Side chain similarity comparisons for integrated drug repositioning and potential toxicity assessments in epidemic response scenarios: The case for COVID-19

Nur Syatila Ab Ghani [a], Reeki Emrizal [b], Haslina Makmur [b], Mohd Firdaus-Raih [a,b,*]

[a] Institute of Systems Biology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia
[b] Department of Applied Physics, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia

ARTICLE INFO

ABSTRACT

Structures of protein-drug-complexes provide an atomic level profile of drug-target interactions. In this work, the three-dimensional arrangements of amino acid side chains in known drug binding sites (substructures) were used to search for similarly arranged sites in SARS-CoV-2 protein structures in the Protein Data Bank for the potential repositioning of approved compounds. We were able to identify 22 target sites for the repositioning of 16 approved drug compounds as potential therapeutics for COVID-19. Using the same approach, we were also able to investigate the potentially promiscuous binding of the 16 compounds to off-target sites that could be implicated in toxicity and side effects that had not been provided by any previous studies. The investigations of binding properties in disease-related proteins derived from the comparison of amino acid substructure arrangements allows for effective mechanism driven decision making to rank and select only the compounds with the highest potential for success and safety to be prioritized for clinical trials or treatments. The intention of this work is not to explicitly identify candidate compounds but to present how an integrated drug repositioning and potential toxicity pipeline using side chain similarity searching algorithms are of great utility in epidemic scenarios involving novel pathogens. In the case of the COVID-19 pandemic caused by the SARS-CoV-2 virus, we demonstrate that the pipeline can identify candidate compounds quickly and sustainably in combination with associated risk factors derived from the analysis of potential off-target site binding by the compounds to be repurposed.

1. Introduction

Epidemics caused by novel infectious agents result in situations where no known treatment regimens are in practice. Case management would therefore first rely on treating and alleviating the symptoms. The focus of the treatment would then move on to eradication of the infectious agent from the host and more in-depth therapeutic management. Such an epidemic scenario presented itself in the city of Wuhan, Hubei Province, China in late 2019 [1]. The causative pathogen for the observed acute respiratory distress was later identified to be a novel human coronavirus (nCoV19) named as Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) [2]. Although, many coronaviruses are found in bat reservoirs, it is probable that SARS-CoV-2 also has intermediate hosts such as pangolins and snakes [3].

Three months after it was first reported, the disease, named coronavirus disease 2019 (COVID-19), had progressed into a global pandemic. The fast spread of the disease was however paralleled by the speed that data regarding the disease and its causative agent were generated. In mid-January 2020, the first genome sequence was deposited into GenBank (https://www.ncbi.nlm.nih.gov/genbank/); by mid-July 2020, more than 40,000 complete genomes with high coverage from samples throughout the world had been deposited in the GISAID database (http://www.gisaid.org/; http://epicov.org). While the rate of genome sequencing and data sharing is unprecedented, the rapid availability of structure data has also been equally impressive. In late September 2020, more than 400 structures of SARS-CoV-2 proteins had been deposited in the Protein Data Bank (PDB) [4].

* Corresponding author at: Department of Applied Physics, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia.
E-mail address: firdaus@mfrlab.org (M. Firdaus-Raih).

Despite the number of confirmed cases passing 32.9 million with more than 1 million fatalities worldwide in early October 2020, treatment options are still lacking for COVID-19 although several vaccines have recently started their trials in July 2020. This dire but data-rich scenario has led investigators to resort to drug repurposing strategies. Although such efforts to reposition approved drugs to a new target can be explored in a clinical setting, we focus specifically on how computational approaches can feature prominently in the identification of the candidate compounds.

Various approaches have been deployed to explore the repertoire of known and approved compounds for COVID-19. Zhou *et al.* utilized network-based analyses of drug targets and the virus-host interactions in the human interactome to list 16 potential drugs to prioritize for repurposing [5]. An even larger effort that generated a SARS-CoV-2-Human protein-protein interaction map was able to identify 66 druggable human proteins that could be targeted by 69 currently available FDA approved compounds to be used as COVID-19 treatments [6].

Side chain similarity comparisons [7,8] have been reported to be a potential starting point in drug repurposing efforts [9]. For such an approach, the 3D arrangements of known drug binding sites are collected as a search database to identify similar sites in non-homologous structures thus implying the capacity to bind similar ligands. Drug-target interaction prediction using structural data has remained a largely unexplored niche [10]. The identification of possible alternative binding sites for an approved drug can also provide insights into their possible off-target effects. There is a clear urgency to discover and deploy suitable candidates that can be repositioned against targets associated with COVID-19. Nevertheless, it is prudent to steer clear of adverse effects resulting from the poly-pharmacological actions of promiscuous drugs with the ability to bind to other targets [11].

In this work, amino acid side chain similarity searching was utilized to propose alternative target sites in SARS-CoV-2 protein structures for drug repositioning. These searches were based on the premise that if a known drug binding site could be found in a SARS-CoV-2 protein, then that protein could also serve as an alternative target for the same drug. This same principle was then used to identify off-target sites that could present as side effects or result in some form of toxicity. The list of potential drugs derived from the side chain arrangements similarity searches was then used to propose structurally similar compounds that could also target the sites already identified for repositioning. Our approach differs significantly from that reported by Zhou *et al.* [5] and Gordon *et al.* [6] which can serve as additional confirmatory analysis and complement the gaps in existing work. The details of these differences will be discussed in a later section.

2. Materials and methods

2.1. SARS-CoV-2 protein structure coordinate data and drug compounds dataset

All SARS-CoV-2 protein structure coordinate data were sourced from the RCSB Protein Data Bank (PDB) [4]. The most recent structure used in this study was retrieved on 28th August 2020, resulting in a total of 351 PDB structures. The associated protein sequences and annotations were also retrieved from the PDB. The downloaded sequences were clustered at 90% sequence similarity cut-off using the CD-HIT program [12]. Members of individual clusters were sorted according to the X-ray crystallography resolution; the SARS-CoV-2 protein sequence with the higher resolution structure was selected as the cluster's representative. The PDB structures containing representative sequences were compiled together for further similarity searches against the dataset of known drug binding sites derived from protein-drug complexes in the PDB.

For the selection of drug compounds, we further selected drug compounds that are: (i) currently undergoing clinical trials for COVID-19, and (ii) those that have PDB ligand identifiers. Binding site-ligand contacts for these compounds were obtained from Drug ReposER application [9] and the binding sites were compiled for sub-structural similarity searching using the ASSAM (Amino Acid Pattern Search for Substructures and Motifs) computer program [7].

2.2. Searching for sites in SARS-CoV-2 protein structures similarly arranged as binding sites for approved drug molecules through sub-structural similarity searches and molecular docking

Binding sites for selected drug compounds derived from protein-drug complexes (in Section 2.1) were used as inputs for the computer program ASSAM [13] to find similar arrangements of amino acids in a set of representative SARS-CoV-2 protein structures. Amino acid residues that are within 4.0 Å of a drug molecule were considered to be binding site residues. Both, the inputs (SARS-CoV-2 protein structures) and outputs (similar site, matched protein-drug complex structure) of the sub-structural similarity searches were then used for molecular docking.

For individual matches of sites between the SARS-CoV-2 protein (query protein) and the matched protein-drug complex (hit protein), a Python script was designed to set up the automatic molecular docking to be performed using the Autodock Vina module [14] embedded in the UCSF Chimera [15] molecular visualization program. The drug molecule from the hit protein was used as the ligand and the SARS-CoV-2 protein was used as the receptor structure for docking. The Python script contains all the necessary commands that will be executed in the UCSF Chimera command line to automatically pre-process structures and perform blind molecular docking. The pre-processing steps of the ligand and receptor structures include the removal of water molecules and ligands, assigning the partial charges for both standard and non-standard residues, as well as an additional energy-minimization step. The atomic partial charges for standard residues including standard amino acids, water and know ligands, as well as non-standard residues were assigned based on the AMBER ff14SB force field (default), while the partial charges for non-standard residues were calculated using the Antechamber module based on the AM1-BBC method. In the case of residues with missing side chains, the amino acid side chains were replaced based on information from a rotamer library. Energy minimization was performed with steps of steepest descent minimization set to 100. Molecular docking was carried out using a local installation of Autodock Vina and linked for use in UCSF Chimera.

Blind docking was carried out instead of using the binding site as a reference point. Therefore, a whole protein structure target was exhaustively searched for potential binding poses using the default settings for parameters such as exhaustiveness value (set to 8) and maximum number of binding modes (set to 9). The default box size was used to sample the ligand orientation where it automatically covers the entire protein receptor thus allowing for matches of binding poses to not only known binding sites, but also to other putative sites that have not been reported elsewhere.

Upon completion of the docking run using the Python script, UCSF Chimera loads a selection of docking poses for visualization where the docking poses are ranked according to the docking scores reported in kcal/mol with more negative values indicating better binding. The sites found from the sub-structural similarity search is also visualized. The UCSF Chimera session for individual script runs were saved for further curation and analysis. The sites

from the sub-structural similarity search were compared against the sites in pre-computed binding poses from molecular docking, where an overlap of at least three matched residues with poses of docking scores more negative than −6.5 kcal/mol selected for further analyses.

2.3. Searching for potential off-targets from human for selected drugs proposed for COVID-19

Potential off-targets for selected drugs proposed for COVID-19 were identified using three different methods. First, known human proteins bound to the selected drug compounds were obtained from the PDB through the 'Advanced Search' interface in the RCSB using the ligand PDB ID as a query. The list of PDB structures retrieved were filtered to only contain PDB with organism denoted as 'Homo Sapiens'. Second, human proteins with similarly arranged sites to drug binding sites for the selected drugs were retrieved from pre-compiled results for sub-structural similarity searches in Drug ReposER web server. Third, human proteins with more than 30% sequence similarity to individual SARS-CoV-2 protein structures were retrieved from blastp searches against the PDB. The list of proteins retrieved was filtered to only contain proteins with sequences more than 30% sequence identity to the query SARS-CoV-2 protein.

These human structures were then used for molecular docking against the selected compounds. Molecular docking runs were conducted based on the above-mentioned protocol using Python scripts executed in UCSF Chimera. A compound's involvement in specific biological mechanisms and potential adverse effects upon interaction with the selected compounds were manually assessed and extracted from information available in UniProtKB [16] and literature mining.

2.4. Screening for novel drugs for COVID-19 using drug ReposER

Structurally similar ligands to the set of drugs retrieved in this study were identified using the chemical component search feature available in the RCSB PDB (https://www.rcsb.org/pdb/ligand/chemAdvSearch.do?chemCompId=) with a structure similarity threshold of 70%. Similar ligands annotated as approved drugs in DrugBank were further selected. For validation, both the queried and similar ligands were structurally aligned in the UCSF Chimera interface [15].

The queried and the similar ligands were individually searched against the Drug ReposER application database to retrieve results for sub-sructural similarity searches. Both sets of results were compared and shared SARS-CoV-2 protein targets from the list of proteins (proteins containing sites similar to binding sites for both queried and similar ligands) were obtained for molecular docking against the corresponding ligand molecules with Autodock Vina using the above-mentioned protocol [15].

3. Results and discussion

In this study, sub-structural similarity searches and docking analyses were carried out to: (i) identify potential targets and drug binding sites in SARS-CoV-2 proteins; (ii) identify off-targets for proposed drug compounds for COVID-19; (iii) identify other approved drugs with similar structure to proposed drugs that are potentially useful for COVID-19 treatment. A total of 351 SARS-CoV-2 proteins were obtained from the PDB that included the following proteins; ADP ribose phosphatase (PDBID: 6w02), spike protein (PDBID: 6vsb), main protease (PDBID: 6lu7), nucleocapsid (PDBID: 6m3m), NSP7-NSP8 complex (PDBID: 6yhu), NSP9 replicase (PDBID: 6w4b), NSP10-NSP16 complex (PDBID: 6w4h),

NSP15 (PDBID: 6w01), ORF7a encoded accessory protein (PDBID: 6w37) and RNA-dependent RNA polymerase or NSP12 (PDBID: 6m71).

The substructure similarity searching used in this work utilized the ASSAM computer program which solves a maximal common subgraph problem to match similar 3D arrangements of amino acids in a dataset of protein structures [7]. The arrangements of amino acids in 3D space are represented as graphs, where the graph nodes are the pseudo-atoms representing side chain groups and the graph edges are distances between the side chain groups. Using this scheme, it is possible to match similar 3D arrangements, such as catalytic sites and ligand binding sites, in non-homologous structures. Drug ReposER is an extended application of the ASSAM program that focuses on sub-structures that constitute the binding sites for approved drug molecules [9].

At the time of writing, approximately a third of the proteins encoded in the SARS-CoV-2 genome have corresponding PDB structures. In anticipation that more structures will be deposited, we have enabled the analysis pipeline to be deployed to process new structures as and when they become available, based on the clustering of protein sequences and comparison to readily available structures. The results from the analyses reported in this work and those that will be carried out by the pipeline for new structures will be made accessible via a dedicated module of the Drug ReposER web application – http://mfrlab.org/drugreposer/covid19/. The list of PDB IDs with pre-computed results from sub-structural similarity searches and the sequence clusters are also available at the same resource.

The search for COVID-19 treatments has resulted in the registration of more than 3000 clinical trials in the ClinicalTrials.gov database to explore the repurposing of more than twenty readily available drugs (https://clinicaltrials.gov/ct2/results?cond=COVID-19) [17]. This number includes completed studies, ongoing studies currently under recruitment, or those currently enrolling by invitation. There are also a number of clinical trials registered in clinicaltrials.gov that have not yet recruited any participants at this point in time.

3.1. Approved drugs as potential treatment for COVID-19 based on sub-structural similarity to known drug binding sites

Searching for sites in the SARS-CoV-2 protein structures (hit sites) that are geometrically similar to sites for approved drug compounds (query sites) using the Drug ReposER application [9] had identified matches that included 22 sites from protein-drug complexes with sequence identities lesser than 30% to the corresponding SARS-CoV-2 proteins (Table 1). These results show that the computational approach adopted in this study is able to find similarly arranged sites in unrelated proteins which could be an advantage when there are limited numbers of homologous structural models to be used for comparison of binding sites. In addition, the selection of matches to proteins with lesser than 30% sequence similarity could be indicative of function differences, thus potentially distinct pathways where the bound drugs could be repurposed to.

The sites identified in the SARS-CoV-2 proteins were then docked with their corresponding drug compounds derived from the protein-drug complex data. Molecular docking runs resulted in the identification of several poses with docking scores ranging from −6.0 kcal/mol up to −17.6 kcal/mol, which are congruent with the results of the Drug ReposER searches (Table 1, Fig. 1). Of these 22 potential interactions, six have been reported in other studies [18–20].

The sub-structural similarity searches carried out revealed that six of the nine analysed SARS-CoV-2 contain multiple potential alternative binding sites for different compounds. For example,

Table 1
Sub-structural similarities of known drug binding sites in SARS-CoV-2 protein structures.

Alternate target in SARS-CoV-2	PDBID	Drug ID[1]	Known target	% seq. identity	Docking Score
ADP ribose phosphatase	6w02B	CLQ *[18]	4fglB (Quinone reductase 2)	21.54	−7.5
		LOC	3ut5B (Tubulin beta chain)	18.02	−7.6
		017 *[18]	6dh3A (HIV protease)	17.92	−9.5
		AB1	2qhcA (HIV protease retropepsin)	18.50	−17.6
		RIT	1rl8A (HIV protease retropepsin)	17.92	−12.0
	6w6yB	NPS	3nt1 (Prostaglandin-endoperoxide synthase 2)	14.09	−6.7
Spike protein	6vybC	017 *[18]	3lzvA (HIV protease)	6.21	−6.6
	6w41C/H	RIT *[18]	1rl8A (HIV protease retropepsin)	21.88	−12.2
Main protease	6lu7A	VIA *[19]	2h42A (PDE5A)	22.34	−7.8
	6y2gA/B	FOL	4i13A (Dihydrofolate reductase)	17.92	−8.2
NSP10/16	6w75B/D	RIT *[20]	1sh9 (Pol polyprotein)	20.00	−7.6
NPS10	6w4hB	IMN	1z9hD (Membrane-associated prostaglandin E synthase-2)	20.00	−6.7
NSP15	6vwwA	LSN	5x24A (Cytochrome P450)	21.59	−7.5
	6w01B	NPS	4po0A (Serum albumin)	19.09	−7.3
Nucleocapsid	6m3mA/D	LSN	5x24A (cytochrome P450)	16.34	−8.5
	6m3mB	RIT	3tneB (Aspartic protease)	19.01	−7.2
	6m3mB/C	017	3so9 (HIV protease)	23.85	−8.6
NSP7/NSP8	6yhuB	AB1 *[21]	2rkgB (Protease retropepsin)	N.A	−17.9
	6yhuC/D	IMN	4coxA (Cyclooxygenase-2)	N.A	−7.2
	6wiqA/B	FOL	1rb2B (Dihydrofolate reductase)	N.A	−6.9
NSP8	7bv1D	FOL	3tqbA (Dihydrofolate reductase)	N.A	−7.5

[1] Drug ID is represented as follows: CLQ: chloroquine; LOC: colchicine; RIT: ritonavir; 017: darunavir; AB1: lopinavir; VIA: sildenafil; NPS: naproxen; LSN: losartan; AIN: aspirin; IMN: indomethacin; FOL: folic acid. *Denotes that the interaction had been previously reported by the accompanying citation.
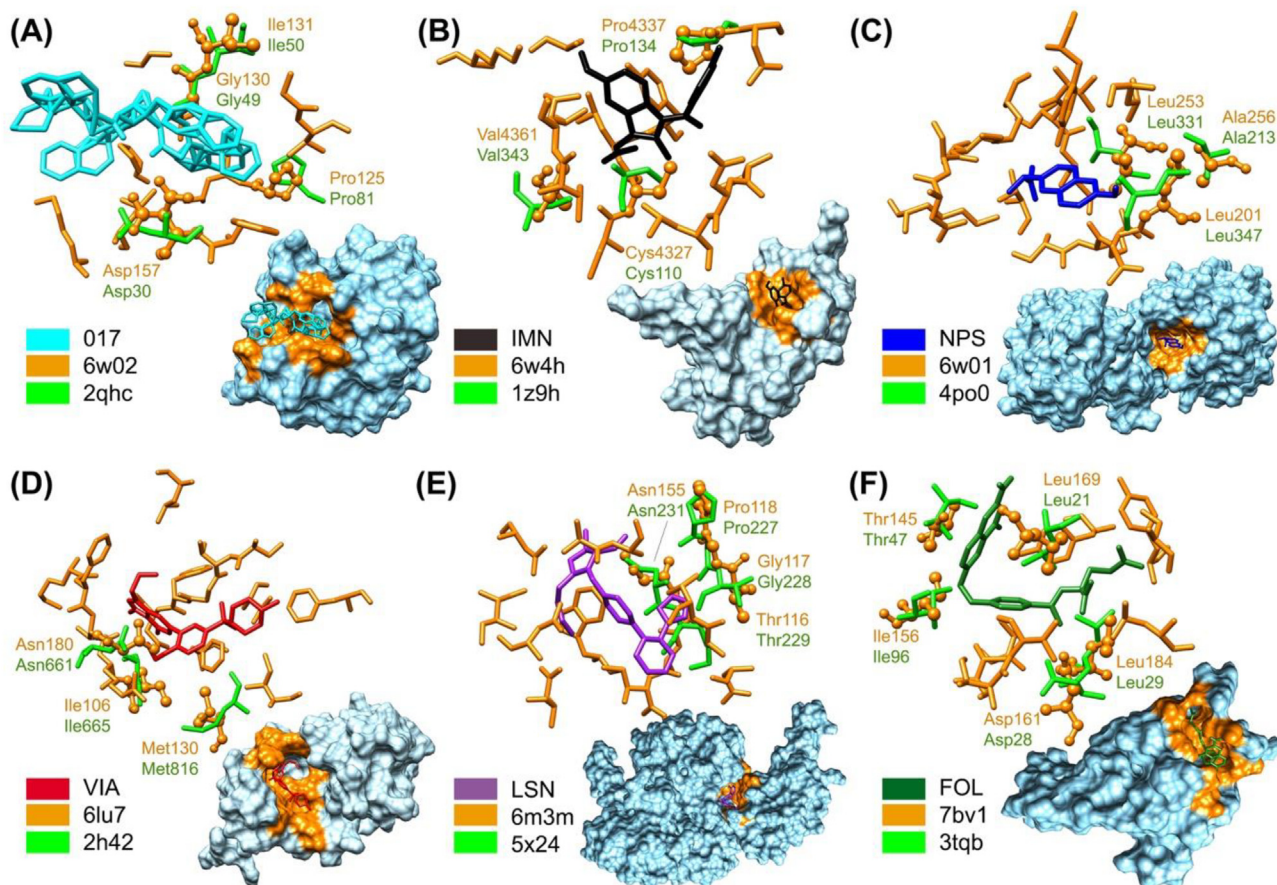
Fig. 1. Sub-structural similarity and poses of docked ligands from Autodock Vina. Predicted binding residues to docked ligands are indicated in orange, while ball and stick representations of atoms colored in orange indicate the residues identified by Drug ReposER that are similarly arranged to binding sites in known targets (green). The docked ligand is presented on the potential target protein from SARS-CoV-2 (light blue). **(A)** ADP ribose phosphatase (PDBID: 6w02) bound to docked darunavir (017) with green colored stick representation of similarly arranged residues from HIV protease retropepsin (PDBID: 2qhc). **(B)** NSP10 (PDBID: 6w4h) bound to docked indomethacin (IMN) with green colored stick representation of similarly arranged residues from Membrane-associated prostaglandin E synthase-2 (PDB: 1z9h). **(C)** NSP15 (PDBID: 6w01) bound to docked naproxen (NPS) with similarly arranged residues from serum albumin highlighted in green **(D)** Main protease (PDBID: 6lu7) bound to docked sildenafil (VIA) with superposed residues from PDE5A (PDBID: 2h42). **(E)** Docked losartan (LSN) in nucleocapsid (PDBID: 6m3m) with superposed losartan binding residues cytochrome P450 (PDBID: 5x24) indicated in green. **(F)** Docked folic acid (FOL) bound to NSP8 (PDBID: 7bv1) that has similar arrangement to folic acid sites in dihydrofolate reductase (PDBID: 3tqb). The locations of proposed binding sites are highlighted in orange color. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the nucleocapsid protein contains potential sites for losartan, ritonavir, darunavir and aspirin (Table 1). The Drug ReposER searches also identified several compounds that had potential binding sites in different SARS-CoV-2 structures; for example, binding sites to ritonavir (RIT) could be found in four different structures – ADP ribose phosphatase, spike protein, NSP10/16 and nucleocapsid (Table 1). Interestingly, six of the twenty-two matches are to Human Immunodeficiency Virus (HIV) structures bound to anti-retrovirals such as darunavir, lopinavir and ritonavir.

The HIV protease inhibitors - darunavir (017), ritonavir (RIT) and lopinavir (AB1) - inhibit the HIV aspartyl protease and prevents the cleavage of Gag and Pol proteins into their subsequent protein components [22]. The potential antiviral activity of such inhibitors against coronaviruses had been previously studied; nelfinavir for example, had been reported to inhibit the replication of SARS-CoV and prevent cytopathic effects [23]. Lopinavir and ritonavir had been shown to improve clinical outcomes from SARS-CoV infections and are hypothesized to bind to the 3-chymotrypsin-like protein (3CLpro) or main protease [24]. Our analysis also demonstrated the potential ability of lopinavir (PDBID: 2qhc) to bind to ADP ribose phosphatase (PDBID: 6w02) with a docking score of −17.6 kcal/mol at a position close to the known substrate binding site (Fig. 1A).

We found that the NSP10-16 complex (PDBID: 6w75) may potentially bind to ritonavir (RIT) in a manner similar to that observed in the HIV protease (PDBID: 1sh9) while ADP ribose phosphatase (PDBID: 6w02) could potentially bind to lopinavir (PDBID: 2qhc) with a high docking score (−17.6 kcal/mol) (Fig. 1A). A potential site for folic acid (FOL) binding that is similar to the arrangement found in dihydrofolate reductase (PDBID: 4i13) was also found at the interaction site between domain III (residue 201–303) of two monomers, where dimerization is crucial for protease function took place (PDBID: 6y2g). We also found that the nucleocapsid might bind to losartan, darunavir and aspirin at the dimerization site between two monomers in a similar manner to the SARS-CoV-2 main protease.

The Drug ReposER searches also identified similarly arranged sites between the indomethacin-bound prostaglandin E synthase 2 (PDBID: 1z9h) and the NSP10 protein (PDBID: 6w4h) (Fig. 1B). An arrangement of amino acid residues that make up the indomethacin binding site in cyclooxygenase-2 (COX-2), also known as prostaglandin synthase 2 (PDBID: 4cox), was also found to be similar to residues at the vicinity of docked indomethacin in the NSP7-NSP8 complex (PDBID: 6yhu) (Table 1). The docking results suggests that indomethacin (IMN) could bind to NSP10 (PDBID: 6w4h) and the NSP7-NSP8 complex (PDBID: 6yhu) of SARS-CoV-2.

The NSP10 protein is a co-factor that can activate the 2′-O methyltransferase activity of NSP16, or the 3′-5′ exoribonuclease activity of the NSP14. The NSP10-NSP16 and NSP10-NSP14 complexes are key elements in the RNA transcription machinery of SARS-CoV-2 [25]. The NSP7 and NSP8 homologs in SARS-CoV are co-factors for NSP12, which is a key element of viral replication or transcription machinery that acts as a RNA dependent RNA polymerase [26]. Recent studies have reported the antiviral activity of indomethacin (IMN) *in vivo*, where it is able to disrupt RNA synthesis and abbreviate the damage to host cells [27]; however no insights in terms of their binding activity in a structural context have been reported. Sequence and fold comparisons revealed that the two sets of proteins, are unrelated in terms of sequence or structure.

Zhu et al. [28] had reported that elevation of prostaglandin synthase activity during viral infection of cytomegalovirus led to an increasing level of prostaglandin E2 which in turn caused an inflammatory response. In this context, the binding of indomethacin to these protein structures (NSP7/NSP8 or NSP10) may also prevent potential inflammatory events. The same mechanism could be adopted by other NSAIDs like naproxen (NPS), that might recognize similar sites from COX-2 (PDBID: 3nt1) in the ADP ribose phosphatase (PDBID: 6w6y), as indicated from the sub-structural similarity we have uncovered. These sub-structural similarities to a known indomethacin binding site may explain the mechanism for studies that have reported the ability of NSAIDs to bind to SARS-CoV-2 proteins [29] although the atomic level details of such interactions have not yet been reported.

3.2. ADP ribose phosphatase of NSP3 as potential target in SARS-CoV-2

Our analysis revealed that the ADP ribose phosphatase of NSP3 from SARS-CoV-2 has the most number of 3D residue arrangements that are similar to the binding sites in known drug targets compared to other SARS-CoV-2 proteins (Table 1, Fig. 2). All the identified sites are within the substrate binding sites with the docking scores for the different poses ranging from −6.7 to −17.0. In this case, the known ADP ribose phosphatase – APR complex was used as a control to obtain reasonable docking scores that could be considered acceptable based on predicted binding poses between the ADP ribose phosphatase and the substrate, APR. The molecular docking with energy minimization steps resulted in several binding poses with docking scores ranging from −7.9 to −9.8, with all sites located within the actual binding site for APR.

The ADP ribose phosphatase of non-structural protein 3 (NSP3) is likely to be targeted by anti-retrovirals and several other drugs more than any other SARS-CoV-2 structures, particularly at the active site of the structure (Fig. 2A). This finding is in agreement with recent computational screening for the drug binding ability of SARS-CoV-2 proteins which highlighted the promiscuity of NSP3 in binding to other molecules at the ADP ribose binding site [21,28]. The de-ADP ribosylation activity of NSP3 suppresses the expression of host innate immunity genes such as interferon and interleukin related genes [30]. Disruption of NSP3 function will allow for the host immune system to respond normally to the infection.

Sub-structural similarity searches and molecular docking runs have revealed the potential binding sites for darunavir (017) that originally targeted HIV protease (PDBID: 6dh3), as well as chloroquine (CLQ) that originally targeted quinone reductase 2 (PDBID: 4fgl) and indicated for malaria and rheumatoid arthritis, onto the ADP ribosylation site of NSP3 (PDBID: 6w02) (Table 1, Fig. 2). Despite the similarity of these sites in terms of their 3D arrangements, the similarity of their molecular functions is unlikely to be related.

The docking results indicate that HIV protease inhibitors and NSAIDs are among the existing drugs that could potentially be repositioned against ADP ribose phosphatase and several non-structural proteins for treatment of COVID-19. The similarly arranged residue patterns observed between the binding poses in SARS-CoV-2 proteins from docking simulations and those from available drug-bound protein complexes allow us to infer the similarities of the binding mechanisms shared by these proteins despite the lack of sequence similarities.

3.3. Potential off-targets of approved drugs proposed for COVID-19

The binding of drug compounds to off-target sites in proteins other than their intended targets can lead to unexpected pharmacological outcomes including the activation or disruption of molecular functions that cause adverse effects or other unexpected conditions [11,31]. However, off-target effects are not necessarily negative and it is this same concept that is in use to repurpose approved compounds for alternative indications based on the availability of similar of binding sites shared among proteins involved in distinct disease pathways [11,31]. We deployed the
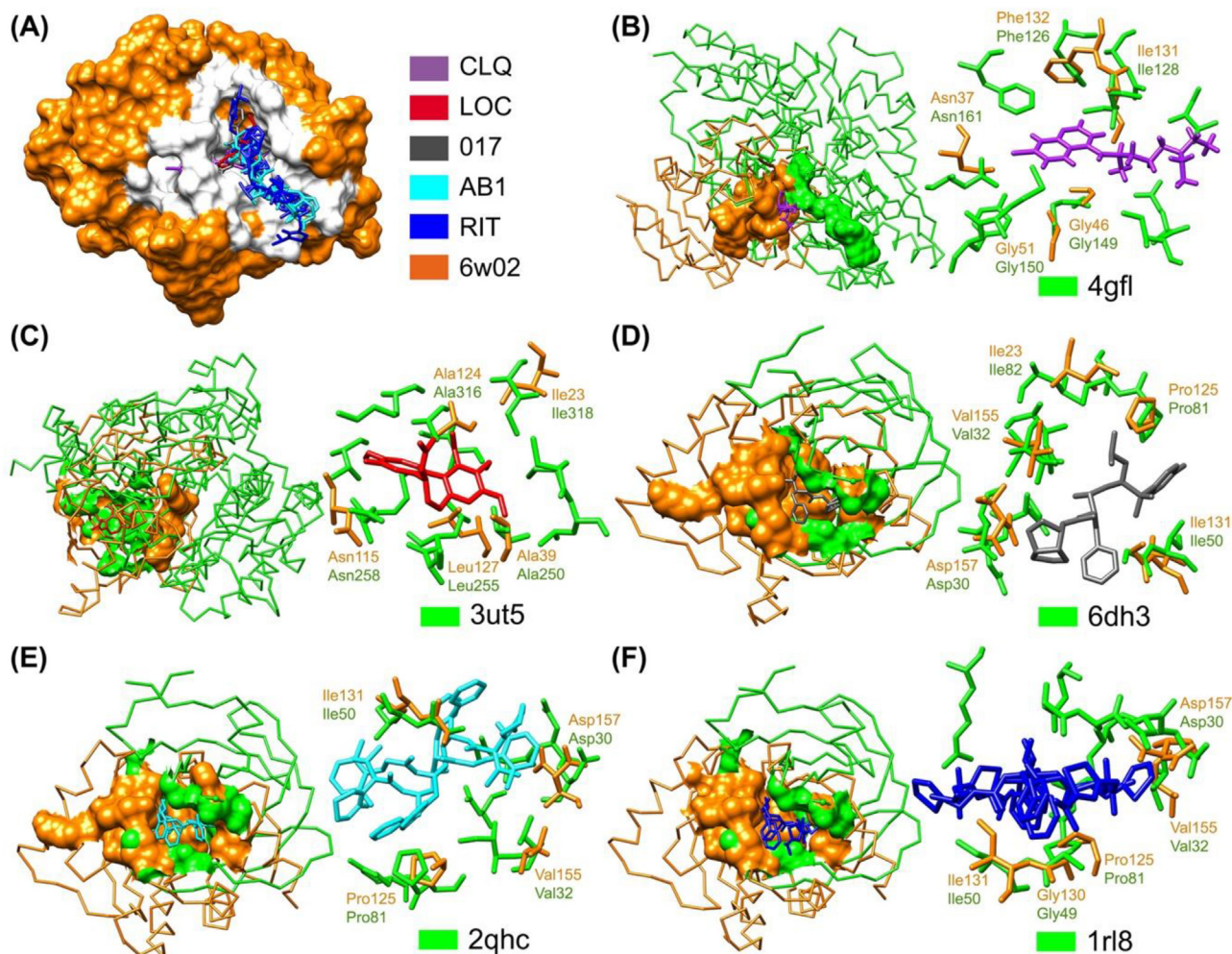
Fig. 2. Docked drug molecules on SARS-CoV-2 ADP ribose phosphatase (PDBID: 6w02). **(A)** Superposed drugs in ADP ribose phosphatase obtained from docking simulations in Autodock Vina. White shaded areas indicate that the residues are within 4.0 Å to the docked drug molecules. **(B-F, left)** Superpositions of known drug targets to ADP ribose phosphatase based on sub-structural similarity of drug binding sites. **(B-F, right)** Residues that are similarly arranged in ADP ribose phosphatase and binding sites for known drugs derived from protein-drug complexes. **(B)** Chloroquine (CLQ) binding site in quinone reductase (PDBID: 4fgl). **(C)** Colchicine (LOC) binding site in tubulin chain B (PDBID: 3ut5). **(D)** Darunavir (017) binding site in HIV protease (PDBID: 6dh3). (E) Lopinavir (AB1) binding site in HIV protease retropepsin (PDBID: 2qhc). **(F)** Ritonavir (RIT) binding site in HIV protease retropepsin (PDBID: 1rl8). The location of proposed binding sites are highlighted in surface representation colored in orange. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

same substructure searching methodology to identify off-target sites for the drugs being explored as COVID-19 treatments.

3.3.1. Human proteins bound to proposed drugs from PDB repository as potential off-targets

An ASSAM search of the human protein structures in the PDB using the drug binding sites we have identified was used as a means to investigate whether the use of these drugs could alter other pathways. The searches led us to a compilation of potential off-target sites and/or effects for eleven approved compounds (Table 2).

3.3.2. Human proteins with similar arrangements of amino acids to binding sites for proposed drugs as potential off-targets

The substructural similarity searches for potential off-target sites in human proteins using the Drug ReposER application was able to identify several proteins that have similar geometry to the binding site of a drug proposed for repositioning against SARS-CoV-2 targets (Section 3.3.2). The same data also allowed us to compile potential repurposing opportunities of these drugs for other indications including COVID-19 (Tables 2 and 3).

Non-homologous proteins that share similarly arranged sites for a particular drug molecule are more likely to be considered as off-targets because they may have different molecular function and are involved in distinct pathways that may not be associated with the original target disease. Recent computational studies have proposed several HIV protease inhibitors [18,20,21], NSAIDs [29], and losartan [41] as potential therapeutic agents for COVID-19. Although we can confirm the presence of potential binding sites to these drugs on SARS-CoV-2 proteins, we were also able to identify potential off-target sites where these drugs may alternatively bind in the structures of human proteins (Table 3).

Side effects on neurological systems have been common for approved drugs. Our study revealed potential neurological complications due to the usage of approved drugs such as HIV protease inhibitors [53], colchicine [54], naproxen [55] and losartan [56] (Table 3). Peripheral neuropathy due to the neurotoxicity of HIV protease inhibitors have been reported as complications resulting from anti-HIV treatment [53]. One potential off-target protein that might cause such symptoms is the HERC2 protein (PDBID: 3kci). Disruption of this protein causes reduction of E6AP activity that has been implicated in neurodevelopment disorders such as Angelman syndrome and autism [57].

Table 2

Human protein-drug complexes available in the PDB for the approved drugs proposed for COVID-19.

Bound drug (Drug ID[1])	PDB ID of protein-drug complex	Protein structure annotation	Known / potential effects from drug binding
CLQ	4v2o	Saposin B	Impaired lipid degradation [32]
	4fgl	Quinone reductase 2	Known treatment for malaria [33]
LOC	4lzr	Bromodomain-containing protein	Potential repurposing for cancer [34]
	5nkn	Neutrophil gelatinase-associated lipocalin	Potentially reduce poisoning effects of colchicine [35]
017	No drug-bound proteins from human is available.		
AB1			
RIT			
VIA	Conserved in phosphodiesterase family		Potential repurposing of sildenafil in multiple diseases related to PDE activities [36]
FOL	1drf	Dihydrofolate reductase	Known target for folic acid [37]
	4lrh	Folate receptor alpha	
	4kmz	Folate receptor beta	
LSN	5x24	Cytochrome P50 2C9	Known enzyme that binds to losartan [38]
NPS	3r58	Aldo-keto reductase family 1 member C3	Potential repurposing of naproxen for prostate cancer [38]
	4jq1	Aldo-keto reductase family 1 member C2	
AIN	No drug-bound proteins from human is available.		
IMN	2zb8	Prostaglandin reductase 2	Potential use of indomethacin to improve insulin sensitivity [39]
	3ads	PPAR-gamma	Potential repurposing of indomethacin for obesity and lipodystrophy [40]

Drug IDs are indicated as in Table 1.

Losartan targets the angiotensin type II receptor, however, it may also bind to the drug metabolizing cytochrome P450 (PDBID: 5×24) that has a similarly arranged site in ceruloplasmin (PDBID: 1kcw) (Table 3, Fig. 3B). Ceruloplasmin has been implicated with Parkinson's disease where disruption of the oxidative activity by ceruloplasmin causes increased iron levels in the brain that is correlated to Parkinson's [47,56]. On the other hand, it was also reported that losartan could be useful for Parkinson's where it might be able to reduce oxidative stress and neurodegeneration [58] thus warranting further investigations regarding the neuroprotective benefits of losartan.

The function inhibition of certain off-target proteins may provide coincidental antiviral effects (Table 3). Other than potentially targeting the SARS-CoV-2 proteins, NSAIDs such as naproxen (NPS), indomethacin (IMN) and aspirin (AIN) may also interact with host proteins involved in mounting the defense against viral infections. For example, we found that naproxen might be able to bind polypyrimidine tract-binding protein 1 (PTBP1) (PDBID: 1qm9) based on the similarity of the binding site for naproxen in serum albumin (PDBID: 4po0) (Table 4, Fig. 3C).

The PTBP1 protein had been shown to activate the replication of picornaviruses and coronaviruses through binding to its RNA binding domain [49,59], thus binding of naproxen to its binding site could potentially block viral replication. Other NSAIDs like the indomethacin and aspirin might also induce antiviral properties by binding to myeloperoxidase, which is a part of host defense system (Table 3). The protein acts as tissue damage factor that induces secondary bacterial lung infections causing the acute respiratory distress syndrome seen in influenza [51]. Decreased function of myeloperoxidase had been shown to potentially decrease inflammatory damage and lung viral load [51].

3.3.3. Human proteins with more than 30% sequence similarity to SARS-CoV-2 proteins as potential off-targets for the proposed COVID-19 drugs

SARS-CoV-2 proteins that may share a similar fold to human proteins were also considered as potential off-targets. In this case, SARS-CoV-2 proteins that retrieved a human protein by blastp alignment with more than 30% sequence identity is a possible indication of fold similarity. These protein structures were then analyzed to ascertain whether they contained a similar sub-structure

arrangement as the SARS-CoV-2 protein that is being targeted for drug repositioning (Table 4).

The SARS-CoV-2 ADP-ribose phosphatase from NSP3 has a similar sequence to human ADP-ribose binding protein and both share a similar mechanism of ADP ribose binding (Fig. 4). The SARS-CoV-2 spike protein is found have sequence similarities to the IRAP protein with both being associated to the renin-angiotensin pathway. No human sequences with possible fold similarities were detected for the main protease and non-structural proteins that include the NSP7, NSP8 and NSP10 which are conserved in viruses.

The binding of approved drugs or inhibitors to interleukin 17a and mineralocorticoid receptors that have similar sequences to the nucleocapsid and NSP15 respectively, could prevent inflammation by the immune response [65,66], a known complication of COVID-19 (Table 3). This would mean that such drugs could target both the virus and the host in parallel with potentially therapeutic results.

3.4. Other compounds with potential as COVID-19 therapeutics based on ligand structure similarity

It is known that similar drugs may require a similar binding environmentand can have similar inhibitory effects, thus making it possible that a target protein can interact with a set of drug molecules with similar structures [67]. With this premise, the structures of the drugs proposed for repositioning against SARS-CoV-2 targets (proposed drugs) were used as a reference point to find other drug molecules with similar structures (matched drugs). This was carried out using the ligand search interface in the PDB that is based on the comparison of pharmacophores. The search identified 6 matches with similar structures to the input queries – quinacrine, vardenafil, lenalidomide, pomalidomide, amprenavir and methotrexate (Table 5). With the exceptions of methotrexate, which has structural similarities to folic acid (ClinicalTrials.gov IDs: NCT04352465 and NCT04434118), and lenalidomide, which is related to thalidomide (ClinicalTrials.gov ID: NCT04361643), none of these compounds are involved in any known clinical trials for COVID-19 at the time of writing. Molecular docking targeting the SARS-CoV-2 proteins using both the proposed and matched drugs (shared SARS-CoV-2 protein targets) resulted in several binding poses that is indicative that the matched drugs can poten-

Table 3
Potential off-target effects for selected drugs according to sub-structural similarity of binding patterns from Drug ReposER application.

Drug ID[1]	Query PDBID of structure with known binding site	Hit PDB of structure with a potential alternate / off-target site (Docking score)	Macromolecule and its associated pathways or mechanisms for off-target sites / involvement in antiviral activity	Potential/reported outcomes associated with the off-target site
Potential off-targets that may cause adverse effects				
LOC	5nm5B	2gk1I (−7.5)	Beta-hexosaminidase subunit alpha (HEXA) - Tay Sachs disease (TSD) [42]	Neurodegeneration
017	6dh0A	3kciA (−9.3)	E3 ubiquitin-protein ligase (HERC2) - neurodevelopmental disorder [42]	Neurological complications
017	3oxwB	2r7eB (−7.6)	Coagulation factor VIII (C8) - hemophilia [43]	Increased risk of hemorrhage in hemophilia patients
RIT	5veuA	3d7uA (−8.5)	Tyrosine protein kinase (CSK) - suppress SRC tyrosine kinase (SFK) activity that cause cancer such as colorectal cancer [44]	Increased risk of colorectal cancer
RIT	1rl8A	3hhdA (−13.1)	Fatty acid synthase (FAS/FASN) - lipid mechanism [45]	Lipodystrophy
NPS	2vdbA	3fgqA (−6.1)	Neuroserpin (SERPIN1) - stroke [46]	Increased risk of stroke
LSN	5x24A	1kcwA (−7.4)	Ceruloplasmin (CP) - Parkinsonism [47]	Worsens the effects of parkinsonism
LSN	5x24A	3gzdA (−6.7)	Selenocysteine lyase (SCLY) - glucose and lipid metabolism [48]	Prevents metabolic syndromes
Potential off-targets with reported antiviral properties				
NPS	4po0A	1qm9A (−6.9)	Polypyrimidine tract-binding protein 1 (PTBP1) - allow viral replication through its RNA binding domain [49]	Potentially inhibit viral replication
AIN	2qqtA	1d7wC (−6.1)	Myloperoxidase (MPO) - antiviral properties toward influenza virus [50], induce secondary bacterial lung infection [51]	Potentially decrease lung viral loads
IMN	3ogwA	6azpA (−8.2)		
RIT	1rl8A	3hhdA (−13.1)	Fatty acid synthetase (FAS/FASN) - facilitate viral replication through generation of membrane compartments [52]	Potentially inhibit viral replication

[1] Drug IDs are indicated as in Table 1.
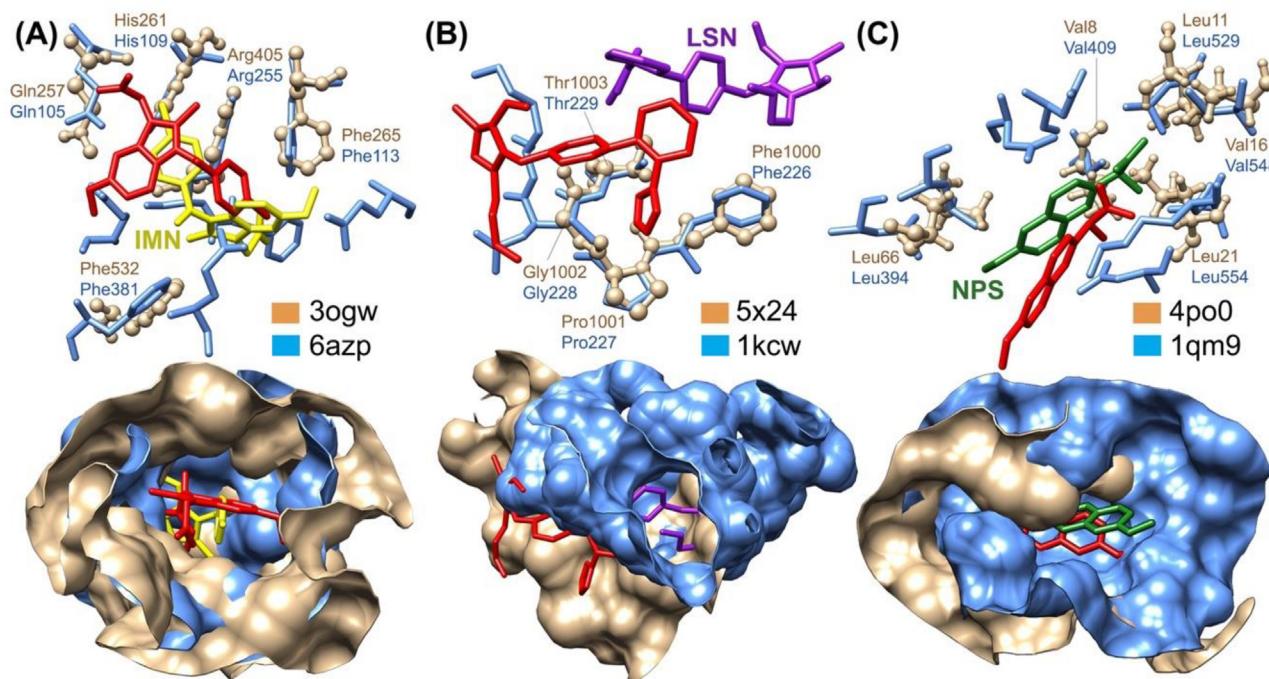
Fig. 3. Examples of potential off-target sites for selected drug molecules. Residues in cornflower blue are within 4.0 Å to a drug molecule derived from protein-drug complexes in the PDB; tan residues are potential protein targets that are similarly arranged to binding residues from protein-drug complexes. All docked ligands are colored in red. **(A)** The indomethacin binding site in lactoperoxidase that are similarly arranged to residues in myeloperoxidase. **(B)** Superposition of the losartan binding site in cytochrome 450 docked onto the binding pose in ceruloplasmin **(C)** Similarly arranged residue patterns between the binding site of naproxen in serum albumin and residues within 4.0 Å to the docked naproxen in polypyrimidine tract-binding protein 1. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

tially bind to SARS-CoV-2 proteins in a similar manner as the proposed drugs (Table 5, Fig. 5).

Our analyses found that both darunavir and amprenavir can potentially bind to the same SARS-CoV-2 site (P125, G130, I131, V155, and D157) in NSP3 (Fig. 4E, Table 5). Darunavir, when docked on NSP3, has a molecular binding affinity of −9.4 kcal/mol. Amprenavir, when docked at the similar site (Fig. 4E), also has a molecular binding affinity of −9.4 kcal/mol (Table 5).

Table 4

Human proteins with more than 30% sequence identity to SARS-CoV-2 proteins retrieved by a blastp search of the PDB database.

SARS-CoV-2 protein	PDBID of structure homolog	Protein function / disease mechanism Potential side effects / benefits
NSP3 / ADP ribose phosphatase	3q6zA (31.25%)	Poly [ADP-ribose] polymerase 14, catalyze the mono-ADP-ribosylation of STAT1, functions in innate immune response [60]
Spike protein	4z7iA (48.51%)	Insulin-regulated amino peptidase – binds angiotensin IV in the brain [61]
	4bkfC (72.00%)	EPHRIN-B3 – serves as receptor for Nipah virus [62]
	5ojmA (94.44%)	Gamma-aminobutyric acid receptor subunit alpha-5 – implicated in neurological disorders [63]
NSP15	4ewqA (41.67%)	Mitogen-activated protein kinase 14 – plays a role in neuroinflammatory responses [64]
	4tntA (94.44%)	Mineralocorticoid receptor - plays a role in inflammatory responses through regulation of macrophage and T-cells, and is implicated in cardiac hypertrophy [65]
Nucleocapsid	2vxsA (37.84%)	Interleukin-17a – involves in inflammatory responses and plays a role in cardiovascular complications [66]
Main protease NSP7 NSP8 NSP10 NSP16	No human homolog found, conserved in viruses.	

[1]Drug IDs are indicated as in Table 1.

Sequence similarity percentages are provided in brackets.

Some structurally similar drug molecules are intended for similar indications. Both darunavir and amprenavir (Fig. 4E) are protease inhibitors that have been used for the treatment of HIV. However, amprenavir is useful against infections that exhibit resistance to other protease inhibitors used in HIV treatment [68]. Thus, amprenavir might confer an advantage in a scenario where the protein target from SARS-CoV-2 develops resistance towards darunavir. Both chloroquine and quinacrine (Fig. 4A) have been indicated for the treatment of systemic lupus erythematosus as well as other diseases [69].

A study comparing the oculotoxicity of chloroquine and quinacrine in the management of lupus erythematosus found that quinacrine exhibits less oculotoxicity compared to chloroquine if taken at low doses [70]. Thus, quinacrine might be a less toxic alternative compared to chloroquine with regard to any ophthalmologic side effects.

Sildenafil and vardenafil (Fig. 4D) have been used in the treatment of erectile dysfunction [36,71]. Due to vardenafil's weaker inhibition of PDE6 compared to sildenafil, the use of vardenafil is less likely to cause abnormal color perception unlike sildenafil [72]. In cases where the patients are afflicted by this sildenafil side effect, switching to vardenafil might still provide the desired therapeutic outcomes. Thalidomide, lenalidomide, and pomalidomide (Fig. 4B-C) have been used in the treatment for multiple myeloma [73,74]. Both lenalidomide and pomalidomide are shown to be more potent compared to thalidomide with pomalidomide exhibiting the highest potency among the three [73,74]. Therefore, lenalidomide and pomalidomide are good alternatives for thalidomide due to their higher potency.

There are also structurally similar drug molecules that are utilized for different indications. For example, folic acid has been indicated for folic acid deficiency [37] while methotrexate has been indicated for rheumatoid arthritis [75] and certain forms of cancer [76]. This structural similarity makes methotrexate a folate analog (anti-folate) that is able to antagonize the biological action of folic acid [37]. Due to the severe side effects that are associated with methotrexate, it should only be indicated in scenarios where the primary drug for a particular treatment has failed to alleviate the patient's condition [77].

Vardenafil, amprenavir and methotrexate had been reported to potentially bind to SARS-CoV-2 proteins through structural analyses [19,78]. To our knowledge, the potential use of quinacrine, lenalidomide, and pomalidomide for COVID-19 have not been
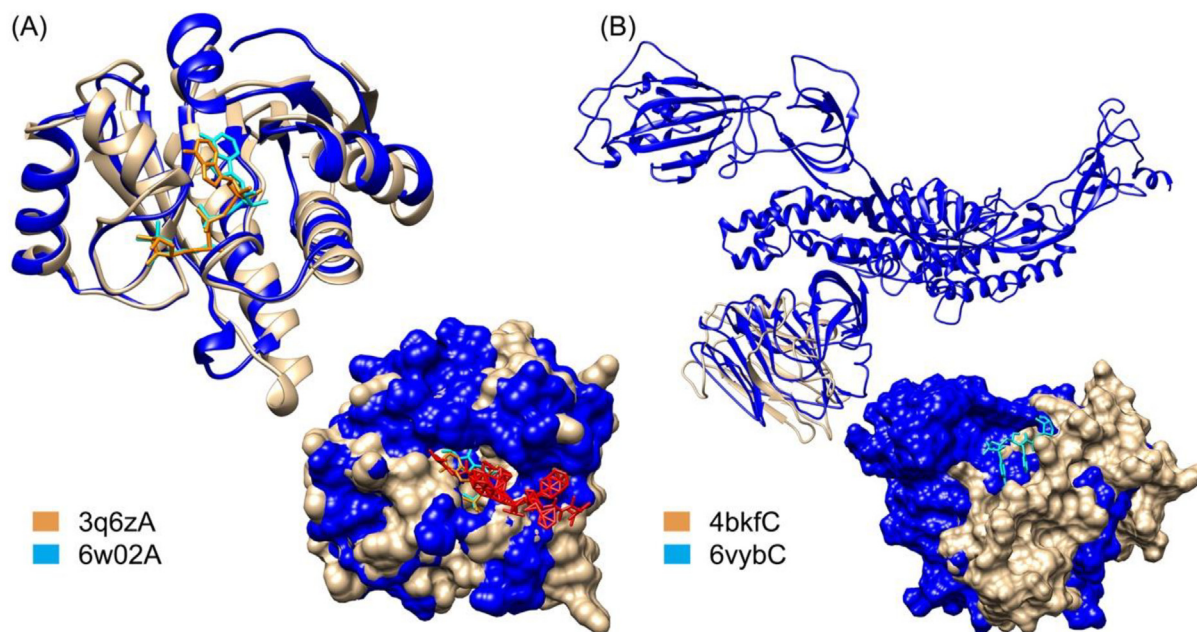
Fig. 4. Examples of potential off-target sites for selected drug molecules based on a blastp search for human proteins with more than 30% sequence identity to SARS-CoV-2 proteins. **(A)** A superposition of structures between the Poly [ADP-ribose] polymerase 14 from human (tan color) and ADP ribose in NSP3 of SARS-CoV-2 (blue color) resulting in a shared binding site for substrate (APR) and proposed site for binding to ritonavir (red). **(B)** Superposition of human EPHRIN-B3 (tan color) and spike protein (blue color) with the proposed site for darunavir (cyan). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 5

The drug molecules in Drug ReposER that exhibit similar structure with the proposed drug molecules for COVID-19 clinical trials that have the similar potential ability of binding to SARS-CoV-2 proteins.

Drug molecules proposed for COVID-19 clinical trials (proposed drugs)	Drug molecules exhibiting similar structure with drug molecules undergoing trials (matched drugs)	Shared SARS-CoV-2 protein targets predicted by Drug ReposER	Similar binding sites on the target structures from SARS-CoV-2 identified by Drug ReposER	Molecular docking analysis		
				Presence of binding conformation that is close to the predicted binding site (<4 Å)		Binding affinity (kcal/mol)
Chloroquine (CLQ)	Quinacrine (QUN)	Angiotensin-converting enzyme 2 (ACE2) (6m17)	–	CLQ	None	−6.2
				QUN	None	−6.5
Sildenafil (VIA)	Vardenafil (VDN)	Non-structural protein 16 (NSP16) (6w75)	MET A 6929 ILE A 6951 TYR A 6979 ALA A 6990	VIA	Yes	−7.9
			HIS A 7023	VDN	Yes	−7.8
Thalidomide (EF2)	Lenalidomide (LVY)	ACE2 and Receptor Binding Domain (RBD) (6m17)	TYR B 41 ASN E 439 PHE E 497 PRO E 507	EF2	None	−7.6
				LVY	None	−7.4
	Pomalidomide (Y70)	NSP16 (6w4h)	HIS A 6867 THR A 6891 TRP A 6922	EF2	None	−7.7
			PHE A 6954	Y70	None	−7.6
Darunavir (017)	Amprenavir (478)	NSP3 (6w02)	PRO B 125 GLY B 130 ILE B 131 VAL B 155	17	Yes	−9.4
			ASP B 157	478	Yes	−9.4
Folic Acid (FOL)	Methotrexate (MTX)	Main protease (M^{pro}) (7buy)	THR A 199 LEU A 205 VAL A 233 SER A 267	FOL	Yes	−7.5
			LEU A 271	MTX	Yes	−7.3

reported elsewhere in the context of binding ability through structural analyses. Furthermore, the finding that quinacrine is a readily available compound that has yet to be explored or proposed for COVID-19 is novel to this work. Should the current candidate drug molecules proposed for COVID-19 clinical trials fail at any stage of the process, these structurally similar drug molecules can be investigated as potential alternatives. It is not unexpected that the use of these structurally similar compounds could be used in concert as a cocktail for more effective therapy [79].

*3.5. Distinction from other COVID-19 drug repurposing efforts and future directions*

In this work, all drugs that have been proposed for clinical trials were analyzed using the Drug ReposER pipeline to find their potential binding sites in any SARS-CoV-2 protein by virtue of having similar 3D arrangements of amino acid residues to the known target sites. It is not unexpected that our results will overlap or have parallels with the outcomes of other studies that have been recently published or are ongoing. However, the results presented here and in the COVID-19 Drug ReposER resource, will also provide the relevant supporting insights regarding why or how a particular drug may be effective while at the same time, have the added advantage of presenting the potential capacity for off-target interactions that may cause or explain any side effects upon administration.

The use of computational substructure comparisons to identify alternative sites for the repositioning of approved drug compounds is different from other studies that report drug repurposing efforts for COVID-19 such as those by Zhou *et al.* [5] and Gordon *et al.* [6] that employed protein network analyses. Zhou *et al.* compared the

network of interaction between SARS-CoV-2 and the human proteins drug-target network in the human interactome in order to search for common protein-protein interactions and functional pathways and from there predict existing drugs involved in such pathways [5]. Compared to our findings, the study proposed 16 existing drugs to be repurposed as anti-HCoV (human coronavirus) where two of the 16 drugs matched our set of proposed drugs. Gordon *et al.* had identified 29 approved drugs bound to 66 druggable human proteins based on the analysis of association networks between human and SARS-CoV-2 proteins [6]. In comparison to our analyses, there are two drugs that overlap with our results, chloroquine (targeting sigma1-receptor:NSP6) and indomethacin (targeting PTGES2:NSP7).

This study was intended to develop a pipeline to identify drug compounds that could be repositioned against SARS-COV-2 targets using the available structural information in the PDB. This pipeline was also able to identify potential side effects or toxicity associated with those compounds that arose from off-target binding. Integrating the data to pharmacophore matching tools allowed other similarly structured drug compounds to be identified that also had the potential to be repositioned against SARS-CoV-2 targets. The information derived from such analyses could be used as a means of decision making to prioritize down-stream experimental validation and assays. This study does not provide any experimental evidence validating the binding of the proposed repositioned drugs to SARS-CoV-2 proteins. The results of this study should not be regarded as an explicit treatment recommendation or protocol for COVID-19.

A limited set of existing drugs extracted from lists of those currently undergoing or planned for COVID-19 trials was used in this work. The analyses reported only utilized data of compounds that
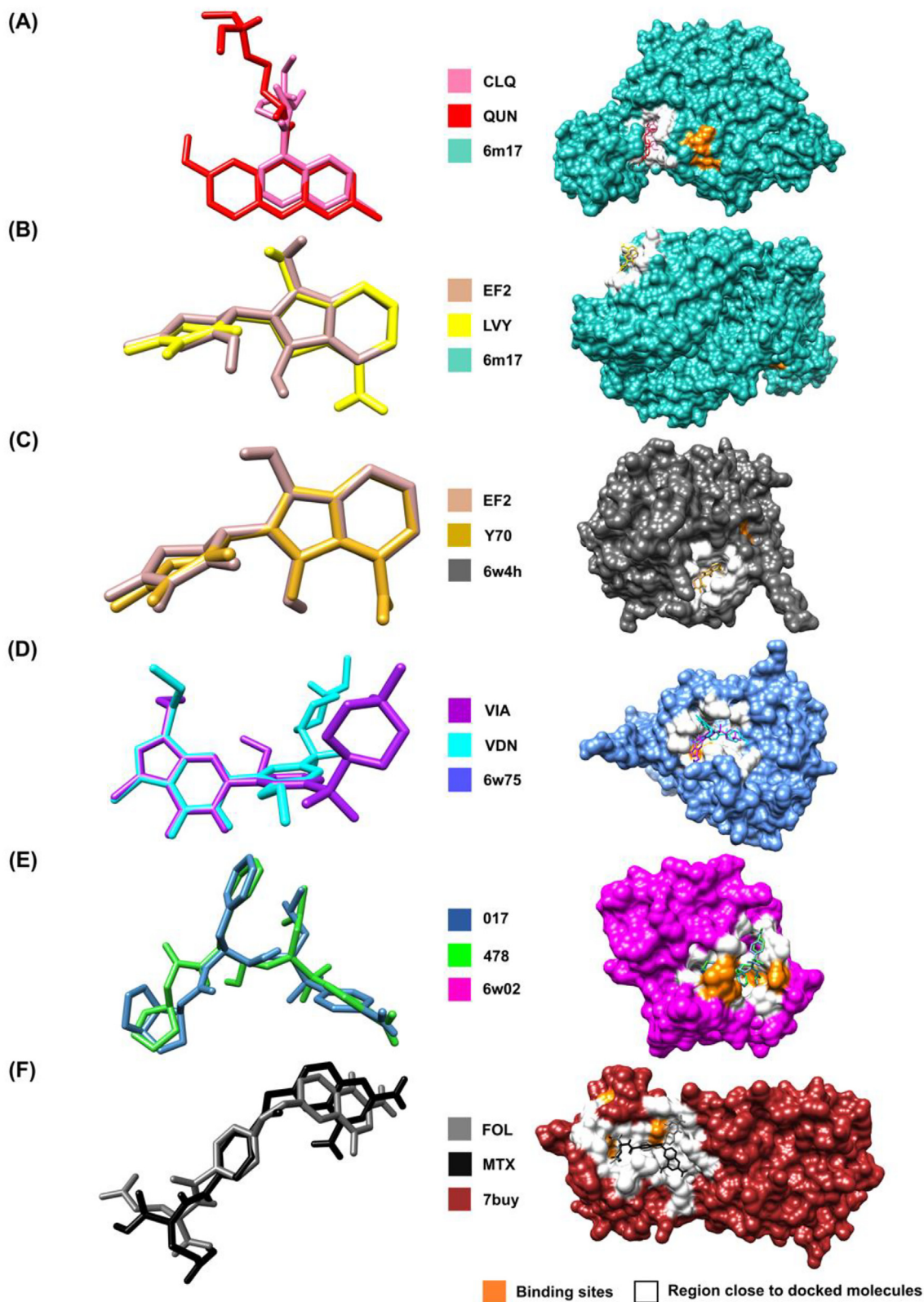
Fig. 5. Quinacrine (QUN), vardenafil (VDN), lenalidomide (LVY), pomalidomide (Y70), amprenavir (478), and methotrexate (MTX) share structural similarities with the corresponding drug molecules that have been proposed for COVID-19 clinical trials: chloroquine (CLQ), sildenafil (VIA), thalidomide (EF2), darunavir (017) and folic acid (FOL) respectively. Structural alignment of QUN (red) **(A)**, LVY (yellow) **(B)**, Y70 (gold) **(C)**, VDN (cyan) **(D)**, 478 (green) **(E)**, and MTX (black) **(F)** with CLQ (pink) **(A)**, EF2 (brown) **(B, C)**, VIA (purple) **(D)**, 017 (blue) **(E)**, and FOL (gray) **(F)** respectively. Molecular docking for the structurally similar drug molecules and the corresponding drug molecules that have been proposed for COVID-19 clinical trials on their shared protein target from SARS-CoV-2 predicted by Drug ReposER (ACE2 and RBD complex (PDBID: 6 m17, green) **(A, B)**, NSP16 (PDBID: 6w4h, gray) **(C)**, NSP16 (PDBID: 6w75, blue) **(D)**, NSP3 (PDBID: 6w02, magenta) **(E)**, and M^pro (PDBID: 7buy, brown) **(F)**). The white shaded areas indicate regions containing residues within less than 4.0 Å to docked drug molecules. The orange shaded areas indicate regions containing residues that form the binding sites identified by Drug ReposER. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

were structurally present as a standalone ligand in the PDB. Both these factors restricted the number of potential candidates that could be proposed for repurposing. Despite these limitations, our analyses yielded 22 target sites for repurposing of which only 6 had been mentioned in other studies. It is clear that the work reported here could be extended to include all known drug binding sites in the PDB. Although the current targets for repositioning in this study considers only SARS-CoV-2 proteins, the pipeline can be integrated to network analyses methods to identify human proteins that could also yield therapeutic effects for COVID-19. Furthermore, this study can also be extended to include other SARS-CoV-2 structures as and when they become available. Such data will be updated via the specific Drug ReposER resource for COVID-19.

4. Conclusions

The fastest and safest route to providing drug treatments for COVID-19 would be to reposition approved compounds against targets from this newly described disease. At the time of writing, the search for effective COVID-19 treatments is still ongoing. Despite being subject to the availability of associated protein coordinate structure data in the PDB, the use of amino acid 3D side chain based sub-structure comparisons have proven to be a feasible means of identifying candidate compounds to be repositioned for COVID-19. Our analyses yielded 22 potential sites in SARS-CoV-2 proteins and 16 drug compounds that could be repurposed for COVID-19. It is clear that the use of structural data from the PDB is able to provide high quality mechanistic level details for strategizing the selection of candidate compounds to be repurposed. The capacity to not only identify new target sites, but also identify potential off-target sites, provide a deeper level of context for the decision making process to safely proceed with exploring specific compounds to be repurposed for the new disease.

Declaration of Competing Interest

All authors have participated in (a) conception and design, or analysis and interpretation of the data; (b) drafting the article or revising it critically for important intellectual content; and (c) approval of the final version.

This manuscript has not been submitted to, nor is under review at, another journal or other publishing venue.

The authors have no affiliation with any organization with a direct or indirect financial interest in the subject matter discussed in the manuscript.

Acknowledgments

References

[1] Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020;395:497–506. https://doi.org/10.1016/S0140-6736(20)30183-5.

[2] Gorbalenya AE, Baker SC, Baric RS, de Groot RJ, Drosten C, Gulyaeva AA, et al. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nat Microbiol 2020;5:536–44. https://doi.org/10.1038/s41564-020-0695-z.

[3] Abd El-Aziz TM, Stockand JD. Recent progress and challenges in drug development against COVID-19 coronavirus (SARS-CoV-2) - an update on the

status. Infect Genet Evol 2020;83:104327. https://doi.org/10.1016/j.meegid.2020.104327.

[4] Burley SK, Berman HM, Bhikadiya C, Bi C, Chen L, Di Costanzo L, et al. Protein Data Bank: the single global archive for 3D macromolecular structure data. Nucl Acids Res 2019;47:D520–8. https://doi.org/10.1093/nar/gky949.

[5] Zhou Y, Hou Y, Shen J, Huang Y, Martin W, Cheng F. Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. Cell Discov 2020;6. https://doi.org/10.1038/s41421-020-0153-3.

[6] Gordon DE, Jang GM, Bouhaddou M, Xu J, Obernier K, O'Meara MJ, et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature 2020.

[7] Nadzirin N, Gardiner EJ, Willett P, Artymiuk PJ, Firdaus-Raih M. SPRITE and ASSAM: web servers for side chain 3D-motif searching in protein structures. Nucl Acids Res 2012;40:W380–6. https://doi.org/10.1093/nar/gks401.

[8] Nadzirin N, Willett P, Artymiuk PJ, Firdaus-Raih M. IMAAAGINE: a webserver for searching hypothetical 3D amino acid side chain arrangements in the Protein Data Bank. Nucl Acids Res 2013;41:432–40. https://doi.org/10.1093/nar/gkt431.

[9] Ab Ghani NS, Ramlan EI, Firdaus-Raih M. Drug ReposER: a web server for predicting similar amino acid arrangements to known drug binding interfaces for potential drug repositioning. Nucl Acids Res 2019;47:W350–6. https://doi.org/10.1093/nar/gkz391.

[10] Parisi D, Adasme MF, Sveshnikova A, Bolz SN, Moreau Y, Schroeder M. Drug repositioning or target repositioning: a structural perspective of drug-target-indication relationship for available repurposed drugs. Comput Struct Biotechnol J 2020;18:1043–55. https://doi.org/10.1016/j.csbj.2020.04.004.

[11] Haupt VJ, Daminelli S, Schroeder M. Drug promiscuity in PDB: protein binding site similarity is key. PLoS One 2013:8. https://doi.org/10.1371/journal.pone.0065894.

[12] Huang Y, Niu B, Gao Y, Fu L, Li W. CD-HIT Suite: a web server for clustering and comparing biological sequences. Bioinformatics 2010;26:680–2. https://doi.org/10.1093/bioinformatics/btq003.

[13] Nadzirin N, Gardiner EJ, Willett P, Artymiuk PJ, Firdaus-raih M. SPRITE and ASSAM: web servers for side chain 3D-motif searching in protein structures. Nucl Acids Res 2012;40:380–6. https://doi.org/10.1093/nar/gks401.

[14] Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 2010;31:455–61. https://doi.org/10.1002/jcc.

[15] Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera–a visualization system for exploratory research and analysis. J Comput Chem 2004;25:1605–12. https://doi.org/10.1002/jcc.20084.

[16] Boutet E, Lieberherr D, Tognolli M, Schneider M, Bairoch A. UniProtKB/Swiss-Prot. Methods Mol Biol 2007;406:89–112.

[17] Zarin DA, Tse T, Williams RJ, Carr S. Trial Reporting in ClinicalTrials.gov — the final rule. N Engl J Med 2016;375:1998–2004. https://doi.org/10.1056/NEJMsr1611785.

[18] Saadat S, Mansoor S, Naqvi N, Fahim A, Rehman Z, Khan SY, et al. Structure based drug discovery by virtual screening of 3699 compounds against the crystal structures of six key SARS-CoV-2 proteins 2020. DOI:10.21203/rs.3.rs-28113/v1.

[19] Qiao Z, Zhang H, Ji H-F, Chen Q. Computational view toward the inhibition of SARS-CoV-2 spike glycoprotein and the 3CL protease. Computation 2020;8:53. https://doi.org/10.3390/computation8020053.

[20] Shankar U, Jain N, Majee P, Mishra SK, Rathi B, Kumar A. Potential Drugs Targeting Nsp16 Protein May Corroborates a Promising Approach to Combat SARSCoV-2 Virus 2020. DOI:10.26434/CHEMRXIV.12279671.V1.

[21] Parida PK, Paul D, Chakravorty D. Nature to Nurture- Identifying Phytochemicals from Indian Medicinal Plants as Prophylactic Medicine by Rational Screening to Be Potent Against Multiple Drug Targets of SARS-CoV-2 2020. DOI:10.26434/CHEMRXIV.12355937.V1.

[22] Li G, De Clercq E. Therapeutic options for the 2019 novel coronavirus (2019-nCoV). Nat Rev Drug Discov 2020;19:149–50. https://doi.org/10.1038/d41573-020-00016-0.

[23] Yamamoto N, Yang R, Yoshinaka Y, Amari S, Nakano T, Cinatl J, et al. HIV protease inhibitor nelfinavir inhibits replication of SARS-associated coronavirus. Biochem Biophys Res Commun 2004;318:719–25. https://doi.org/10.1016/j.bbrc.2004.04.083.

[24] Zumla A, Chan JFW, Azhar EI, Hui DSC, Yuen K-Y. Coronaviruses — drug discovery and therapeutic options. Nat Rev Drug Discov 2016;15:327–47. https://doi.org/10.1038/nrd.2015.37.

[25] Bouvet M, Lugari A, Posthuma CC, Zevenhoven JC, Bernard S, Betzi S, et al. Coronavirus Nsp10, a critical co-factor for activation of multiple replicative enzymes. J Biol Chem 2014;289:25783–96. https://doi.org/10.1074/jbc.M114.577353.

[26] Kirchdoerfer RN, Ward AB. Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. Nat Commun 2019;10. https://doi.org/10.1038/s41467-019-10280-3.

[27] Marinella MA. Indomethacin and resveratrol as potential treatment adjuncts for SARS-CoV-2/COVID-19. Int J Clin Pract 2020;74. https://doi.org/10.1111/ijcp.v74.910.1111/ijcp.13535.

[28] Zhu H, Cong J-P, Yu D, Bresnahan WA, Shenk TE. Inhibition of cyclooxygenase 2 blocks human cytomegalovirus replication. PNAS 2002;99:3932–7. https://doi.org/10.1073/pnas.052713799.

[29] Milad L, Reza V, Majid S, Fatemeh R, Akbar H-O, Massoud A, et al. Repurposing naproxen as a potential antiviral agent against SARS-CoV-2 2020. DOI:10.21203/RS.3.RS-21833/V1.

[30] Fehr AR, Channappanavar R, Jankevicius G, Fett C, Zhao J, Athmer J, et al. The conserved coronavirus macrodomain promotes virulence and suppresses the innate immune response during severe acute respiratory syndrome coronavirus infection. mBio 2016;7. https://doi.org/10.1128/mBio.01721-16.

[31] Chartier M, Morency L-P, Zylber MI, Najmanovich RJ. Large-scale detection of drug off-targets: hypotheses for drug repurposing and understanding side-effects. BMC Pharmacol Toxicol 2017;18. https://doi.org/10.1186/s40360-017-0128-7.

[32] Huta BP, Mehlenbacher MR, Nie Y, Lai X, Zubieta C, Bou-Abdallah F, et al. The lysosomal protein saposin B binds chloroquine. ChemMedChem 2016;11:277–82. https://doi.org/10.1002/cmdc.201500494.

[33] Becker K, Tilley L, Vennerstrom JL, Roberts D, Rogerson S, Ginsburg H. Oxidative stress in malaria parasite-infected erythrocytes: host–parasite interactions. Int J Parasitol 2004;34:163–89. https://doi.org/10.1016/j.ijpara.2003.09.011.

[34] Lucas X, Wohlwend D, Hügle M, Schmidtkunz K, Gerhardt S, Schüle R, et al. 4-Acyl pyrroles: mimicking acetylated lysines in histone code reading. Angew Chem Int Ed 2013;52:14055–9. https://doi.org/10.1002/anie.201307652.

[35] Barkovskiy M, Ilyukhina E, Dauner M, Eichinger A, Skerra A. An engineered lipocalin that tightly complexes the plant poison colchicine for use as antidote and in bioanalytical applications. Biol Chem 2019;400:351–66. DOI:10.1515/hsz-2018-0342.

[36] Andersson K-E. PDE5 inhibitors - pharmacology and clinical applications 20 years after sildenafil discovery: PDE5 inhibitors. Br J Pharmacol 2018;175:2554–65. https://doi.org/10.1111/bph.14205.

[37] Kamen B. Folate and antifolate pharmacology. Semin Oncol 1997;24:S18-30-S18-39.

[38] Zhou S-F, Zhou Z-W, Yang L-P, Cai J-P. Substrates, inducers, inhibitors and structure-activity relationships of human cytochrome P450 2C9 and Implications in Drug Development. Curr Med Chem 2009;16:3480–675. https://doi.org/10.2174/092986709789057635.

[39] Wu Y-H, Ko T-P, Guo R-T, Hu S-M, Chuang L-M, Wang A-J. Structural basis for catalytic and inhibitory mechanisms of human prostaglandin reductase PTGR2. Structure 2008;16:1714–23. https://doi.org/10.1016/j.str.2008.09.007.

[40] Puhl AC, Milton FA, Cvoro A, Sieglaff DH, Campos JCL, Bernardes A, et al. Mechanisms of peroxisome proliferator activated receptor γ regulation by non-steroidal anti-inflammatory drugs. Nucl Receptor Signaling 2015;13. https://doi.org/10.1621/nrs.13004. e004.

[41] Jazzi AS, Mahnam K, Hejazi SH, Damavandi MS, Sadeghi P, Zeinalian M, et al. Inhibition of Viral Macrodomain of COVID-19 and Human TRPM2 by losartan 2020. DOI:10.20944/PREPRINTS202003.0457.V1.

[42] Dersh D, Iwamoto Y, Argon Y, Gilmore R. Tay–Sachs disease mutations in HEXA target the α chain of hexosaminidase A to endoplasmic reticulum–associated degradation. Mol Biol Cell 2016;27:3813–27. https://doi.org/10.1091/mbc.E16-01-0012.

[43] Mannully ST, Ramya LN, Pulicherla KK. Perspectives on progressive strategies and recent trends in the production of recombinant human factor VIII. Int J Biol Macromol 2018;119:496–504. https://doi.org/10.1016/j.ijbiomac.2018.07.164.

[44] Kunte DP, Wali RK, Koetsier JL, Hart J, Kostjukova MN, Kilimnik AY, et al. Down-regulation of the tumor suppressor gene C-terminal Src kinase: An early event during premalignant colonic epithelial hyperproliferation. FEBS Lett 2005;579:3497–502. DOI:10.1016/j.febslet.2005.05.030.

[45] Gibellini L, De Biasi S, Nasi M, Carnevale G, Pisciotta A, Bianchini E, et al. Different origin of adipogenic stem cells influences the response to antiretroviral drugs. Exp Cell Res 2015;337:160–9. https://doi.org/10.1016/j.yexcr.2015.07.031.

[46] Gelderblom M, Neumann M, Ludewig P, Bernreuther C, Krasemann S, Arunachalam P, et al. Deficiency in Serine Protease Inhibitor Neuroserpin Exacerbates Ischemic Brain Injury by Increased Postischemic Inflammation. PLoS One 2013;8. DOI:10.1371/journal.pone.0063118.

[47] Kristinsson J, Snaedal J, Tórsdóttir G, Jóhannesson T. Ceruloplasmin and iron in Alzheimer's disease and Parkinson's disease: a synopsis of recent studies. Neuropsychiatr Dis Treat 2012;8:515–21. https://doi.org/10.2147/NDT.S34729.

[48] Seale LA, Hashimoto AC, Kurokawa S, Gilman CL, Seyedali A, Bellinger FP, Raman AV, Berry MJ. Disruption of the selenocysteine lyase-mediated selenium recycling pathway leads to metabolic syndrome in mice. Mol Cell Biol 2012;32:4141–54. https://doi.org/10.1128/mcb.00293-12.

[49] Kafasla P, Lin H, Curry S, Jackson RJ. Activation of picornaviral IRESs by PTB shows differential dependence on each PTB RNA-binding domain. RNA 2011;17:1120–31. https://doi.org/10.1261/rna.2549411.

[50] Sugamata R, Dobashi H, Nagao T, Yamamoto K ichi, Nakajima N, Sato Y, et al. Contribution of neutrophil-derived myeloperoxidase in the early phase of fulminant acute respiratory distress syndrome induced by influenza virus infection. Microbiol Immunol 2012;56:171–82. DOI:10.1111/j.1348-0421.2011.00424.x.

[51] Ishikawa H, Fukui T, Ino S, Sasaki H, Awano N, Kohda C, Tanaka K. Influenza virus infection causes neutrophil dysfunction through reduced G-CSF production and an increased risk of secondary bacteria infection in the lung. Virology 2016;499:23–9. https://doi.org/10.1016/j.virol.2016.08.025.

[52] Pombo JP, Sanyal S. Perturbation of intracellular cholesterol and fatty acid homeostasis during flavivirus infections. Front Immunol 2018;9:1276. https://doi.org/10.3389/fimmu.2018.01276.

[53] Kranick SM, Nath A. Neurologic complications of HIV-1 infection and its treatment in the Era of antiretroviral therapy. ContinLifelong Learn Neurol 2012;18:1319–37. https://doi.org/10.1212/01.CON.0000423849.24900.ec.

[54] Sil S, Ghosh R, Sanyal M, Guha D, Ghosh T. A comparison of neurodegeneration linked with neuroinflammation in different brain areas of rats after intracerebroventricular colchicine injection. J Immunotoxicol 2016;13:181–90. https://doi.org/10.3109/1547691X.2015.1030804.

[55] Park K, Bavry AA. Risk of stroke associated with nonsteroidal anti-inflammatory drugs. Vasc Health Risk Manag 2014;10:25–32. https://doi.org/10.2147/VHRM.S54159.

[56] Sarma GRK, Kamath V, Mathew T, Roy AK. A case of parkinsonism worsened by losartan: a probable new adverse effect: Letters to the Editor. Mov Disord 2008;23:1055. https://doi.org/10.1002/mds.21945.

[57] Abraham JR, Barnard J, Wang H, Noritz GH, Yeganeh M, Buhas D, Natowicz MR. Proteomic investigations of human HERC2 mutants: Insights into the pathobiology of a neurodevelopmental disorder. Biochem Biophys Res Commun 2019;512:421–7. https://doi.org/10.1016/j.bbrc.2019.02.149.

[58] Wright JW, Harding JW. Importance of the brain angiotensin system in Parkinson's disease. Parkinson's Dis 2012;2012:1–14. https://doi.org/10.1155/2012/860923.

[59] Sola I, Galan C, Mateos-Gomez PA, Palacio L, Zuniga S, Cruz JL, et al. The polypyrimidine tract-binding protein affects coronavirus RNA accumulation levels and relocalizes viral RNAs to novel cytoplasmic domains different from replication-transcription sites. J Virol 2011;85:5136–49. https://doi.org/10.1128/jvi.00195-11.

[60] Iwata H, Goettsch C, Sharma A, Ricchiuto P, Goh WWB, Halu A, et al. PARP9 and PARP14 cross-regulate macrophage activation via STAT1 ADP-ribosylation. Nat Commun 2016;7. https://doi.org/10.1038/ncomms12849.

[61] Albiston AL, McDowall SG, Matsacos D, Sim P, Clune E, Mustafa T, et al. Evidence that the angiotensin IV (AT 4) receptor is the enzyme insulin-regulated aminopeptidase. J Biol Chem 2001;276:48623–6. https://doi.org/10.1074/jbc.C100512200.

[62] Negrete OA, Wolf MC, Aguilar HC, Enterlein S, Wang W, Mühlberger E, et al. Two key residues in EphrinB3 are critical for its use as an alternative receptor for Nipah virus. PLoS Pathog 2006;2:0078–86. DOI:10.1371/journal.ppat.0020007.

[63] Hernandez CC, XiangWei W, Hu N, Shen D, Shen W, Lagrange AH, et al. Altered inhibitory synapses in de novo GABRA5 and GABRA1 mutations associated with early onset epileptic encephalopathies. Brain 2019;142:1938–54. DOI:10.1093/brain/awz123.

[64] Roy SM, Minasov G, Arancio O, Chico LW, Van Eldik LJ, Anderson WF, et al. A selective and brain penetrant p38αMAPK inhibitor candidate for neurologic and neuropsychiatric disorders that attenuates neuroinflammation and cognitive dysfunction. J Med Chem 2019;62:5298–311. https://doi.org/10.1021/acs.jmedchem.9b00058.

[65] Belden Z, Deiuliis JA, Dobre M, Rajagopalan S. The role of the mineralocorticoid receptor in inflammation: focus on kidney and vasculature. Am J Nephrol 2017;46:298–314. https://doi.org/10.1159/000480652.

[66] Raucci F, Mansour AA, Casillo GM, Saviano A, Caso F, Scarpa R, et al. Interleukin-17A (IL-17A), a key molecule of innate and adaptive immunity, and its potential involvement in COVID-19-related thrombotic and vascular mechanisms. Autoimmun Rev 2020;19:102572. https://doi.org/10.1016/j.autrev.2020.102572.

[67] Kumar A, Zhang KYJ. Advances in the development of shape similarity methods and their application in drug discovery. Front Chem 2018;6. DOI:10.3389/fchem.2018.00315.

[68] Conway B, Shafran SD. Pharmacology and clinical experience with amprenavir. Expert Opin Invest Drugs 2000;9:371–82. https://doi.org/10.1517/13543784.9.2.371.

[69] Plantone D, Koudriavtseva T. Current and future use of chloroquine and hydroxychloroquine in infectious, immune, neoplastic, and neurological diseases: a mini-review. Clin Drug Investig 2018;38:653–71. https://doi.org/10.1007/s40261-018-0656-y.

[70] Zuehlke RL, Lillis PJ, Tice A. Antimalarial therapy for lupus erythematosus: an apparent advantage of quinacrine. Int J Dermatol 1981;20:57–60. https://doi.org/10.1111/j.1365-4362.1981.tb05295.x.

[71] Hellstrom WJG, Gittelman M, Karlin G, Segerson T, Thibonnier M, Taylor T, et al. Vardenafil for treatment of men with erectile dysfunction: efficacy and safety in a randomized, double-blind, placebo-controlled trial. J Androl 2002;23:763–71.

[72] Doggrell SA. Comparison of clinical trials with sildenafil, vardenafil and tadalafil in erectile dysfunction. Expert Opin Pharmacother 2005;6:75–84. https://doi.org/10.1517/14656566.6.1.75.

[73] McCurdy AR, Lacy MQ. Pomalidomide and its clinical potential for relapsed or refractory multiple myeloma: an update for the hematologist. Ther Adv Hematol 2013;4:211–6. https://doi.org/10.1177/2040620713480155.

[74] Anderson KC. Lenalidomide and thalidomide: mechanisms of action—similarities and differences. Semin Hematol 2005;42:S3–8. https://doi.org/10.1053/j.seminhematol.2005.10.001.

[75] Inoue K, Yuasa H. Molecular basis for pharmacokinetics and pharmacodynamics of methotrexate in rheumatoid arthritis therapy. Drug Metab Pharmacokinet 2014;29:12–9. https://doi.org/10.2133/dmpk.DMPK-13-RV-119.

[76] Bleich HL, Boro ES, Frei III E, Jaffe N, Tattersall MHN, Pitman S, et al. New approaches to cancer chemotherapy with methotrexate. N Engl J Med 1975;292:846–51. https://doi.org/10.1056/NEJM197504172921607.

[77] Przekop PRJ, Tulgan H, Przekop AA, Glantz M. Adverse drug reaction to methotrexate: pharmacogenetic origin. J Am Osteopath Assoc 2006;106:706–7.

[78] Liu S, Zheng Q, Wang Z. Potential covalent drugs targeting the main protease of the SARS-CoV-2 coronavirus. Bioinformatics 2020;36:3295–8. DOI:10.1093/bioinformatics/btaa224.

[79] Ma L, Kohli M, Smith A. Nanoparticles for combination drug therapy. ACS Nano 2013;7:9518–25. https://doi.org/10.1021/nn405674m.