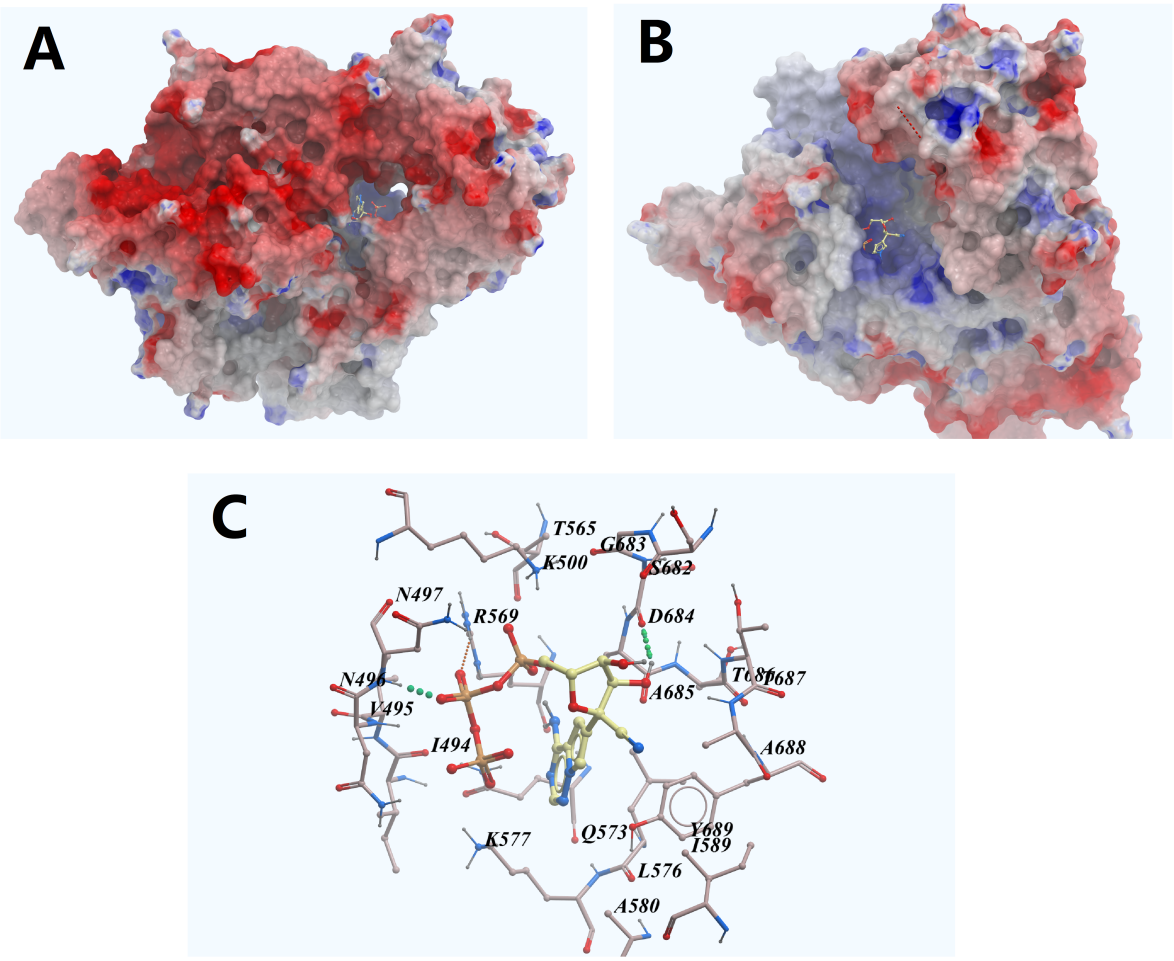


B

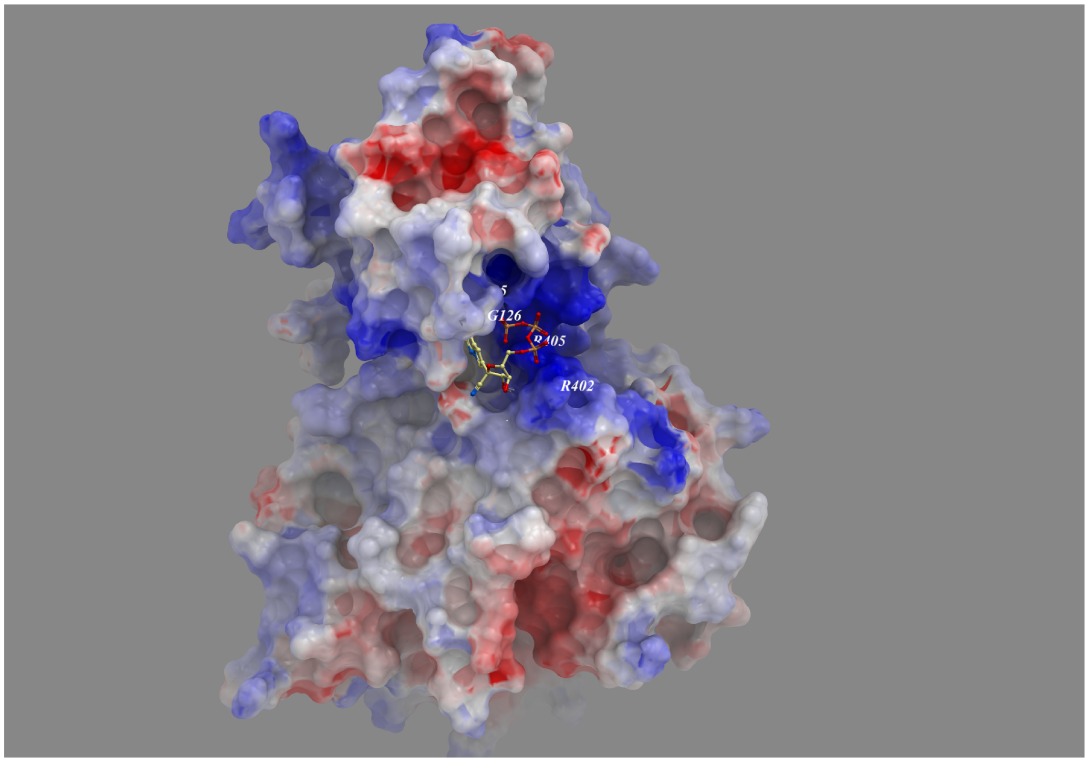


C

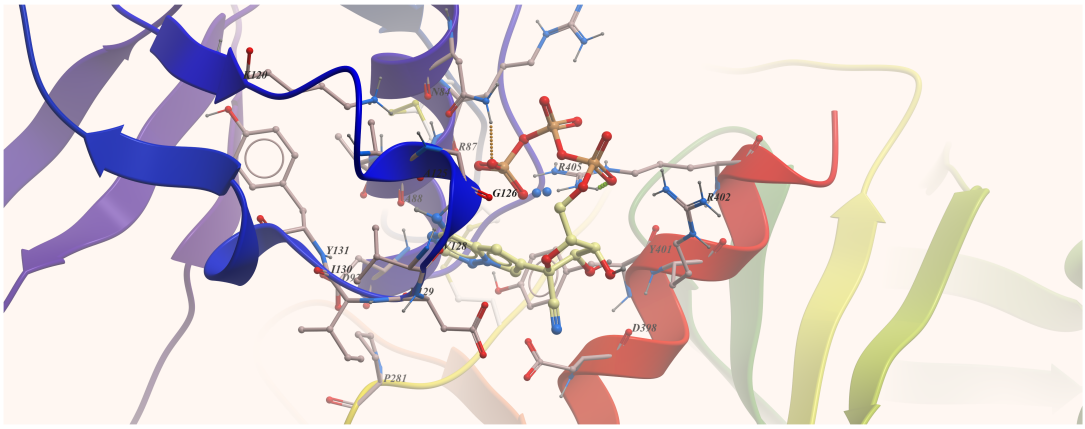



A
B
C
T565
G683
K500
S682
N497
R569
D684
N496
T686
V687
V495
A685
I494
A688
K577
Q573
V689
I589
L576
A580

A


G126
R405
R402

B


L120
R87
R405
G126
R402
Y131
I130
L29
Y401
D398
P281