

DrugDevCovid19: An Atlas of Anti-COVID-19 Compounds Derived By Computer-Aided Drug Design

Yang Liu

Sichuan University - Wangjiang Campus: Sichuan University <https://orcid.org/0000-0002-2517-9436>

Jianhong Gan

Sichuan University - Wangjiang Campus: Sichuan University <https://orcid.org/0000-0001-7519-6851>

Rongqi Wang

Sichuan University - Wangjiang Campus: Sichuan University <https://orcid.org/0000-0002-1824-3751>

Xiaocong Yang

Sichuan University - Wangjiang Campus: Sichuan University <https://orcid.org/0000-0002-9903-9380>

Zhixiong Xiao

Sichuan University - Wangjiang Campus: Sichuan University <https://orcid.org/0000-0003-2504-5742>

Yang Cao (✉ cao@scu.edu.cn)

Sichuan University - Wangjiang Campus: Sichuan University <https://orcid.org/0000-0002-1925-2123>

Research article

Keywords: COVID-19, SARS-CoV-2, Computer-Aided Drug Design, Bioinformatics, Database

Posted Date: November 9th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1045472/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Since the outbreak of SARS-CoV-2, numerous compounds against COVID-19 have been derived by computer-aided drug design (CADD) studies. They are valuable resources for the development of COVID-19 therapeutics. In this work, we reviewed these studies and analyzed 779 compounds against 16 target proteins from 181 CADD publications. We performed unified docking simulations and neck-to-neck comparison with the solved co-crystal structures. We computed their chemical features and classified these compounds aiming to provide insights for subsequent drug design. Through detailed analyses, we recommended a batch of compounds that are worth further study. Moreover, we organized all the abundant data and constructed a freely available database, DrugDevCovid19 (<http://clab.labshare.cn/covid/php/index.php>), to facilitate the development of COVID-19 therapeutics.

1 Introduction

On 31 December 2019, a previously unknown coronavirus was reported in Wuhan (China) to the World Health Organization. It is now known as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which can manifest itself in the disease named COVID-19. At the time of writing, nearly 239 million cases were reported worldwide, with over 4 million deaths. SARS-CoV-2 has a genome of approximately 30,000 nucleotides, which can be translated into approximately 30 proteins[1, 2]. The encoded spike (S) protein employs the host's angiotensin-converting enzyme 2 (ACE2) and extracellular protease transmembrane protease serine 2 (TMPRSS2) for S attachment and priming, respectively[3–5]. After that, it starts the fusion of the virus envelope and cell membrane. Researchers have now revealed that the virus can enter host cells through endocytosis, and use the pH-sensitive endosomal cysteine proteases cathepsin B and L (CatB/L) for S protein priming in TMPRSS2- cell lines[4, 6]. Recently, it was found that another entry cofactors called neuropilin-1 (NRP1) on the host cell surface can bind to the S protein to potentiates infectivity[7, 8]. After the virus invades the cell and uncoating, the viral genome RNA is translated to produce several viral non-structural proteins, including RNA-dependent RNA polymerase (RdRp), the papain-like protease (PLpro), and 3-chymotrypsin-like main protease (Mpro). The two proteases are assembled into viral replicase–transcriptase complex[9, 10].

These intensive studies reveal many attractive targets for anti-SARS-CoV-2 drug discovery and repurposing, including the viral non-structural proteins (nsp1-16), viral structural proteins (e.g., S and nucleoprotein), and host receptors and enzymes related to virus infection or reproduction in host cells, such as ACE2, TMPRSS2, DHODH and PAR-1[11, 12]. According to the solved structures in the PDB[13], 82% (18/22) of the potential targets currently have complete or partial structural data, and the mechanism of inhibiting targets has been achieved intuitively and profoundly by the co-crystallization of targets and their potential inhibitors. It provides a strong foundation for the development of anti-SARS-CoV-2 drugs, taking advantages of CADD and high-throughput screening.

So far, a large number of drug candidates have been developed for COVID-19 treatment, some of which are tested clinically and have shown preliminary therapeutic effects[14–21]. We believe that it is the right

time to summarize the discovery of these anti-COVID-19 drug candidates for a better understanding of the strong and weak points of current CADD studies. In this work, we collected 779 drug candidates discovered by CADD methods from 181 publications and conducted comprehensive comparisons on their targets, binding conformation, druggable properties, etc. We also built a publicly available database to illustrate those discoveries, which may profit the drug development against COVID-19.

2 Methods

2.1 Compound collection and preprocessing

The compounds in this work were collected from 181 peer-reviewed articles since the outbreak of SARS-CoV-2. The articles were obtained by searching Google Scholar and PubMed using the keywords of “SARS-CoV-2” or “COVID-19” as well as “CADD”, “Virtual Screening”, “Drug Design”, “Repurposing”, “Screening”, “Drug Discovering”, “Inhibitor”, etc. The majority of the 2D structures were downloaded from PubChem[22] or ZINC[23] databases. The compounds that were not included in either database were manually constructed according to the structure diagram provided in the literature. Their 3D structures were generated using the Experimental-Torsion Basic Knowledge Distance Geometry (ETKDG) method embedded in RDKit 2021.03.01[24]. To conduct the molecular docking calculation based on AutoDock Vina (version 1.1.2)[25] and CB-Dock (version 1.0)[26], the AutoDockTools[27] software was utilized to convert these compounds into the PDBQT format.

2.2 Protein preparation for molecular docking

Among the 16 target proteins from 181 CADD publications, 15 were available in the PDB database [28]. Since most targets have multiple entries, they were selected by the following criteria: 1) structures with full-length or at least complete active sites; 2) structures with co-crystallized ligand; 3) structures used in the reported virtual screening researches. Accordingly, the selected crystal structures include 6W6Y (Mac1), 6W9C (PL-PRO S1/S2 pocket), 7CMD (PL-PRO S3/S4 pocket), 6LU7 (Mpro), 6W4B (nsp9), 7BV2 (RdRp), 6XEZ (nsp13), 6W01 (nsp15), 6W4H (nsp16), 6M0J (S), 6M3M (Nucleoprotein), 1R4L (ACE2 catalytic pocket), 6M0J (ACE2 PPI surface), 7MEQ (TMPRSS2), 7K2U (DHODH), 3VW7 (PAR-1). Particularly, due to the lack of active sites in the experimental structures of nsp1 and nsp14, we obtained their structure models from (<https://zhanglab.ccmh.med.umich.edu/COVID-19/>)[29] for active-site analysis and docking. For each target, water molecules and ligands were deleted, while hydrogens were added with appropriate protonation states using AmberTools[30]. Proteins with missing loops and side chains were fixed using DeepView v4.1[31]. The binding sites of each compound were obtained by referring to the corresponding publications.

2.3 Docking protocol

To investigate the binding capability of the reported compounds, AutoDock Vina 1.1.2 was employed for the docking simulation. Moreover, template-based docking simulations were also performed to illustrate the similarities and differences between them. The latter was carried out using a template-based docking

tool (FitDock, in-house developed software, unpublished). For each compound, the most similar active inhibitor and its corresponding co-crystallized target structure were used as a template to perform the FitDock simulation.

2.4 Structure analysis of compounds

The chemical spatial features of CADD recommended compounds and known inhibitors, including cLogP (calculated lipophilicity), HBA (hydrogen-bond acceptor), HBD (hydrogen-bond donor), MW (molecular weight), and NRotB (Number of rotatable bonds), were evaluated by XLOGP3[32]. We evaluated the chemical space coverage by computing t-Distributed Stochastic Neighbor Embedding (tSNE) maps of chemical spatial features of the CADD compounds and known inhibitors[33]. Exploiting this method, compounds with similar chemical features can be clustered together. The structural similarity between compounds was evaluated with the Tanimoto similarity coefficient of Morgan fingerprint generated by RDKit, the similarity score of LigMate[34] and the PC-score of LS-align[35].

2.5 The web server of DrugDevCovid19

The DrugDevCovid19 web server was implemented in HTML5, CSS3, PHP, and JavaScript. The chart.js was used to visualize the statistical data and the ngl.js[36] was used to visualize the structural data which were collected in this work. The server offered chemical structures and identifiers (PubChem CID or ZINC ID or Custom ID) of compounds, structures, and PDB codes of the corresponding targets, their 3D binding conformations, sources et al. If a compound was reported multiple times, they were merged into one item. All those data were well-organized for comparison studies.

3 Results And Discussion

3.1 Molecular docking is the most popular CADD method against COVID-19

CADD has seen broad application in drug discovery against COVID-19. In our collected publications, most of them employed docking screen[37–43], molecular simulation[44], pharmacophore models[45, 46], or machine learning-based virtual screens[47] for drug discovery. Docking screen was the most popular one, which took over 96% of the works. They took the advantage of well-established docking tools including the academic software AutoDock[27] or AutoDock Vina[25], DOCK6[48], LeDock[49], rDock[50], LigandFit[51], etc., or the commercial software Glide[52], GOLD[53], Surflex-Dock[54], MOE[55], Discovery Studio[56], etc. A typical docking screening conducted by those studies includes three steps: 1) molecular docking, to predict binding modes and affinities of each molecule in a compound library; 2) analyzing ADMET to predict the pharmacokinetic properties; 3) molecular dynamics, to exam the binding stability and free energy. Besides the docking methods, compound libraries were also a key factor for virtual screening. They mainly included FDA-approved, investigational and experimental drug libraries, natural product libraries, and comprehensive compound libraries, e.g. ZINC[23], ChEMBL[57], PubChem[22].

Particularly, almost one-third of our collections employed the FDA-approved, investigational, or experimental drug libraries, indicating the importance of drug repurposing in dealing with this pandemic.

A successful CADD study not only depends on computational methods but also relies on the target proteins. In our collected publications, we observed 16 drug targets used for drug design (Table 1 and **Table S1**). They included the 12 proteins from SARS-CoV-2 but also 4 proteins from human, a host of the virus. The former included the viral non-structural proteins, such as nsp1~16 involved in the viral pathogenicity, replication, modification of viral RNA, or assembly of virions, and structural proteins, such as spike protein (S) related to the viral infection, or nucleoprotein (N) related to viral assembly. The latter included 4 enzymes including angiotensin-converting enzyme 2 (ACE2), transmembrane protease serine 2 (TMPRSS2), dihydroorotate dehydrogenase (DHODH) and proteinase-activated receptor 1 (PARP-1). In all those 16 proteins, 15 had at least one experimentally solved structure, except nsp14. Particularly, as of August 4, 2021, the number of experimental structures for nsp3 (PLpro), nsp5 (Mpro), and S are as many as 308, 369, and 466, respectively in the PDB[28], which are remarkably more than those for the other targets (Table 1 and **Table S1**). Comparative analysis showed that some of the proteins showed significant conformational changes upon ligand binding. For example, the PLpro-inhibitor complex and PLpro apo structure were remarkably different at BL2 loop (**Figure S1a**). And open and closed conformations were observed in the exterior residues (e.g. Q24, D30, E35, E37, D38, Y41, Q42, N53, E56, Y83, Q325, E329, Q330, K353, R393) of ACE2 (**Figure S1b**). Those structural variations have been carefully investigated in some of the CADD studies[58, 59].

Table 1
The main drug targets for the treatment of SARS-CoV-2 used in CADD studies.

Targets	Short name	Organism	Solved structures (Solved complexes)
Host translation inhibitor nsp1	nsp1	SARS-CoV-2	21 (0)
Non-structural protein 3	nsp3, PLpro	SARS-CoV-2	308 (262)
3C-like proteinase	nsp5, 3CLpro, Mpro	SARS-CoV-2	369 (177)
Non-structural protein 9	nsp9	SARS-CoV-2	12 (0)
RNA-directed RNA polymerase	nsp12, RdRp	SARS-CoV-2	28 (9)
Helicase	nsp13, Hel	SARS-CoV-2	67 (58)
Proofreading exoribonuclease	nsp14, ExoN	SARS-CoV-2	11 (0)
Uridylate-specific endoribonuclease	nsp15	SARS-CoV-2	44 (26)
2'-O-methyltransferase	nsp16	SARS-CoV-2	28 (24)
Spike glycoprotein	S	SARS-CoV-2	466 (13)
Nucleoprotein	N	SARS-CoV-2	21 (0)
Angiotensin-converting enzyme 2	ACE2	<i>Homo sapiens</i>	58 (1)
Transmembrane protease serine 2	TMPRSS2	<i>Homo sapiens</i>	1 (1)
Dihydroorotate dehydrogenase	DHODH	<i>Homo sapiens</i>	79 (75)
Proteinase-activated receptor 1	PAR-1	<i>Homo sapiens</i>	5 (1)

Note: Solved structures refer to the number of solved 3D structures in the PDB; Solved complexes refer to the number of solved 3D structures complexed with ligands (binding in the active sites) in the PDB.

3.2 CADD studies predicted diverse compounds against COVID-19

Since the outbreak of SARS-CoV-2, CADD has been widely used either by drug repurposing or screening novel compounds. In the 181 research reports collected, we obtained a total of 1073 compound records, which included molecular sources, research methods, targets, active sites for binding, etc. Among these compounds, 53.6%, 11.4%, 7.5%, 7.1%, 6.6% target Mpro, S, RdRp, PLpro and ACE2 respectively (Figure 1a). Some of the compounds have been recommended by multiple research teams. For example, lopinavir, one of the compounds recommended by nine virtual screening researches, has been proved to inhibit SARS-CoV-2 with the IC₅₀ of 9.12 μ M in the later wet-lab research[60]. Thus, we summarized the compounds recommended more than twice in targeting a receptor, as shown in Figure 1b, suggesting their potential in the following research.

After eliminating the redundant records, we finally obtained 779 recommended compounds, including 457 (59%), 114 (14.6%), 74 (9.5%), 70 (9%) and 69 (8.9%) compounds targeting Mpro, S, PLpro, RdRp, ACE2 and the others respectively (Table 2). We also observed that a target protein may contain two or more binding pockets for drug design. These targets include nsp1, PLpro, nsp9, nsp13, nsp14, nsp16, S protein, N protein and ACE2 (**Table S2**). The different pockets usually play different roles in a target protein. For example, in the papain-like protease domain of PLpro, the S1/S2 pocket is the binding regions of catalytic sites while S3/S4 pocket (the BL2 loop) is substrate-binding related sites[58]. And in nsp13, both the ATP binding pocket and RNA binding pocket are used for screening[61]. In this work, we did not distinguish the detailed impact of the binding pockets, but performed the analysis by comparing the recommended compounds using the same binding pockets but not the same target proteins.

In order to compare compounds obtained from different studies, we performed unified molecular docking against the top five popular drug targets, including 457 recommended compounds against the catalytic site of Mpro, 73 recommended compounds against the PPI surface of S, 68 recommended compounds against the catalytic site of RdRp, 68 recommended compounds against S3/S4 pocket of PLpro and 41 recommended compounds against the catalytic site of ACE2 (Figure 2a). The unified docking was compared to the nine experimentally validated compounds against Mpro, with the corresponding docking scores of \sim -7.0 kcal/mol (**Table S3**). We observed that 76%, 75%, 47%, and 59% recommended compounds in Mpro, S3/S4 pocket of PLpro, RdRp, and S respectively achieved lower docking scores (lower than -7.0 kcal/mol), which suggest stronger binding. Particularly, GRL0617, the ranking-first compounds targeting PLpro S3/S4 pocket (docking score = -10.0 kcal/mol), was confirmed to have strong affinity (K_D = 1.93 μ M)[58], after predicting in July 2020[37]. Unlike the four targets, only 2% recommended compounds against ACE2 achieved scores were lower than -7.0 kcal/mol. It could be attributed to the flat interface between ACE2 and compounds, which originally binds with the viral S protein. Hence it is much more challenging to develop drugs against ACE2 than the others.

We also evaluated the drug-likeness of the recommended compounds using Lipinski's "Rule of Five"[62]. The results showed that 67% and 81% of compounds targeting Mpro, PLpro respectively met at least four of the Lipinski's rules (Figure 2b). However, only 53%, 56%, 57% of compounds targeting S, ACE2, and RdRp respectively were able to fulfill the criteria, implying the difficulties to discover qualified drugs against those targets.

Table 2
Statistics of studies on the SARS-CoV-2 inhibitor discovery using CADD.

Target	#Refs	#Hits	Target	#Refs	Hits
nsp1	2	14	nsp16	2	25
PLpro	12	74	S	34	114
Mpro	124	457	N	5	18
nsp9	2	9	ACE2	16	69
RdRp	23	70	TMPRSS2	6	29
Helicase	2	5	DHODH	1	27
ExoN	2	2	PAR-1	1	2
NendoU	2	12			

3.3 Comparison studies between CADD compounds and co-crystallized inhibitors

So far, a large number of co-crystal structures of SARS-CoV-2 targets in complex with inhibitors are available in PDB, for example, 136, 74, 37 and 16 compounds are co-crystallized with Mpro, DHODH, ATP binding domain of nsp13, and papain-like protease domain of PLpro respectively (**Tables S4**). It is a valuable source for evaluating and improving the CADD recommended compounds by comparing with those experimentally determined inhibitors. In the following sections, we will elaborate the results of our comparison studies on molecular similarity, drug-likeness, and binding modes for those compounds.

3.3.1 Molecular similarity

We first analyzed molecular similarity between the compounds discovered by CADD and the co-crystallized inhibitors in terms of 2D (morgan fingerprint) and 3D (LS-align) similarities, which were evaluated by Tanimoto coefficients and PC-score respectively. The result showed that in the available inhibitors against 11 SARS-CoV-2 targets, the average Tanimoto coefficients were all less than 0.3, indicating that the majority of CADD compounds were dissimilar to the co-crystallized inhibitors (Figure 3a). In other words, CADD compounds showed diverse chemical scaffolds compared with the co-crystallized inhibitors. Nevertheless, we also found that 19, a small number of CADD compounds were highly similar to the Mpro inhibitors (Tanimoto similarity coefficient ≥ 0.5) (**Table S5**). Among them, 14 flavonoid analogues shared the same skeleton as the known inhibitors Myricetin[63], 7-O-methylmyricetin[63], 7-O-methyl-dihydromyricetin[63] and Baicalein[64]. In addition, Peramppanel is the lead compound of the known inhibitor COMPOUND4[65], which also has the same skeleton structure. Flovagatran, DB04692, Compound 23727975 and Caspase-1 Inhibitor VI are peptoid compounds, which have the most similar topological structures with the known inhibitor MPI1[66], MPI6[66], N3[67] and Z-

VAD(OMe)-FMK[68] respectively. Compared to the Tanimoto coefficients, the 3D similarity analysis showed higher similarities that the average PC-scores are all larger than 0.5. And they were even over 0.7 for the compounds targeting Mpro, nsp13 and DHODH, which indicated that the compounds shared rather similar topology in space despite of the different chemical composition (Figure 3b). What is more, 1.5%, 42.4%, 5.9%, 80%, 2.4%, 6.9% and 11.1% CADD compounds targeting PLpro (S3/S4 pocket), Mpro, RdRp, nsp13 (ATP binding site), ACE2 (catalytic pocket), TMPRSS2 and DHODH respectively, showed that maximal PC-score were greater than 0.8. The above results illustrated that the CADD methods not only potentially discovered the variants of co-crystallized inhibitors but also created diversities for further drug discovery.

3.3.2 Drug-likeness

We further performed the comparative study on the drug-likeness of the CADD compounds with the known co-crystallized inhibitors against the corresponding targets in terms of cLogP (calculated lipophilicity), HBA (hydrogen-bond acceptor), HBD (hydrogen-bond donor), MW (molecular weight). In this section, we investigated the compounds targeting Mpro (136 inhibitors, 457 CADD compounds), PLpro S3/S4 pocket (13 inhibitors, 68 CADD compounds) and DHODH (74 inhibitors, 27 CADD compounds) because they have multiple co-crystallized structures, making the comparative analysis more compelling. According to Lipinski's "Rule of Five", the compounds with $MW \leq 500$, $HBA \leq 10$, $HBD \leq 5$ and $-2 \leq cLogP \leq 5$ are predictive of good permeability and absorption[62]. Because the latest evaluation of 204 small molecule oral drugs by Brown et al.[69] showed 30% approved drugs with an MW higher than 500 Da. Therefore, we raised the MW standard in "Rule of Five" to 600 Da in the following analysis. The chemical space comparison between SARS-Cov-2 Mpro inhibitors and CADD compounds was illustrated in a tSNE map (Figure 4). 54% of the CADD compounds met the "Rule of Five". Moreover, the distribution of drug-likeness features was basically the same as that of Mpro inhibitors. The clusters of inhibitors showed that the peptoid inhibitors represented by PF-00835231[70], MPI7[66], and Narlaprevir[71] accounted for the majority (52%) of the currently known Mpro inhibitors. These inhibitors typically had greater MW and binding affinities, which was closely related to the larger binding pocket of the active site in Mpro[65, 70, 72]. More significantly, a cluster of CADD compounds presented in the region at which peptoid inhibitors was distributed, implying their similar drug-likeness properties. The remaining 46% of CADD compounds violated one or more rules in the "Rule of Five". Some of them showed higher hydrophilicity, and the others had larger molecular weights ($MW > 600$) or higher lipophilicity ($cLogP > 5$). For PLpro and DHODH (Figure S2), 78% of CADD compounds targeting PLpro S3/S4 pocket met the "Rule of Five" and had a more diverse chemical space than the known inhibitors. All the CADD compounds of DHODH met the "Rule of Five" and fell within the chemical space of known inhibitors.

3.3.3 Binding modes

To reveal the molecular mechanism, we next compared the binding modes of the CADD compounds with the co-crystallized inhibitors. Our analysis was focused on the 19 CADD recommended compounds of Mpro described above, which shared highly similar topological structures with co-crystallized inhibitors. We compared the docking conformation of these CADD compounds to the corresponding co-crystallized

structures (**Table S5**) in terms of the chemical compositions for favorable binding. Among the 19 recommended compounds, Perampanel, an anti-epileptic drug, possesses a distinctive topology. And compared with the co-crystallized inhibitor (named COMPOUND4 in Zhang's report[65]), Perampanel lacked the polar interaction with His163 and Glu166 and hydrophobic interaction with the S2 pocket in its predicted binding mode (Figure 5a). Moreover, the binding assays also showed that the activity of Perampanel was much lower than that of COMPOUND4 (e.g., IC₅₀ of 100~250 μ M vs. 4.02 μ M[65]). 14 flavonoid analogues shared the similar binding mode and key interactions with the co-crystallized inhibitors. For example, the parent nucleus of Baicalin docked into the Mpro pocket in a similar manner to a co-crystallized inhibitor Baicalein, which possessed approximately 7-fold higher affinity than Baicalin (e.g., IC₅₀ of 0.94 μ M vs. 6.41 μ M[64]). Since the 7-O-glucuronide was unable to fit in the S1' pocket appropriately in this binding mode, thus it resulted in the weaker polar interactions with Leu141, Gly143 and Glu166. However, the interaction of 7-O-glucuronide with Thr24-26 on the outside of the S1' pocket contributed to the binding of Baicalin (Figure 5b). As another example, Quercetin and Robinetin shared the similar binding mode with the covalent inhibitor Myricetin (IC₅₀ of 0.63 μ M[63]), and both could form polar interactions with GLU166 (Figure 5c and Figure 5d). Additionally, since the pyrogallol group in Myricetin worked as an electrophile to covalently modify the catalytic cysteine of Mpro[63], it can be speculated that the pyrocatechol group in Quercetin and the pyrogallol group in Robinetin may also have the ability to covalently bind to catalytic cysteine. In-depth study of these flavonoids will further promote the design of non-peptidomimetic covalent or non-covalent inhibitors against Mpro. The remaining four compounds are peptoids. Among them, DB04692 and Compound 23727975 shared the similar non-covalent docking modes, with identical or similar groups binding in the S1 and S2 pocket as that of co-crystallized inhibitors (Figure 5e and Figure 5f). Moreover, both had a reactive warhead in the S1' pocket, close to Cys145, implying the ability to form covalent bond.

3.4 An atlas of the inhibitors against SARS-CoV-2

To facilitate the analysis of the discovered drug candidates against SARS-CoV-2, we constructed an online database, named DrugDevCovid19, which collected the compound structures, target protein structures, chemical properties, screening methods, experimental data, and 3D docking modes from the primary published computational and experimental reports. The current version contains 779 compounds and 16 drug targets extracted from 181 reports mentioned above. It is publicly available via a user-friendly interface at <http://clab.labshare.cn/covid/php/index.php> (Figure 6). Compounds and targets of interest can be searched in a variety of ways at the above webpage. The 'Entries List' presents all the compound records retrieved from literature. The records were classified by the drug target and displayed in a donut chart. Users can click the regions in the chart to view the detailed records. The 'Candidates List' presents all the non-redundant compounds contained in the records. Similarly, users can click each region to view the compounds which are sorted by record number. Since there may be multiple druggable sites on one target, the compounds were further classified according to the active site and sorted by the binding affinity values predicted by AutoDock Vina[27]. The 'Targets List' presents all the potential drug targets. For each target, the web interface illustrates the full name, PDB code(s), the potential active sites retrieved from literature, the organisms, and the Uniprot ID. Users can click the target name to view more

information, such as the target function retrieved from UniProt and the surface residues in the active site. The 3D structural features of the active site can be viewed via an NGL Viewer. For each CADD compound, we showed its chemical spatial features, drug function (retrieved from DrugBank if it exists), structural information (including SMILES, 2D, and 3D structures), and Vina scores, with an interactive 3D visualization of the docking conformation.

4 Conclusions

To develop therapeutic drugs that specifically target SARS-CoV-2 as soon as possible, scientists have comprehensively studied the SARS-CoV-2 from various aspects, including the viral genome, proteome, the pathway of viral infection, and patient's immune response to the virus. Among them, the proteins encoded by the viral genome and the host proteins that are involved in viral invasion are the most promising anti-SARS-CoV-2 targets. CADD can be used to quickly screen potential compounds to target these proteins. In this work, we collected 779 CADD recommended compounds against 16 targets reported in 181 CADD studies. We analyzed the structures in which the inhibitors are co-crystallized with their target proteins, explored their binding modes and conformational diversities, aiming at gaining structural insights for CADD and identifying features and rules for further drug design and refinement. Our efforts can be briefly summarized into three aspects. First, we analyzed the structures of the targets in complex with active inhibitors. We preliminarily analyzed the structural diversity, binding modes, and the structure-activity relationship of the target proteins in complex with the active inhibitors. Second, we performed docking analyses of the CADD compounds. We identified the most similar active inhibitors of each CADD compound and performed molecular docking to compare their binding modes and affinities with those of the known inhibitors to select the most promising drug candidates. Finally, we created the freely available database to visualize the data generated in the above two aspects.

Among the 16 protein targets, the viral Mpro and PLpro are the most widely studied. Since the two proteases of SARS-CoV-2 are highly similar to those of SARS-CoV (96% and 83% of sequence identity for Mpro and PLpro, respectively)[73], the peptoid compounds designed in previous studies which can covalently bind to SARS-CoV Mpro as well as other proteases, were first attempted to treat SARS-CoV-2 infection[70, 74, 75]. Most of the Mpro inhibitors currently under investigation are peptide-like compounds with a reactive head that can covalently bind Cys145. VIR250 and VIR251, two PLpro inhibitors developed by Rut et al., are also peptoid molecules that can covalently bind to the catalytic residue Cys111[76]. However, such compounds may be ineffective due to the degradation caused by off-target covalent modification. Therefore, some non-covalent and non-peptoid protease inhibitors were also proposed. For example, Zhang et al. designed a series of highly active non-covalent, non-peptoid Mpro inhibitors based on the structure of Perampanel[65]. Su et al. confirmed that Baicalin and its analogue Baicalein can non-covalently bind to the active site of Mpro, exhibiting potent antiviral activities in a cell-based assay[64]. Ratia et al. discovered GRL0617, a non-covalent inhibitor of SARS-CoV PLpro, in 2008 and proved to be effective in inhibiting SARS-CoV-2 infection[77, 78]. Subsequently, Shen et al. designed a series of new PLpro inhibitors based on the scaffold of GRL0617[58]. These non-covalent drugs may become new therapies for the treatment of SARS-CoV-2 infection. By tracing the sources of these active inhibitors, it is

found that some of them are from the early virtual screening studies[65, 79]. However, besides these known inhibitors, there are much more recommended compounds obtained from CADD studies for the further study. Our analysis showed that these CADD compounds exhibit high chemical spatial diversity, which is particularly valuable to guide the development of new drugs.

S, ACE2, and RdRp are also potential targets and were investigated by some CADD studies. Since the interface between S and ACE2 lacks typical pockets or cavities, which are generally required for small-molecule inhibitors, the docking studies on these two targets involve the selection of binding pockets. For S, all the existing reports aimed at targeting the receptor-binding domain (RBD), but focused on different binding sites. For example, Choudhary et al.[80], Sinha et al.[81], Alexpandi et al.[82] and Kalhor et al.[83] directly select the contact surface that binding to ACE2 as the docking site and Mahdian et al.[38], Feng et al.[84] and Wei et al.[85] select pockets near the contact surface for docking. For ACE2, the docking studies targeted the contact surface that binding to S RBD[38, 80, 86, 87] or the catalytic pocket to block the enzyme activity and stabilize the closed conformation of ACE2, thereby to shift the relative positions of the receptor's critical exterior residues recognized by SARS-CoV-2[59, 88, 89]. However, no experimental evidence has shown the effectiveness of these strategies. To date, the promising drug leads targeting S or ACE2 to block SARS-CoV-2 infection are peptide binder and neutralizing antibodies (such as soluble human ACE2)[90–94]. For RdRp, the reported effective inhibitors are mainly nucleotide analog viral polymerase inhibitors, such as Remdesivir and Favipiravir, but the clinical outcomes are unsatisfactory. For example, several randomized clinical trials that investigated the efficacy of Remdesivir for COVID-19 treatment produced inconsistent results[95–98]. Favipiravir, compared with Arbidol, did not significantly improve the clinical recovery rate at Day 7, but it did significantly improve the latency to cough relief and decreased the duration of pyrexia[99]. And the meta-analysis conducted by Shrestha et al. concluded that patients had clinical and radiological improvements following the treatment with favipiravir in comparison to that of the standard care though no significant differences on viral clearance, oxygen support requirement and side effect profile[100]. One possible reason is that the exoribonuclease of SARS-CoV-2 provides the proofreading capacity to RdRp thus exerts relatively high resistance of RdRp on nucleotide analog inhibitors[101, 102]. These nucleotide analogs usually bind to the RdRp active site (NTP binding site)[103], which is also the main site for the development of RdRp inhibitors using CADD methods[104]. Recently, the cryo-electron microscopy structure of the viral RdRp bound to suramin, a non-nucleotide inhibitor, reveals two new promising binding sites, which are the binding sites of the RNA template strand and primer strand[105]. The new structural mechanism is conducive to CADD studies.

In summary, starting from the structures of potential COVID-19 drug targets in complex with known inhibitors, our work first systematically explored the potential binding sites and binding modes to gain insights for drug discovery by CADD. We then collected and analyzed the recommended compounds from the current CADD studies based on the structures of target proteins, aiming to identify novel inhibitors from them. All data can be retrieved from our database, DrugDevCovid19, with a user-friendly interface at <http://clab.labshare.cn/covid/php/index.php>.

Declarations

Availability of data and materials

The data described in this study are freely available at <http://clab.labshare.cn/covid/php/index.php>. FitDock, the template-based docking tool used in this study, is available upon contacting the corresponding author.

Competing interests

The authors declare that they have no conflict of interest.

Funding

This work was supported by the National Natural Science Foundation of China under Grant (number 81973243).

Author Contributions

Y.L. performed the computation; Y.L., J.G., R.W. and X.Y. analyzed data; Y.L. and Y.C. wrote the paper; Y.C. and Z.X. designed the research. All authors have given approval to the final version of the manuscript.

Acknowledgements

The authors wish to thank Prof. Shirley Liu in Dana-Farber Cancer Institute and Prof. Zairong Zhang in Chinese Academy of Sciences for invaluable discussion.

Authors' information

Y. L. is currently a PhD candidate at college of life sciences, Sichuan University, China. He mainly focuses on developing methods for sequence analysis and computer-aided drug design. He is the major developer of cavity detection-guided protein–ligand blind docking method CB-Dock.

J. G. is currently a Master student at college of life sciences, Sichuan University, China. His major research interest is focused on the method development for protein structure prediction.

R. W. is currently a Master student at college of life sciences, Sichuan University, China. Her major research interest is the development of peptide-MHC binding prediction methods.

X. Y. is currently a Master student at college of life sciences, Sichuan University, China. He mainly focuses on developing template-based molecular docking algorithms. He is the major developer of a template-based docking tool FitDock.

Z. X. received his PhD from University of Massachusetts. He is currently a full professor at college of life sciences, Sichuan University, director of the Center of Growth, Metabolism and Aging (CGMA). His group

mainly investigates cancer related cell biology, bio-markers and drug screening.

Y. C. received his PhD from Institute of Biophysics, Chinese Academy of Sciences. He is an associate professor at college of life sciences, Sichuan University and a visiting scholar in Harvard University. His research interests include developing new methods for computer-aided drug design and protein design. He is the major developer of a group of computational methods such as CISRR, CYSCORE, AbRSA, CB-DOCK and LigMate.

References

1. Rouchka EC, Chariker JH, Chung D (2020) Variant analysis of 1,040 SARS-CoV-2 genomes. *PLoS ONE* 15:e0241535. <https://doi.org/10.1371/journal.pone.0241535>
2. Cantini F, Banci L, Altincekic N et al (2020) ¹H, ¹³C, and ¹⁵N backbone chemical shift assignments of the apo and the ADP-ribose bound forms of the macrodomain of SARS-CoV-2 non-structural protein 3b. *Biomol NMR Assign* 14:339–346. <https://doi.org/10.1007/s12104-020-09973-4>
3. Zhou P, Yang X-L, Wang X-G et al (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579:270–273. <https://doi.org/10.1038/s41586-020-2012-7>
4. Mantlo E, Bukreyeva N, Maruyama J et al (2020) Antiviral activities of type I interferons to SARS-CoV-2 infection. *Antiviral Res* 179:104811. <https://doi.org/10.1016/j.antiviral.2020.104811>
5. Huang X, Pearce R, Omenn GS, Zhang Y (2021) Identification of 13 Guanidinobenzoyl- or Aminidinobenzoyl-Containing Drugs to Potentially Inhibit TMPRSS2 for COVID-19 Treatment. *Int J Mol Sci* 22:7060. <https://doi.org/10.3390/ijms22137060>
6. Simmons G, Gosalia DN, Rennekamp AJ et al (2005) Inhibitors of cathepsin L prevent severe acute respiratory syndrome coronavirus entry. *Proc Natl Acad Sci U S A* 102:11876–11881. <https://doi.org/10.1073/pnas.0505577102>
7. Daly JL, Simonetti B, Klein K et al (2020) Neuropilin-1 is a host factor for SARS-CoV-2 infection. *Science* 370:861–865. <https://doi.org/10.1126/science.abd3072>
8. Cantuti-Castelvetri L, Ojha R, Pedro LD et al (2020) Neuropilin-1 facilitates SARS-CoV-2 cell entry and infectivity. *Science* 370:856–860. <https://doi.org/10.1126/science.abd2985>
9. De Wit E, Van Doremalen N, Falzarano D, Munster VJ (2016) SARS and MERS: Recent insights into emerging coronaviruses. *Nat Rev Microbiol* 14:523–534
10. Klemm T, Ebert G, Calleja DJ et al (2020) Mechanism and inhibition of the papain-like protease, PLpro, of SARS-CoV-2. *EMBO J* 39:e106275. <https://doi.org/10.15252/embj.2020106275>
11. Shyr ZA, Gorshkov K, Chen CZ, Zheng W (2020) Drug discovery strategies for sars-cov-2. *J Pharmacol Exp Ther* 375:127–138
12. Zhou H, Fang Y, Xu T et al (2020) Potential therapeutic targets and promising drugs for combating SARS-CoV-2. *Br J Pharmacol* 177:3147–3161. <https://doi.org/10.1111/bph.15092>

13. Burley SK, Bhikadiya C, Bi C et al (2021) RCSB Protein Data Bank: Powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. *Nucleic Acids Res* 49:D437–D451. <https://doi.org/10.1093/nar/gkaa1038>
14. NCT04380688 (2020) Acalabrutinib Study With Best Supportive Care Versus Best Supportive Care in Subjects Hospitalized With COVID-19. <https://clinicaltrials.gov/show/NCT04380688>
15. NCT04535167 (2020) First-In-Human Study To Evaluate Safety, Tolerability, And Pharmacokinetics Following Single Ascending Dose And Extended Infusion of PF 07304814 In Hospitalized Participants With COVID-19. <https://clinicaltrials.gov/show/NCT04535167>
16. NCT04440007 (2020) Study of the Efficacy and Safety of STI-5656 (Abivertinib Maleate) With SOC Versus SOC in Subjects With COVID-19. <https://clinicaltrials.gov/show/NCT04440007>
17. NCT04382586 (2020) Covid-19 Infection and Pulmonary Distress Treatment With Zanubrutinib in Hospitalized Participants. <https://clinicaltrials.gov/show/NCT04382586>
18. NCT04308317 (2020) Tetrandrine Tablets Used in the Treatment of COVID-19. <https://clinicaltrials.gov/show/NCT04308317>
19. NCT04465695 (2020) Dual Therapy With Interferon Beta-1b and Clofazimine for COVID-19. <https://clinicaltrials.gov/show/NCT04465695>
20. NCT04489446 (2020) Sildenafil in COVID-19. <https://clinicaltrials.gov/show/NCT04489446>
21. NCT04261270 (2020) A Randomized,Open,Controlled Clinical Study to Evaluate the Efficacy of ASC09F and Ritonavir for 2019-nCoV Pneumonia. <https://doi.org/10.31525/ct1-nct04261270>. <https://clinicaltrials.gov/show/NCT04261270>
22. Kim S, Chen J, Cheng T et al (2021) PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Res* 49:D1388–D1395. <https://doi.org/10.1093/nar/gkaa971>
23. Sterling T, Irwin JJ (2015) ZINC 15 – Ligand Discovery for Everyone. *J Chem Inf Model* 55:2324–2337. <https://doi.org/10.1021/acs.jcim.5b00559>
24. Landrum G (2021) RDKit: Open-Source Cheminformatics Software. [Http://www.Rdkit.Org/](http://www.Rdkit.Org/)
25. Trott O, Olson AJ (2010) Software news and update AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* 31:455–461. <https://doi.org/10.1002/jcc.21334>
26. Liu Y, Grimm M, Dai W et al (2020) CB-Dock: a web server for cavity detection-guided protein–ligand blind docking. *Acta Pharmacol Sin* 41:138–144. <https://doi.org/10.1038/s41401-019-0228-6>
27. Huey R, Morris GM, Forli S (2012) Using AutoDock 4 and AutoDock Vina with AutoDockTools: A Tutorial. *Scripps Res Inst Mol32*
28. Berman HM, Westbrook J, Feng Z et al (2000) The Protein Data Bank. *Nucleic Acids Res* 28:235–242. <https://doi.org/10.1093/nar/28.1.235>
29. Zhang C, Zheng W, Huang X et al (2020) Protein Structure and Sequence Reanalysis of 2019-nCoV Genome Refutes Snakes as Its Intermediate Host and the Unique Similarity between Its Spike Protein

- Insertions and HIV-1. *J Proteome Res* 19:1351–1360.
<https://doi.org/10.1021/acs.jproteome.0c00129>
30. Salomon-Ferrer R, Case DA, Walker RC (2013) An overview of the Amber biomolecular simulation package. *Wiley Interdiscip Rev Comput Mol Sci* 3:198–210. <https://doi.org/10.1002/wcms.1121>
 31. Johansson MU, Zoete V, Michielin O, Guex N (2012) Defining and searching for structural motifs using DeepView/Swiss-PdbViewer. *BMC Bioinformatics* 13:173. <https://doi.org/10.1186/1471-2105-13-173>
 32. Cheng T, Zhao Y, Li X et al (2007) Computation of Octanol–Water Partition Coefficients by Guiding an Additive Model with Knowledge. *J Chem Inf Model* 47:2140–2148.
<https://doi.org/10.1021/ci700257y>
 33. Van Der Maaten L, Hinton G (2008) Visualizing data using t-SNE. *J Mach Learn Res* 9:2579–2625
 34. Grimm M, Liu Y, Yang X et al (2020) LigMate: A Multifeature Integration Algorithm for Ligand-Similarity-Based Virtual Screening. *J Chem Inf Model* 60:6044–6053.
<https://doi.org/10.1021/acs.jcim.9b01210>
 35. Hu J, Liu Z, Yu D-J, Zhang Y (2018) LS-align: an atom-level, flexible ligand structural alignment algorithm for high-throughput virtual screening. *Bioinformatics* 34:2209–2218.
<https://doi.org/10.1093/bioinformatics/bty081>
 36. Rose AS, Bradley AR, Valasatava Y et al (2018) NGL viewer: web-based molecular graphics for large complexes. *Bioinformatics* 34:3755–3758. <https://doi.org/10.1093/bioinformatics/bty419>
 37. Cavasotto CN, Di Filippo JI (2021) In silico Drug Repurposing for COVID-19: Targeting SARS-CoV-2 Proteins through Docking and Consensus Ranking. *Mol Inform* 40:e2000115.
<https://doi.org/10.1002/minf.202000115>
 38. Mahdian S, Ebrahim-Habibi A, Zarrabi M (2020) Drug repurposing using computational methods to identify therapeutic options for COVID-19. *J Diabetes Metab Disord* 19:691–699.
<https://doi.org/10.1007/s40200-020-00546-9>
 39. Berber B, Doluca O (2021) A comprehensive drug repurposing study for COVID19 treatment: novel putative dihydroorotate dehydrogenase inhibitors show association to serotonin–dopamine receptors. *Brief Bioinform* 22:1023–1037. <https://doi.org/10.1093/bib/bbaa379>
 40. Liu S, Zheng Q, Wang Z (2020) Potential covalent drugs targeting the main protease of the SARS-CoV-2 coronavirus. *Bioinformatics* 36:3295–3298. <https://doi.org/10.1093/bioinformatics/btaa224>
 41. Wang J (2020) Fast Identification of Possible Drug Treatment of Coronavirus Disease-19 (COVID-19) through Computational Drug Repurposing Study. *J Chem Inf Model* 60:3277–3286.
<https://doi.org/10.1021/acs.jcim.0c00179>
 42. Mahmud S, Uddin MAR, Paul GK et al (2021) Virtual screening and molecular dynamics simulation study of plant-derived compounds to identify potential inhibitors of main protease from SARS-CoV-2. *Brief Bioinform* 22:1402–1414. <https://doi.org/10.1093/bib/bbaa428>
 43. Manandhar A, Blass BE, Colussi DJ et al (2021) Targeting SARS-CoV-2 M3CLpro by HCV NS3/4a Inhibitors: In Silico Modeling and In Vitro Screening. *J Chem Inf Model* 61:1020–1032.

<https://doi.org/10.1021/acs.jcim.0c01457>

44. Genheden S, Ryde U (2015) The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin Drug Discov* 10:449–461. <https://doi.org/10.1517/17460441.2015.1032936>
45. Kanhed AM, Patel DV, Teli DM et al (2021) Identification of potential Mpro inhibitors for the treatment of COVID-19 by using systematic virtual screening approach. *Mol Divers* 25:383–401. <https://doi.org/10.1007/s11030-020-10130-1>
46. Kouznetsova VL, Zhang A, Tatineni M et al (2020) Potential COVID-19 papain-like protease PL pro inhibitors: repurposing FDA-approved drugs. *PeerJ* 8:e9965. <https://doi.org/10.7717/peerj.9965>
47. Mohapatra S, Nath P, Chatterjee M et al (2020) Repurposing therapeutics for COVID-19: Rapid prediction of commercially available drugs through machine learning and docking. *PLoS ONE* 15:e0241543. <https://doi.org/10.1371/journal.pone.0241543>
48. Allen WJ, Fochtman BC, Balius TE, Rizzo RC (2017) Customizable de novo design strategies for DOCK: Application to HIVgp41 and other therapeutic targets. *J Comput Chem* 38:2641–2663. <https://doi.org/10.1002/jcc.25052>
49. Zhao H, Caflisch A (2013) Discovery of ZAP70 inhibitors by high-throughput docking into a conformation of its kinase domain generated by molecular dynamics. *Bioorganic Med Chem Lett*. <https://doi.org/10.1016/j.bmcl.2013.08.009>
50. Ruiz-Carmona S, Alvarez-Garcia D, Foloppe N et al (2014) rDock: A Fast, Versatile and Open Source Program for Docking Ligands to Proteins and Nucleic Acids. *PLoS Comput Biol*. <https://doi.org/10.1371/journal.pcbi.1003571>
51. Venkatachalam CM, Jiang X, Oldfield T, Waldman M (2003) LigandFit: A novel method for the shape-directed rapid docking of ligands to protein active sites. *J Mol Graph Model*. [https://doi.org/10.1016/S1093-3263\(02\)00164-X](https://doi.org/10.1016/S1093-3263(02)00164-X)
52. Friesner RA, Banks JL, Murphy RB et al (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem* 47:1739–1749. <https://doi.org/10.1021/jm0306430>
53. Verdonk ML, Cole JC, Hartshorn MJ et al (2003) Improved protein-ligand docking using GOLD. *Proteins Struct Funct Genet* 52:609–623. <https://doi.org/10.1002/prot.10465>
54. Jain AN (2003) Surflex: Fully automatic flexible molecular docking using a molecular similarity-based search engine. *J Med Chem*. <https://doi.org/10.1021/jm020406h>
55. Chemical Computing Group ULC (2019) Molecular Operating Environment (MOE), 2019.01 User Guide. *Sci Comput Instrum*
56. Miyata T (2015) Discovery studio modeling environment. *Ensemble* 17:98–104
57. Gaulton A, Hersey A, Nowotka ML et al (2017) The ChEMBL database in 2017. *Nucleic Acids Res* 45:D945–D954. <https://doi.org/10.1093/nar/gkw1074>
58. Shen Z, Ratia K, Cooper L et al (2021) Potent, Novel SARS-CoV-2 PLpro Inhibitors Block Viral Replication in Monkey and Human Cell Cultures. <https://doi.org/10.1101/2021.02.13.431008>.

59. Teralı K, Baddal B, Gülcan HO (2020) Prioritizing potential ACE2 inhibitors in the COVID-19 pandemic: Insights from a molecular mechanics-assisted structure-based virtual screening experiment. *J Mol Graph Model* 100:107697. <https://doi.org/10.1016/j.jmglm.2020.107697>
60. Jeon S, Ko M, Lee J et al (2020) Identification of Antiviral Drug Candidates against SARS-CoV-2 from FDA-Approved Drugs. *Antimicrob Agents Chemother* 64. <https://doi.org/10.1128/AAC.00819-20>
61. Newman JA, Douangamath A, Yadzani S et al (2021) Structure, mechanism and crystallographic fragment screening of the SARS-CoV-2 NSP13 helicase. *Nat Commun* 12:4848. <https://doi.org/10.1038/s41467-021-25166-6>
62. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings 1PII of original article: S0169-409X(96)00423-1. The article was originally published in *Advanced Drug Delivery Reviews* 23 (1997). *Adv Drug Deliv Rev* 46:3–26. [https://doi.org/10.1016/S0169-409X\(00\)00129-0](https://doi.org/10.1016/S0169-409X(00)00129-0)
63. Su H, Yao S, Zhao W et al (2021) Identification of pyrogallol as a warhead in design of covalent inhibitors for the SARS-CoV-2 3CL protease. *Nat Commun* 12:3623. <https://doi.org/10.1038/s41467-021-23751-3>
64. Su H, Yao S, Zhao W, feng et al (2020) Anti-SARS-CoV-2 activities in vitro of Shuanghuanglian preparations and bioactive ingredients. *Acta Pharmacol Sin* 41:1167–1177. <https://doi.org/10.1038/s41401-020-0483-6>
65. Zhang CH, Stone EA, Deshmukh M et al (2021) Potent Noncovalent Inhibitors of the Main Protease of SARS-CoV-2 from Molecular Sculpting of the Drug Perampanel Guided by Free Energy Perturbation Calculations. *ACS Cent Sci* 7:467–475. <https://doi.org/10.1021/acscentsci.1c00039>
66. Yang KS, Ma XR, Ma Y et al (2021) A Quick Route to Multiple Highly Potent SARS-CoV-2 Main Protease Inhibitors**. *ChemMedChem* 16:942–948. <https://doi.org/10.1002/cmdc.202000924>
67. Jin Z, Du X, Xu Y et al (2020) Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature* 582:289–293. <https://doi.org/10.1038/s41586-020-2223-y>
68. Wang Z, Zhao Y, Wang Q et al (2021) Identification of proteasome and caspase inhibitors targeting SARS-CoV-2 Mpro. *Signal Transduct. Target. Ther*
69. Brown DG, Wobst HJ (2021) A Decade of FDA-Approved Drugs (2010–2019): Trends and Future Directions. *J Med Chem* 64:2312–2338. <https://doi.org/10.1021/acs.jmedchem.0c01516>
70. Hoffman RL, Kania RS, Brothers MA et al (2020) Discovery of Ketone-Based Covalent Inhibitors of Coronavirus 3CL Proteases for the Potential Therapeutic Treatment of COVID-19. *J Med Chem* 63:12725–12747. <https://doi.org/10.1021/acs.jmedchem.0c01063>
71. Kneller DW, Galanie S, Phillips G et al (2020) Malleability of the SARS-CoV-2 3CL Mpro Active-Site Cavity Facilitates Binding of Clinical Antivirals. *Structure* 28:1313–1320e3. <https://doi.org/10.1016/j.str.2020.10.007>
72. Qiao J, Li Y-S, Zeng R et al (2021) SARS-CoV-2 Mpro inhibitors with antiviral activity in a transgenic mouse model. *Science* 371:1374–1378. <https://doi.org/10.1126/science.abf1611>

73. Su H, Zhou F, Huang Z et al (2021) Molecular Insights into Small-Molecule Drug Discovery for SARS-CoV-2. *Angew Chemie* 133:9873–9886. <https://doi.org/10.1002/ange.202008835>
74. Fu L, Ye F, Feng Y et al (2020) Both Boceprevir and GC376 efficaciously inhibit SARS-CoV-2 by targeting its main protease. *Nat Commun* 11:4417. <https://doi.org/10.1038/s41467-020-18233-x>
75. Ma C, Sacco MD, Hurst B et al (2020) Boceprevir, GC-376, and calpain inhibitors II, XII inhibit SARS-CoV-2 viral replication by targeting the viral main protease. *Cell Res* 30:678–692. <https://doi.org/10.1038/s41422-020-0356-z>
76. Rut W, Lv Z, Zmudzinski M et al (2020) Activity profiling and crystal structures of inhibitor-bound SARS-CoV-2 papain-like protease: A framework for anti-COVID-19 drug design. *Sci Adv* 6:eabd4596. <https://doi.org/10.1126/sciadv.abd4596>
77. Shin D, Mukherjee R, Grewe D et al (2020) Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity. *Nature* 587:657–662. <https://doi.org/10.1038/s41586-020-2601-5>
78. Gao X, Qin B, Chen P et al (2021) Crystal structure of SARS-CoV-2 papain-like protease. *Acta Pharm Sin B* 11:237–245. <https://doi.org/10.1016/j.apsb.2020.08.014>
79. Islam R, Parves MR, Paul AS et al (2020) A molecular modeling approach to identify effective antiviral phytochemicals against the main protease of SARS-CoV-2. *J Biomol Struct Dyn* 1–12. <https://doi.org/10.1080/07391102.2020.1761883>
80. Choudhary S, Malik YS, Tomar S (2020) Identification of SARS-CoV-2 Cell Entry Inhibitors by Drug Repurposing Using in silico Structure-Based Virtual Screening Approach. *Front Immunol* 11:1664. <https://doi.org/10.3389/fimmu.2020.01664>
81. Sinha SK, Shakya A, Prasad SK et al (2020) An in-silico evaluation of different Saikosaponins for their potency against SARS-CoV-2 using NSP15 and fusion spike glycoprotein as targets. *J Biomol Struct Dyn* 1–12. <https://doi.org/10.1080/07391102.2020.1762741>
82. Alexpandi R, De Mesquita JF, Pandian SK, Ravi AV (2020) Quinolines-Based SARS-CoV-2 3CLpro and RdRp Inhibitors and Spike-RBD-ACE2 Inhibitor for Drug-Repurposing Against COVID-19: An in silico Analysis. *Front Microbiol* 11:1796. <https://doi.org/10.3389/fmicb.2020.01796>
83. Kalhor H, Sadeghi S, Abolhasani H et al (2020) Repurposing of the approved small molecule drugs in order to inhibit SARS-CoV-2 S protein and human ACE2 interaction through virtual screening approaches. *J Biomol Struct Dyn* 1–16. <https://doi.org/10.1080/07391102.2020.1824816>
84. Feng S, Luan X, Wang Y et al (2020) Eltrombopag is a potential target for drug intervention in SARS-CoV-2 spike protein. *Infect Genet Evol* 85:104419. <https://doi.org/10.1016/j.meegid.2020.104419>
85. Wei T, Wang H, Wu X, qing et al (2020) In Silico Screening of Potential Spike Glycoprotein Inhibitors of SARS-CoV-2 with Drug Repurposing Strategy. *Chin J Integr Med* 26:663–669. <https://doi.org/10.1007/s11655-020-3427-6>
86. Yu J, Wang L, Bao L (2020) Exploring the active compounds of traditional Mongolian medicine in intervention of novel coronavirus (COVID-19) based on molecular docking method. *J Funct Foods* 71:104016. <https://doi.org/10.1016/j.jff.2020.104016>

87. Ren X, Shao X-X, Li X-X et al (2020) Identifying potential treatments of COVID-19 from Traditional Chinese Medicine (TCM) by using a data-driven approach. *J Ethnopharmacol* 258:112932. <https://doi.org/10.1016/j.jep.2020.112932>
88. Xia S, Yan L, Xu W et al (2019) A pan-coronavirus fusion inhibitor targeting the HR1 domain of human coronavirus spike. *Sci Adv* 5:eaav4580. <https://doi.org/10.1126/sciadv.aav4580>
89. Joshi T, Joshi T, Sharma P et al (2020) In silico screening of natural compounds against COVID-19 by targeting Mpro and ACE2 using molecular docking. *Eur Rev Med Pharmacol Sci* 24:4529–4536. https://doi.org/10.26355/eurrev_202004_21036
90. Monteil V, Kwon H, Prado P et al (2020) Inhibition of SARS-CoV-2 Infections in Engineered Human Tissues Using Clinical-Grade Soluble Human ACE2. *Cell* 181:905–913e7. <https://doi.org/10.1016/j.cell.2020.04.004>
91. Kruse RL (2020) Therapeutic strategies in an outbreak scenario to treat the novel coronavirus originating in Wuhan, China. <https://doi.org/10.12688/f1000research.22211.2>. *F1000Research* 9:72
92. Zhang G, Pomplun S, Loftis AR et al (2020) Investigation of ACE2 N-terminal fragments binding to SARS-CoV-2 Spike RBD. *bioRxiv*
93. Ahmed SF, Quadeer AA, McKay MR (2020) Preliminary Identification of Potential Vaccine Targets for the COVID-19 Coronavirus (SARS-CoV-2) Based on SARS-CoV Immunological Studies. *Viruses* 12:254. <https://doi.org/10.3390/v12030254>
94. Huang X, Pearce R, Zhang Y (2020) De novo design of protein peptides to block association of the SARS-CoV-2 spike protein with human ACE2. *Aging* 12:11263–11276. <https://doi.org/10.18632/aging.103416>
95. Wang Y, Zhang D, Du G et al (2020) Remdesivir in adults with severe COVID-19: a randomised, double-blind, placebo-controlled, multicentre trial. *Lancet* 395:1569–1578. [https://doi.org/10.1016/S0140-6736\(20\)31022-9](https://doi.org/10.1016/S0140-6736(20)31022-9)
96. Goldman JD, Lye DCB, Hui DS et al (2020) Remdesivir for 5 or 10 Days in Patients with Severe Covid-19. *N Engl J Med* 383:1827–1837. <https://doi.org/10.1056/nejmoa2015301>
97. Spinner CD, Gottlieb RL, Criner GJ et al (2020) Effect of Remdesivir vs Standard Care on Clinical Status at 11 Days in Patients with Moderate COVID-19: A Randomized Clinical Trial. *JAMA - J Am Med Assoc* 324:1048–1057. <https://doi.org/10.1001/jama.2020.16349>
98. Beigel JH, Tomashek KM, Dodd LE et al (2020) Remdesivir for the Treatment of Covid-19 – Final Report. *N Engl J Med* 383:1813–1826. <https://doi.org/10.1056/NEJMoa2007764>
99. Chen C, Zhang Y, Huang J et al (2020) Favipiravir versus Arbidol for COVID-19: A randomized clinical trial. *medRxiv*
100. Shrestha DB, Budhathoki P, Khadka S et al (2020) Favipiravir versus other antiviral or standard of care for COVID-19 treatment: a rapid systematic review and meta-analysis. *Virol J* 17:141. <https://doi.org/10.1186/s12985-020-01412-z>
101. Smith EC, Blanc H, Vignuzzi M, Denison MR (2013) Coronaviruses Lacking Exoribonuclease Activity Are Susceptible to Lethal Mutagenesis: Evidence for Proofreading and Potential Therapeutics. *PLoS*

102. Ferron F, Subissi L, De Morais ATS et al (2018) Structural and molecular basis of mismatch correction and ribavirin excision from coronavirus RNA. *Proc Natl Acad Sci U S A* 115:E162–E171. <https://doi.org/10.1073/pnas.1718806115>
103. Yin W, Mao C, Luan X et al (2020) Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. *Science* 368:1499–1504. <https://doi.org/10.1126/science.abc1560>
104. Elfiky AA (2020) Anti-HCV, nucleotide inhibitors, repurposing against COVID-19. *Life Sci* 248:117477. <https://doi.org/10.1016/j.lfs.2020.117477>
105. Yin W, Luan X, Li Z et al (2021) Structural basis for inhibition of the SARS-CoV-2 RNA polymerase by suramin. *Nat Struct Mol Biol* 28:319–325. <https://doi.org/10.1038/s41594-021-00570-0>

Figures

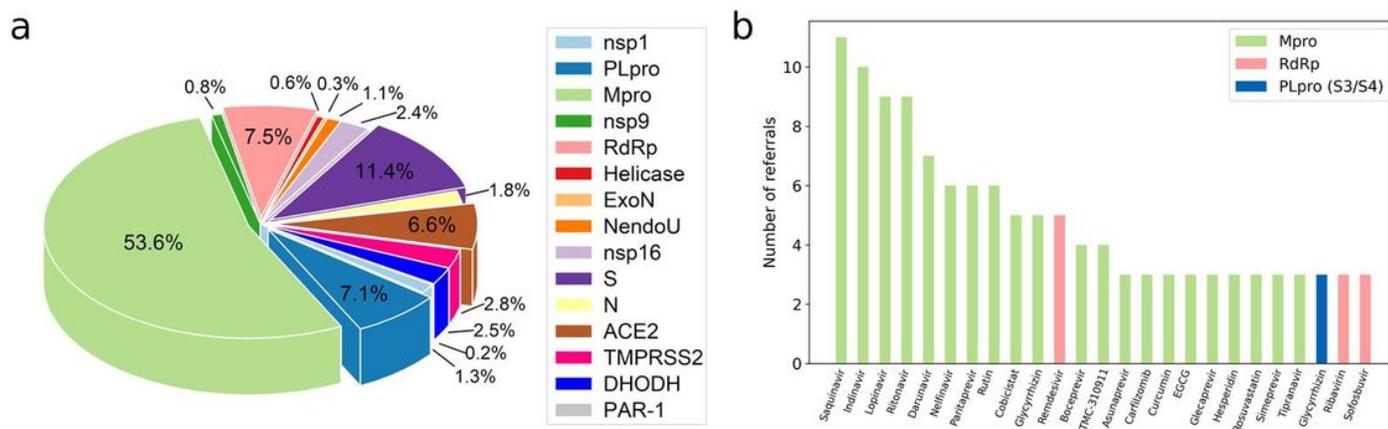


Figure 1

The statistics of compound records retrieved from CADD studies. a. The distribution of compound record numbers classified by drug targets. b. The CADD recommended compounds reported in more than two publications. Their targets included Mpro, RdRp and PLpro. S3 and S4 are the two binding sites of PLpro.

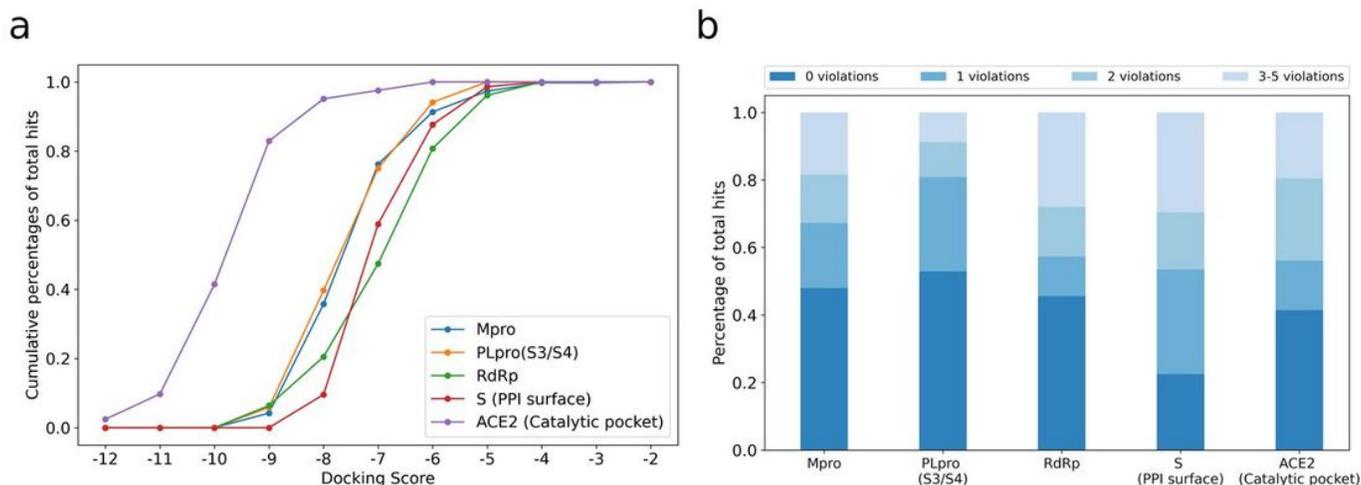


Figure 2

The distribution of docking scores and drug-like properties of CADD recommended compounds. a. The cumulative percentage of CADD recommended compounds with docking scores increasing. b. The percentage of CADD compounds following the Lipinski's rules. 0 violations indicates the compounds that follow the Lipinski's "Rule of Five": $MW \leq 600$, $HBA \leq 10$, $HBD \leq 5$, $NrotB \leq 10$ and $-2 \leq cLogP \leq 5$; 1 violation and 2 violations indicates the compounds that violate one and two of the above five rules respectively; 3-5 violations indicates the compounds that violate at least three of the five rules.

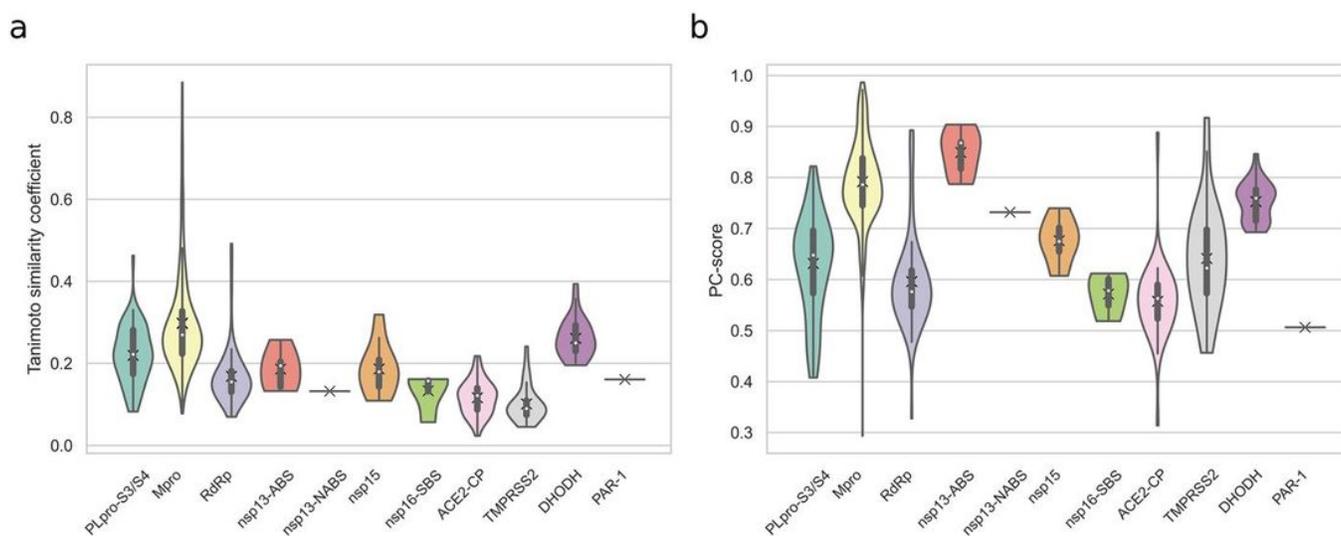


Figure 3

The violin plots of molecular similarities between the CADD recommended compounds and known inhibitors for each drug target. a. The distribution of Tanimoto similarity coefficients between the CADD

recommended compounds and known inhibitors. b. The distribution of PC-score between the CADD recommended compounds and known inhibitors. The targets and binding sites include PLpro S3/S4 pocket (PLpro-S3/S4), catalytic pocket of Mpro, RdRp, nsp15, ACE2 (ACE2-CP), TMPRSS2 and DHODH, the ATP binding site (nsp13-ABS) and nucleic acids binding site (nsp13-NABS) of nsp13, the SAM binding site of nsp16 (nsp16-SBS) and the active site of PAR-1. The position of 'X' and the white points represent the average values and median values respectively and the violin range is limited within the range of the observed data.

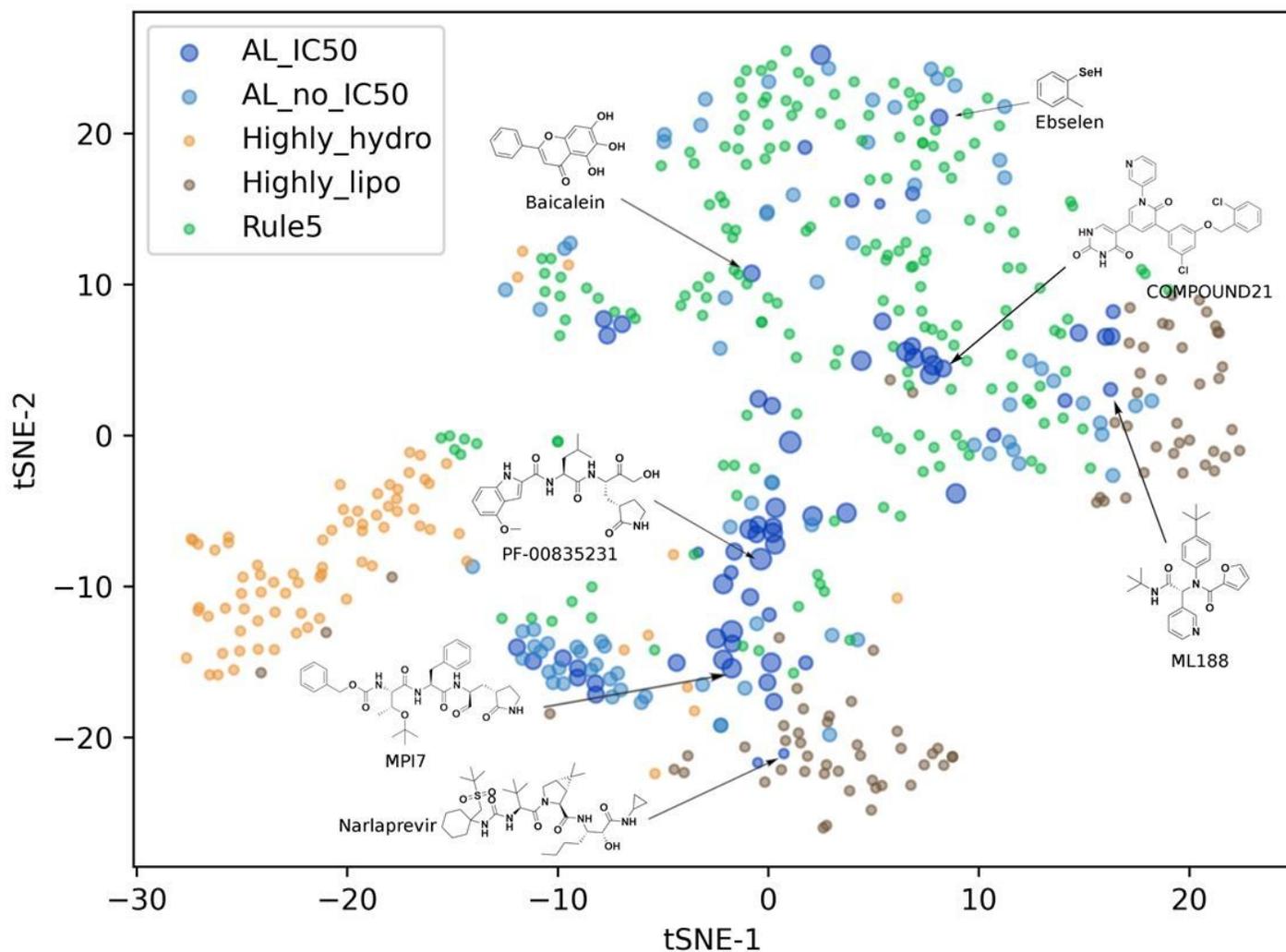


Figure 4

The t-distributed stochastic neighbor embedding (tSNE) map of the chemical spatial features including cLogP (calculated lipophilicity), HBA (hydrogen-bond acceptor), HBD (hydrogen-bond donor) and MW (molecular weight) for each CADD recommended compounds and the co-crystallized inhibitors against Mpro. AL_IC50 represents the inhibitors with known IC50. The size of the circles refers to the IC50 of the compounds (The larger size indicates the smaller IC50). AL_no_IC50 represents the inhibitors with unknown activity data; Rule5 represents the CADD compounds that meet the criteria of $MW \leq 600$, $HBA \leq 10$, $HBD \leq 5$, and $-2 \leq cLogP \leq 5$; Highly_lipo refers to the CADD recommended compounds that violate one or more of the above criteria and $cLogP > 2$ (i.e., compounds with high lipophilicity);

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Graphicalabstracts.png](#)
- [FigS1.tif](#)
- [FigS2.tif](#)
- [TableS1.xlsx](#)
- [TableS2.xlsx](#)
- [TableS3.docx](#)
- [TableS4.xlsx](#)
- [TableS5.docx](#)