



OPEN

## In silico analysis of SARS-CoV-2 proteins as targets for clinically available drugs

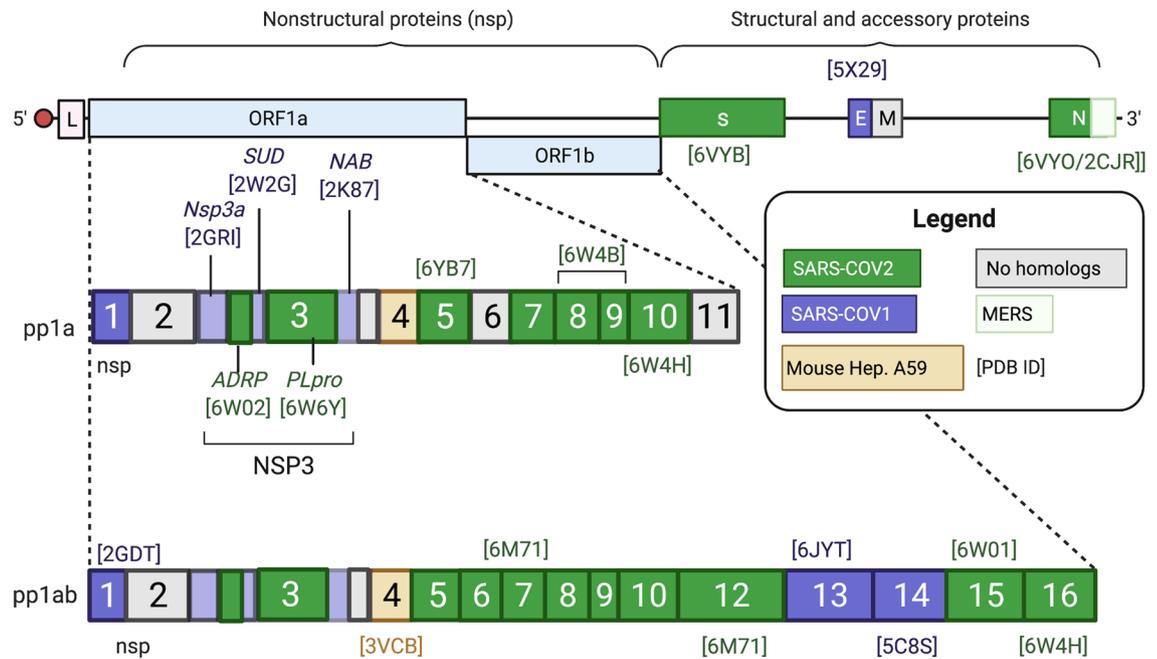
Wallace K. B. Chan<sup>1,2,5</sup>, Keith M. Olson<sup>1,2,5</sup>, Jesse W. Wotring<sup>3</sup>, Jonathan Z. Sexton<sup>3,4</sup>, Heather A. Carlson<sup>3</sup> & John R. Traynor<sup>1,2,3✉</sup>

The ongoing pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) requires treatments with rapid clinical translatability. Here we develop a multi-target and multi-ligand virtual screening method to identify FDA-approved drugs with potential activity against SARS-CoV-2 at traditional and understudied viral targets. 1,268 FDA-approved small molecule drugs were docked to 47 putative binding sites across 23 SARS-CoV-2 proteins. We compared drugs between binding sites and filtered out compounds that had no reported activity in an in vitro screen against SARS-CoV-2 infection of human liver (Huh-7) cells. This identified 17 “high-confidence”, and 97 “medium-confidence” drug-site pairs. The “high-confidence” group was subjected to molecular dynamics simulations to yield six compounds with stable binding poses at their optimal target proteins. Three drugs—amprenavir, levomefolic acid, and calcipotriol—were predicted to bind to 3 different sites on the spike protein, domperidone to the Mac1 domain of the non-structural protein (Nsp) 3, avanafil to Nsp15, and nintedanib to the nucleocapsid protein involved in packaging the viral RNA. Our “two-way” virtual docking screen also provides a framework to prioritize drugs for testing in future emergencies requiring rapidly available clinical drugs and/or treating diseases where a moderate number of targets are known.

The rapidly progressing coronavirus 2019 (COVID-19) pandemic, caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), requires therapeutic strategies that can quickly enter clinical trials to minimize the human, social, and economic impact. Repurposing of FDA-approved drugs has allowed potential anti-SARS-CoV-2 drugs to rapidly enter clinical trials—such as remdesivir, darunavir, and lopinavir<sup>1–6</sup>; repurposing efforts reduce drug development time, cost, and early failures due to safety concerns<sup>7</sup>. Unfortunately, most drug repurposing studies examine only a few classical viral targets/sites or use phenotypic screens that do not identify the drug target<sup>8–11</sup>. Pragmatically, drug repurposing success is limited by the number of clinically approved drugs and the number of SARS-CoV-2 targets for those drugs to act<sup>12</sup>. Moreover, most drugs repurposed for use against SARS-CoV-2 remain unproven in clinical trials, produce conflicting clinical results, or show marginal effectiveness<sup>13–16</sup>. Many people have now been vaccinated against the virus, but there are many who, for some reason or another, are not receiving vaccines and the possible scenario that variants of the virus may be resistant to current vaccines. Thus, there is an ongoing need for therapies, especially those acting at new or understudied SARS-CoV-2 proteins and sites.

The SARS-CoV-2 genome primarily consists of 4 structural proteins that form the virion and mediate cell entry and 16 nonstructural proteins (Nsps) that form the viral replication-transcription complex and inhibit host immune responses; both classes offer numerous potential antiviral targets. Notable structural proteins include the nucleocapsid (N), which packs viral RNA<sup>17–19</sup>, and the spike (S), a ~200 kDa homotrimeric protein that mediates viral entry through recognition by its receptor-binding domain (RBD) of the host angiotensin-converting enzyme 2 (ACE2) receptor<sup>20</sup>. After entry, the host ribosome synthesizes the precursor polyprotein, which the Nsp3 (or PLPro) and Nsp5 (or MPro) proteases activate via auto-cleavage into the 16 different Nsps<sup>21</sup>. Most drug repurposing studies screen the same classical antiviral targets, such as the active sites of Nsp3 or Nsp5, the spike RBD, or the RNA-dependent-polymerase active site (Nsp12) of the replication-transcription complex, using either in silico or experimental methods<sup>16,20,22–24</sup>.

<sup>1</sup>Department of Pharmacology, University of Michigan, 2301 MSRBIII, 1150 W Medical Center Dr, Ann Arbor, MI 48190-5606, USA. <sup>2</sup>Edward F Domino Research Center, University of Michigan, Ann Arbor, MI 48190, USA. <sup>3</sup>Department of Medicinal Chemistry, University of Michigan, Ann Arbor, MI 48190, USA. <sup>4</sup>Department of Internal Medicine, University of Michigan, Ann Arbor, MI 48190, USA. <sup>5</sup>These authors contributed equally: Wallace K. B. Chan and Keith M. Olson. ✉email: jtraynor@umich.edu



**Figure 1.** SARS-CoV-2 RNA Genome and Selected PDB Structures. Simplified SARS-CoV-2 RNA genome shows the proteins and PDB structures used for the *in-silico* drug repurposing screen. i. SARS-CoV-2 structures are in green. In the absence of a SARS-CoV-2 structure, homology models were built based on the closely related SARS-CoV-1 (purple), MERS (green outline), or Mouse Hepatitis A59 (yellow) [E value < 0.001]. Proteins with no SARS-CoV-2 or homologous structures were included in subsequent steps. The [PDB ID] of the chosen structure(s) is included above or below the appropriate gene. Created with Biorender.com.

Potential drugs for repurposing could target numerous additional protein sites. For example, drugs acting at allosteric sites or protein–protein interfaces (PPIs) on the spike protein might reduce viral entry to mammalian host cells<sup>25</sup>. Similarly, attractive targets include less studied domains/proteins, such as the Mac1 domain of Nsp3<sup>26,27</sup> and the endonuclease Nsp15<sup>28</sup>, both likely involved in blocking the host immune response. Docking-based virtual screens enable the quick identification of leads by screening many drugs across single targets<sup>26,27</sup>, or single drugs across many targets<sup>29</sup>, or a combination of both<sup>30</sup>. For example, virtual screens have been used to identify plant products that may inhibit important viral proteins such as the main protease (Nsp 5 or Mpro)<sup>31</sup>, Nsp1<sup>32</sup>, RNA-dependent polymerase<sup>33</sup>, the spike protein binding domain<sup>34</sup>, Nsp15<sup>33,35</sup> and Nsp16<sup>36</sup>. On the other hand, many virtual screens against SARS-CoV-2 targets are limited because they use single or relatively few sites, do not refine or prioritize drug/site poses, and/or ignore available biological data<sup>37,38</sup>. These factors prevent an objective comparison across different drugs and prioritize computational methods over biological data.

To address these drawbacks, we developed a "two-way" multi-ligand and multi-target virtual docking screen to study 1268 FDA-approved drugs at 48 established or predicted sites across 23 SARS-CoV-2 proteins to identify drug repurposing candidates for SARS-CoV-2 treatment. This strategy increases the probability of finding treatments by expanding the number of potential target sites and thus the effective library size. This approach also offers the possibility to develop drug combinations acting at two or more viral sites that could provide better anti-Covid-9 therapies compared to single target treatments<sup>39</sup>. We refined and normalized the docking scores into Z scores, and further, we compared our computational leads with experimental hits from a large phenotypic screen for anti-SARS-CoV-2 infection of human Huh-7 cells<sup>10</sup>. This resulted in the identification of 97 unique drug/site pairs across 45 sites on 23 SARS-CoV-2 proteins, from which we prioritized 17 compounds. We further assessed these hits for stability within their respective proposed binding sites using molecular dynamics and identified six potential repurposed drugs with direct antiviral potential against SARS-CoV-2.

## Results

**Identification of target proteins and binding sites.** Predicted SARS-CoV-2 protein sequences from the .NCBI.org database (Fig. 1) were subjected to a protein database (PDB) BLAST search to identify SARS-CoV-2 or related protein structures with an E value < 0.001. We selected SARS-CoV-2 protein/domain structures based on the lowest E value, then structural resolution if multiple SARS-CoV-2 structures existed (Fig. S1A). If no SARS-CoV-2 structures were available (as of May 2020), homology models were constructed based on solved structures from other coronaviruses, namely SARS-CoV-1, Middle East Respiratory Syndrome (MERS), or Mouse Hepatitis A59. Homology models were refined using the GB/VI force field model in Molecular Operating Environment (MOE)<sup>40</sup>, and the best scoring model for each protein was used in docking studies (Table S1). In all, this provided structures of 23 SARS-CoV-2 proteins, including 14 known structures and 9 homology models, for docking the repurposing library.

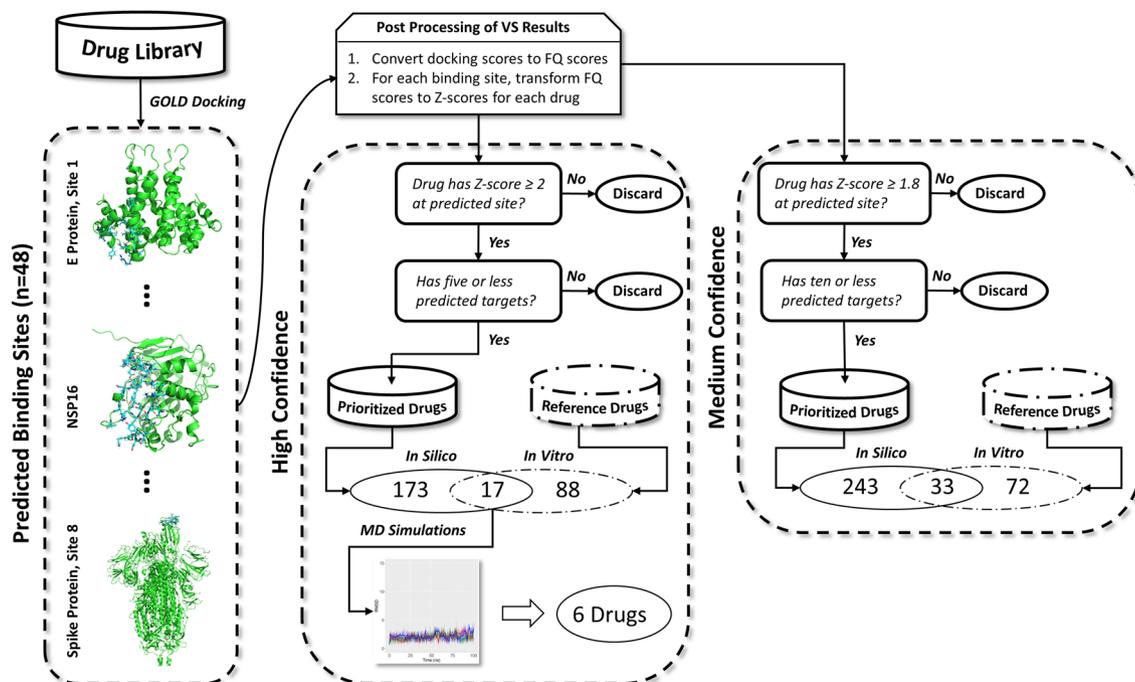
Protein Domain	Species	E-value	PDB ID	Protein Domain	Species	E-value	PDB ID
NSP1	SARS-CoV-1	E-52	2GDT	NSP9	SARS-CoV-2	–	6W4B
NSP3				NSP10	SARS-CoV-2	E-79	6W4H
NSP3a (834–930)	SARS-CoV-1	E-24	2GRI	NSP12/RdRp	SARS-CoV-2	–	6M71
ADRP/MAC-1 (1001–1183)	SARS-CoV-2	0	6W6Y	NSP13	SARS-CoV-1	E-54	6JYT
SUD (1184–1540)	SARS-CoV-1	E-105	2W2G	NSP14	SARS-CoV-1	E-33	5C8S
PLPro (1640–1887)	SARS-CoV-2	E-6	6W6Y	NSP15	SARS-CoV-2	–	6W01
NAB (1888–1997)	SARS-CoV-1	E-41	2K87	NSP16	SARS-CoV-2	–	6W4H
NSP 4	Mouse Hep. A59	E-24	3VCB	S–Spike	SARS-CoV-2	–	6VYB (open), 6VXX (closed), 6VW1 (w/hACE2)
NSP5	SARS-CoV-2	–	6YB7	E Protein	SARS-CoV-1	E-9	5X29
NSP7	SARS-CoV-2	E-25	6M71	Nucleoprotein	SARS-CoV-2	0	6VYO
NSP8	SARS-CoV-2	E-25	6M71	CTD (248–265)	MERS	E-59	2CJR

**Table 1.** SARS-CoV-2 proteins, domains, and structure.

To virtually screen the repurposing library, we identified known and potential drug binding sites on each SARS-CoV-2 protein, including “classical” sites (e.g., spike protein RBD or protease active site), protein–protein interfaces (PPIs), sites with co-crystallized ligands, and putative allosteric sites. The “site finder” function in MOE was used to uncover additional potential binding sites on each protein based on their “Propensity for Ligand Binding (PLB)” indices, which scores binding pockets based on cavity size, depth, and amino-acid composition<sup>41,42</sup>. Sites that scored well with available crystal structures of SARS-CoV-2, SARS-CoV-1, MERS, or Mouse hepatitis A59 proteins were designated as PPIs. Other sites that scored well but had no apparent binding partner were designated as putative allosteric sites. Notably, this method identified most classical antiviral sites such as the active sites on Nsp3 and Nsp5 and the Nsp12 RNA binding sites, providing an internal validation of the approach; the only site the method did not identify was the spike protein recognition site for ACE2, most likely due to its flat and exposed surface. Lastly, sites with co-crystallized small molecules or nucleotides with an MW > 150 present were included as potential sites. In total, these analyses resulted in 48 binding sites (Table 1; see Table S2 for details of the amino-acid composition of each site).

**Docking-based “two-way” virtual screening campaign.** The pipeline used for our virtual screening campaign is illustrated in Fig. 2. Using GOLD software (Cambridge Crystallographic Data Center, see methods), we docked 1268 FDA-approved small molecules from DrugBank (see methods) to all 48 binding sites across 23 SARS-CoV-2 proteins. Docking scores were normalized to fit quality (FQ) scores to minimize bias towards higher molecular weight compounds and rescue low MW drugs that would otherwise have been discarded (Fig. S3). Subsequently, FQ scores were transformed into Z scores (Fig. S2) for each drug at each binding site to directly compare across drug-site pairs. Both “high confidence” as defined by a Z score of 2 or greater and “medium confidence” as defined by a Z score of 1.8 or greater strategies were employed to identify compounds. Of the drugs that had  $Z \geq 2$  at any given site, a decreasing trend was observed in the number of predicted sites bound with respect to frequency of drugs; 5 sites were chosen as the cutoff point. Additionally, all 48 predicted sites from the SARS-CoV-2 proteins bound at least one predicted drug with  $Z \geq 2$ , and these drugs appeared relatively evenly distributed among the sites. From this, we characterized 190 drugs (15% of the library) as high confidence hits. By relaxing the screening criteria to  $Z \geq 1.8$  and limiting the number of interacting sites for any one drug to 10 different sites, we identified an additional 86 compounds for a total of 276 hits (22% of the virtual library). These medium confidence compounds were predicted to interact with various proteins and binding sites yielding 97 unique drug-binding site pairs across 45 different sites on the 23 SARS-CoV-2 proteins investigated (Table S3). The criteria chosen for the high confidence hits allowed us to select a reasonably sized pool of drugs while minimizing promiscuity. We loosened the cut-offs to identify medium confidence hits that would allow for more false negatives to enter the hit pool, while limiting the number of false positives; further lowering of the criteria would have caused the inclusion of many false positives This is illustrated graphically in Fig. S2.

**Comparison of in silico hits with experimental data.** A recent report of drug repurposing identified 15 drugs in our library as having high (20–765 nM) potency at reducing SARS-CoV-2 infection in human Huh-7 cells<sup>10</sup>. Only one of these compounds, domperidone, was identified in our virtual screen using the high confidence criteria. However, the experimental study identified an additional 88 hits that were effective at 2  $\mu$ M or better that have not yet been further characterized (Table S4). Of these compounds, 17, including domperidone, overlapped with the high confidence group of drugs. Our analysis predicted these 17 drugs to bind 25 different sites with 33 different drug-site combinations (Table S5). A heat map of these 17 compounds at each of the 48 predicted or actual SARS-CoV-2 protein binding sites is shown in Fig. 3. Of the hits, 8 had Z values  $\geq 2$  at more than one site, for example, amprenavir at Nsp15 and Nsp5, and avanafil at two distinct sites on Nsp15. Overall, 72 of the high priority drugs identified by docking did not map with the experimentally determined hits, and 87 of the experimental hits did not appear in our high confidence list (Fig. 3). When the medium confidence hits were included, a further 16 drugs were identified as active in both the in silico and experimental screen (Table S3).



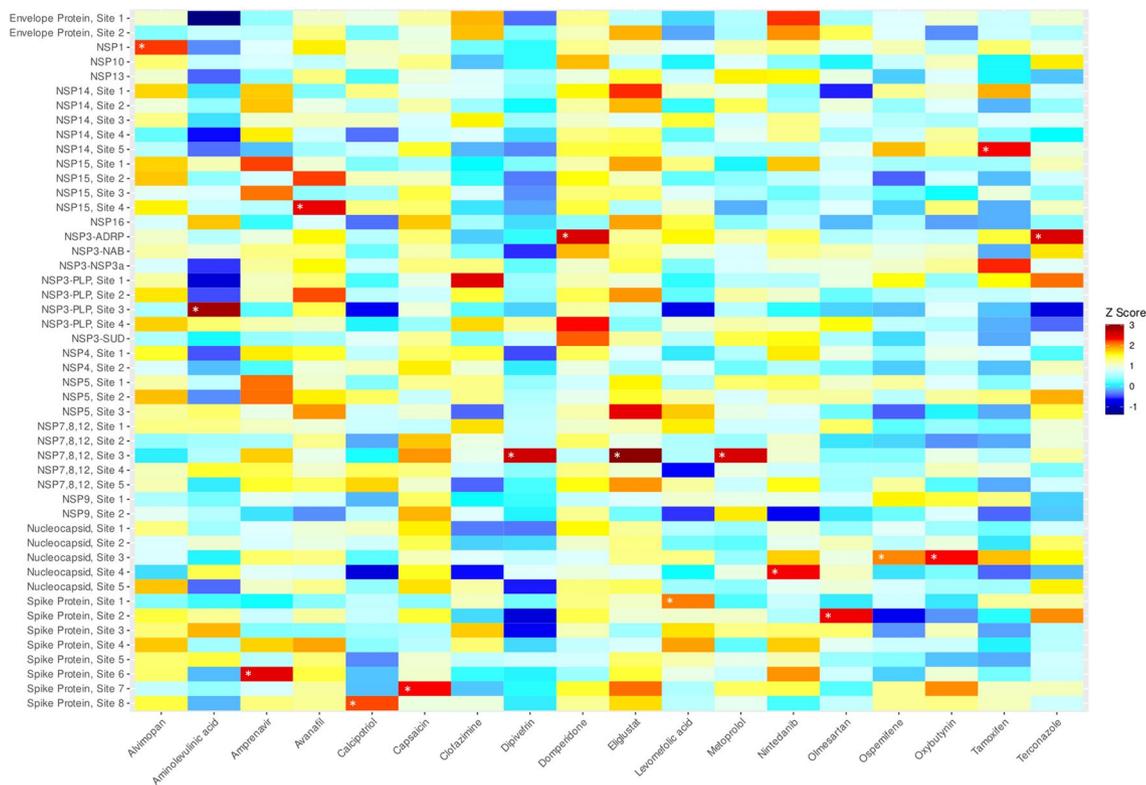
**Figure 2.** Pipeline for docking-based, “two-way” virtual screening strategy. GOLD software was used to dock the drug library against each of the 48 predicted binding sites. All docking scores were converted to fit quality (FQ) scores. Subsequently, the FQ scores of each drug were transformed into Z scores for each site. Out of the 1,268 FDA-approved drugs, 190 were prioritized with a Z score  $\geq 2$  at 5 or fewer target sites; these were considered high confidence hits. Consequently, 17 overlapped with a set of drugs previously known to have marginal to high potency against SARS-CoV-2 in vitro and were considered virtual hits<sup>10</sup>. These were then subjected to MD simulations, in which 6 drugs were observed to bind stably to their respective target protein. A medium confidence set of drugs was also prepared based on a Z score  $\geq 1.8$ , which prioritized 276 drugs; 33 of these overlapped with the reference drugs. Overall, there were 34 and 97 unique drug-site pairs for the high- and medium confidence hits respectively that overlapped with the reference drugs. In-silico, prioritized drugs are shown in cylinders with solid lines and reference drugs with in vitro activity against SARS-CoV-2 in cylinders with dashed-dotted lines. Prepared using PowerPoint Presentation Software.

Indeed, five of these drugs were potent inhibitors of SARS-CoV-2 infection of Huh-7 cells<sup>10a,b</sup>, namely Bosutinib,  $IC_{50} = 20$  nM; Fedratinib,  $IC_{50} = 24$  nM; domperidone,  $IC_{50} = 44$  nM; remdesivir,  $IC_{50} = 97$  nM; amiodarone,  $IC_{50} = 167$  nM; lomipatide,  $IC_{50} = 765$  nM).

**Molecular dynamics simulations and trajectory analysis.** To account for the limitations of molecular docking, which include the absence of solvent, minimal sidechain and backbone flexibility, we performed molecular dynamics simulations on our 17 top in-silico hits using the drug-binding site pairing that gave the highest Z score (Table S5) to evaluate the stability of the bound drugs in a quasi-physiological environment. Simulations were run for 100 ns in three independent trials followed by a molecular dynamics trajectory analysis to assess protein and drug stability. From these simulations, we further refined our list down to 6 drugs that met the following criteria: (1) the final snapshot shows the same drug pose in at least 2 replicates, (2) the drug RMSD converges for at least two of the replicates, and (3) the protein remains stable during the simulation. 6 drug/site pairs met these criteria. (Figs. S4 and S5). Drug RMSDs from those systems that did not meet these criteria are shown in Fig. S6.

Most proteins remained stable throughout the simulation (Fig. S4). Nsp14 moderately diverged from the original structure, likely because we used a homology model. In addition, in the presence of ospemifene, the nucleocapsid complex partially or entirely destabilized in all three replicates (Fig. S4), although this did not occur in the presence of nintedanib or oxybutynin. Lastly, the spike protein showed a moderate shift in all tests; we ascribe this to the numerous loops in the structure, which are not observed in the X-ray analysis and had to be modeled.

In contrast to the stability of the proteins, only six drugs (domperidone, avanafil, nintedanib, levomefolic acid, amprenavir, and calcipotriol) remained in their predicted binding pockets throughout the simulations (Fig. S5). Most drugs either migrated to another site or diffused into the solvent early in the simulation, including aminolevulinic acid and metoprolol, or showed instability in two or more replicates. With terconazole, the ligand trajectories converged in two trials, achieving a similar ligand orientation but outside the predicted site. With alvimopan and dipivefrin, the drugs fully dissociated from the protein in one replicate, migrated to another site in another, and remained stable in yet another. Overall, the study predicted that: domperidone and avanafil



**Figure 3.** Heat map of hits from composite virtual screening campaign. Under our prioritization scheme, 17 high confidence, FDA-approved drugs were predicted to interact in 34 unique drug-site combinations. White asterisks denote the systems selected for molecular dynamics simulations. The heat map was generated with R language using the ggplot2 package.

bound stably with different Nsps, that levomefolic acid, amprenavir, and calcipotriol bound to the spike protein, but at three different sites, and that nintedanib bound stably with the nucleocapsid (Fig. 4).

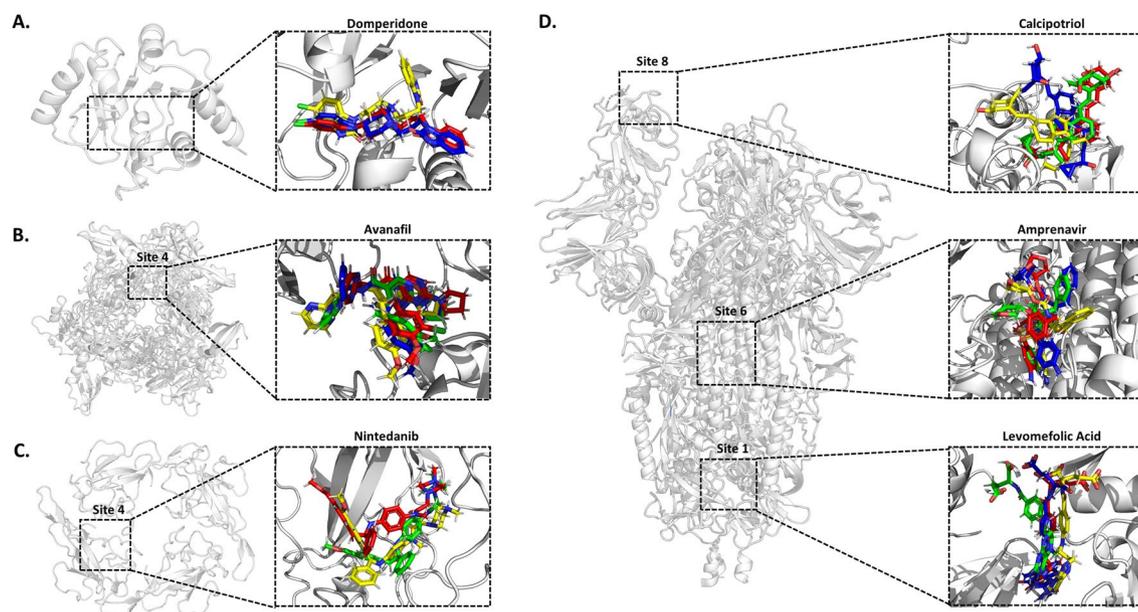
Domperidone bound to the ADP-ribose-1-monophosphatase (ADRP or Mac1) active site of Nsp3<sup>41-44</sup> (Fig. S7A). In two of three simulations, a cryptic pocket opened to accommodate the 1,3-benzodiazol-2-one moiety of the drug (Moiety 1B) (Fig. S7B/C). This small cavity was formed by Leu160 moving outward and Phe156 rotating inward to stabilize drug binding with a side-to-face aromatic interaction (Fig. S7E). In the presence of domperidone, the binding pocket adopted a more open conformation in the 'clamp' region at the ADP ribose site, by an outward movement of Ile131 (Fig. S7D), perhaps due to relaxation from the crystal structure. The avanafl pose was at the interface between Nsp15 homomers (Figs. 4, and S8A/B), buried in a hydrophilic pocket with access to a solvent channel that leads to the protein surface (Fig. S8C). While the simulations with avanafl employed a single site, the enzymatically active form of Nsp15 is a homohexamer such that avanafl could bind at any of the six comparable interfaces<sup>28</sup> to disrupt the quaternary structure. However, our MD timescale was not long enough to predict complete protein disassociation.

The spike protein is an obvious target for antiviral activity. Levomefolic acid showed binding to a cavity between the spike protein monomers in the trimeric state (Fig. S9). Amprenavir exhibited modest conformational flexibility within a putative hydrophobic pocket on the spike protein subunit 2, located near the S2' cleavage site between residues 815/816<sup>45</sup> (Fig. S10). In this pocket, the sulfonamide-containing moiety remained stable and buried in the pocket, although the benzyl and oxolan-3-yl acetyl groups were more dynamic (Fig. S10B/C). The drug also creates several contacts with the fusion peptide<sup>46</sup> (Fig. S10A). Calcipotriol interacted with the RBD. Calcipotriol was positioned flat against the protein, though simulations revealed that the (1R,3S)-4-methylidenecyclohexane-1,3-diol-containing portion of the drug was mobile and preferred to be in the solution phase (Fig. S11A/C), while the rest of the drug was firmly sequestered within a hydrophobic pocket (Fig. S11B/C). This resulted in calcipotriol partially blocking the ACE2 receptor binding site.

Finally, nintedanib was observed to bind stably to the RNA-binding pore of the nucleocapsid (Fig. S12). As the nucleocapsid is a homotetramer, it could be predicted to bind 4 molecules of nintedanib.

## Discussion

To identify potential new drug repurposing candidates with anti-SARS-CoV-2 activity, we designed a "two-way" virtual docking screen of FDA-approved drugs at 48 diverse sites, including classic antiviral sites, PPIs, sites with co-crystallized small molecules, and predicted allosteric sites to maximize the chance of success. To compare drug/site pairs, we report a prioritization scheme that normalized all docking scores to Z scores<sup>38</sup>. To the best of our knowledge, this is the first application of such a methodology to a virtual screening pipeline and allowed us

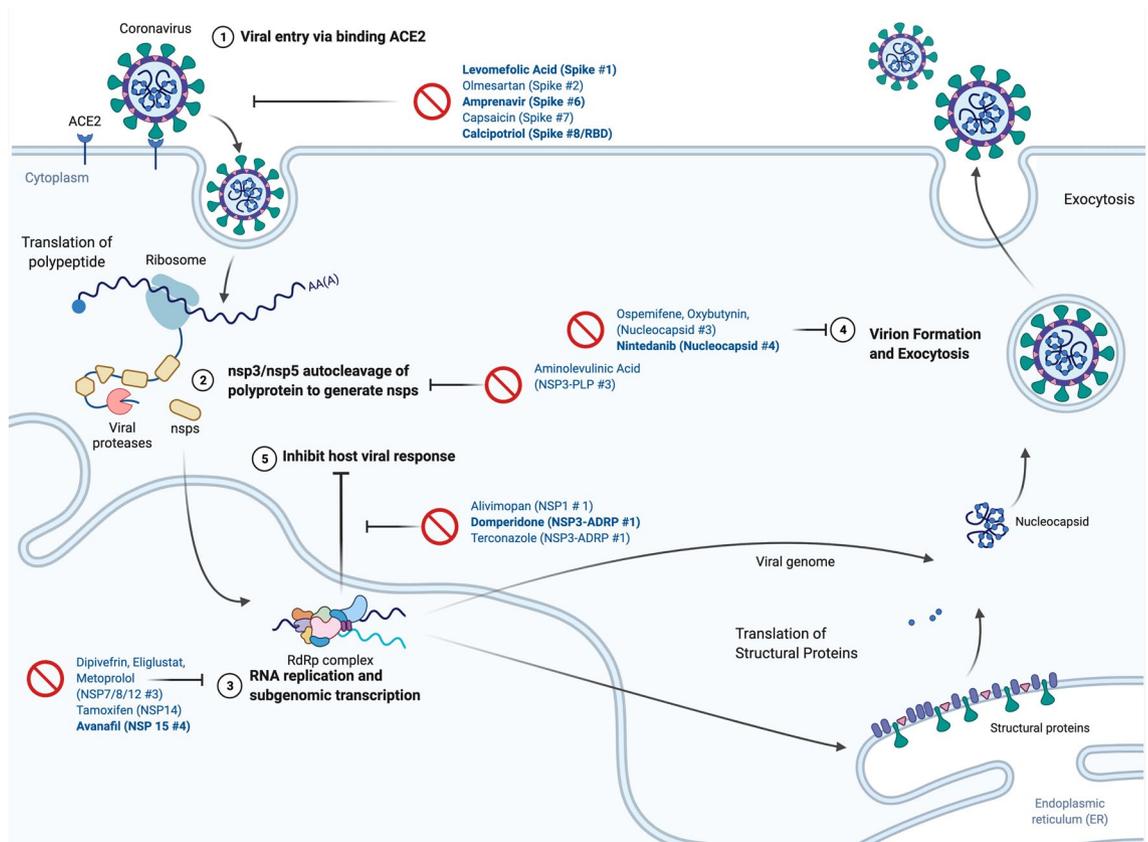


**Figure 4.** Final snapshots from molecular dynamics simulations for stable drug-site systems. The docked drug pose is shown in yellow, while the final pose after three independent 100 ns molecular dynamics simulations is shown in red (1), green (2), and blue (3). (A) ADRP domain of NSP3 with domperidone docking and final molecular dynamic simulation poses. (B) Site 4 of NSP15 with Avanafil docking and final molecular dynamic simulation poses shown. ADP is shown in dark red. (C) Site 4 of the RNA-binding domain of the nucleocapsid protein with nintedanib docking and final molecular dynamic simulation poses shown. (D) Spike protein sites 1, 6, and 8 with levomefolic acid, amprenavir, and calcipotriol docking and final molecular dynamic simulation pose shown. In one trial, domperidone and nintedanib migrated to another part of the protein (not shown). Please see Figs. S7–S12 for additional descriptions of the trajectories. Molecular graphics were generated using PyMOL and the figure was prepared using PowerPoint Presentation Software.

to identify a diverse selection of potential anti-SARs-CoV-2 drugs. This resulted in the *in silico* characterization of many unique drug-site pairs while retaining information from 48 single-site screens. Although we do not yet have experimental evidence that all of these sites, particularly the ones that are not classical targets for antiviral drugs, are able to modulate viral protein activity we did identify 15 compounds from our screen that have anti-SARS-CoV-2 activity in an *in vitro* screen in human Huh-7 cells (Fig. 5), and 6 of these showed stable interactions with SAR-CoV-2 proteins in molecular dynamics simulations (Fig. 5). Unfortunately, domperidone, which was highly potent in the experimental screen and a high confidence hit in the *in silico* screen, was inactive against the virus in a follow-up study that examined anti-viral activity in four other cell lines<sup>10</sup>, which may suggest interactions with mammalian cell targets rather than viral targets, or that binding at the Mac3 site of NSP3 does not produce viral inhibition. This host cell target is unlikely to be the dopamine D2 receptor, the clinical target for domperidone in treating nausea and vomiting, since there is no evidence that the D2 receptors are expressed in Huh-7 cells; this suggests a novel, but cell-specific, mechanism of antiviral action<sup>47</sup>. Nonetheless, our approach does provide a framework for the *in silico* screening of repurposed drug libraries against many protein targets and a starting point to develop novel antiviral drugs from leads predicted in this study.

Previous virtual drug repurposing screens against multiple SARS-CoV-2 targets suffer from several limitations, including the use of relatively few sites and/or too many sites obtained from automated methods, reliance on docking scores alone to prioritize drugs that favorably bias large molecular weight ligands, biased comparison of drugs between binding sites, the absence of molecular dynamics simulations to further prioritize lead docked structures<sup>37,38</sup>, and a lack of use of known biological data on anti-SARS-CoV-2<sup>29,30</sup> activity. Since many of the putative sites and drug-target interactions we have identified are novel we were unable to retrospectively benchmark our methodology. However, an original aspect of the present study, in comparison to previous *in silico* work, is that we employed biological data in our decision tree. These data were obtained from a phenotypic cell-based anti-viral assay. Such a methodology is expected to discover drugs that bind/modulate both viral and/or host factors. The cross-over hit rate between our virtual screen and the phenotypic screen was not high, but this is most likely because the *in silico* methodology only identified the subset of compounds acting at viral proteins and not on host factors. Additionally, binding sites among the proteins analyzed may not have been pre-formed in the crystal structures. For example, a spike protein crystal structure shows linoleic acid bound in a novel pocket not previously observed in other crystal structures<sup>48</sup>.

A second novel aspect of our study was the conversion of docking scores to fit quality (FQ) scores, which is essentially a scaled ligand efficiency metric that has not previously been reported in virtual screening campaigns. Additionally, we normalized FQ scores to Z scores to prioritize different drug-site pairs, a method that had previously only been employed to analyze adverse drug reactions<sup>38</sup>. From the Z-scores, we identified a group



**Figure 5.** Predicted Sites and Mechanism of 17 “High-Confidence” drug/site pairs. The 17 drugs with Z score  $\geq 2$  at 5 or fewer target sites and previously identified anti-SARS-CoV-2 activity in vitro. The predicted drug/site pairs and commonly drugged steps in the (simplified) viral life cycle including (1) blockade of viral entry mediated via the S protein recognizing the host ACE2; (2) inhibition of viral proteases NSP3 and NSP5 which auto-cleave and activate the 16 nsps, (3) inhibition of the viral replication/transcription complex; (4) blockade of virion formation, RNA packaging, and exocytosis, and (5) blockade of viral proteins that attenuate the host immune response. Drugs names in bold indicate are “High-Confidence” hits predicted to form stable drug/receptor interaction in molecular dynamics simulations, non-bolded drugs failed to form stable complexes in the MD simulations. Drug lists ordered as, “Drug (Protein Site #)”. Created with biorender.com.

of potential anti-SARS-CoV-2 drugs. In traditional virtual screening methods, the goal is to minimize false positives<sup>49</sup>, but since the size of the FDA-approved drug library is fixed, we included 97 “medium-confidence” drug/site pairs using less-stringent criteria. These 97 drug/site pairs represent 0.002% of the possible 60,864 combinations (48 sites  $\times$  1268 drugs). This provides a resource for further investigating potential SARS-CoV-2 drugs identified in phenotypic screens without an obvious target, where target deconvolution remains a major barrier<sup>50</sup>. Moreover, several of the identified drugs are predicted to act at multiple sites across the same or different proteins, suggesting drug combinations may produce additive or even synergistic antiviral activity<sup>29,30</sup>.

One caveat to our work is that we do not know if occupancy of many of putative binding sites we identified will lead to modulation of viral protein activity and it will be important in the future to substantiate the role of each of the putative sites in the inhibition of the target. Such a large study is outside the scope of this study. Nonetheless, Table S2 lists all the sites by their type, whether classical (i.e., well-validated), putative allosteric (i.e., not necessarily validated but predicted site without an established mechanism), or protein–protein interactions. Inclusion of this information will provide a starting point for future studies, especially to help study newly discovered anti-SARS-CoV-2 compounds that may not have a known mechanism of action.

Overall, we identified 17 compounds in the high confidence group that overlapped with compounds that showed some degree of antiviral activity against SARS-CoV-2 infection in Huh-7 cells. These compounds bound to 48 sites on the virus proteins. The highest scoring drug-binding site interactions were subjected to molecular dynamics to provide six stable drug-binding site pairs. Moreover, the predicted binding site for each drug provides a logical rationale as to why the compounds could interfere with the activities of their target proteins and so possess antiviral activity. Domperidone bound to the ADP-ribose site on the Mac1 domain of Nsp3. This would compete with the hydrolysis of the natural substrate, which is associated with virulence and the ability to evade host immune response<sup>43,44</sup>. In fact, preclinical models of SARS-CoV-2, SARS-CoV-1, and MERS-CoV<sup>27</sup> suggest Mac1 inhibitors may act as broad-spectrum antivirals. Avanafil, a phosphodiesterase type 5 inhibitor bound to a putative site at the interior of the Nsp15 hexamer complex. This complex is an uridylyate-specific endoribonuclease that processes viral RNA to avoid activation of host double-stranded RNA (dsRNA) sensors<sup>51</sup>. Active sites

on the enzyme are located externally<sup>28</sup>, so if it were active, avanafil would have to allosterically modulate Nsp15 activity, although some evidence indicates Nsp15 can jeopardize the immune response independent of NendoU activity<sup>50,52</sup>. Nintedanib is a tyrosine kinase inhibitor bound to the nucleocapsid, a protein essential for packaging RNA during viral assembly, replication, and RNA transcription. It is also worth noting that ospemifene, a selective estrogen receptor modulator, was identified as binding to the nucleocapsid protein. This was not included in the final choice of compounds since molecular dynamics simulations of the binding site suggested drug binding decreased the protein's stability. On the other hand, this effect may be an important property because recent reports show that destabilization of the native oligomers by stabilizing non-native oligomers of the N-terminal domain with small molecules, produced antiviral activity<sup>53</sup>.

The spike protein is a principal target for COVID-19 therapeutics due to its essential role in viral entry into the host cell. Amprenavir, levomefolic acid, and calcipotriol were predicted to interact with the spike protein at different locations: a putative allosteric site, a PPI, and the RBD, respectively. Amprenavir bound stably to a potential allosteric hydrophobic pocket on the spike protein subunit 2, which the host TMPRSS2 protease cleaves to enable viral entry, forming predicted contacts with the fusion peptide<sup>44,46</sup>. This acts as a grappling hook, moving into the host membrane and mediating fusion. Levomefolic acid, by binding to a cavity between spike protein oligomer interface, could function as a PPI to prevent the formation of the active heterotrimer. Calcipotriol was predicted to interact stably with the RBD of the spike protein and could ostensibly interfere with the recognition of the host ACE2 receptor. It is also notable that amprenavir was a medium confidence hit at Nsp5 (also known as main protease, MPro, or 3CLPro), which fits with its use as an HIV protease inhibitor, providing two potential mechanisms for this drug.

Unfortunately, although shown to have activity in Huh-7 cells including domperidone which was particularly potent, none of our top six repurposed candidates showed activity against SARS-CoV-2 infection in follow-up experiments using a variety of cell lines<sup>10</sup>, although, in addition to Huh-7 cells, amprenavir has been reported to have moderate anti-SARS-CoV-2 activity against infected Vero E6 cells<sup>52,54</sup>. This could mean the compounds are not effective against the degree of viral loads and/or ease of viral infection in all cell lines or, and more likely, the drugs were acting on host factors in the Huh-7 cells. In support of this, an experimental comparison of repurposed drugs across five different human-derived cell lines showed domperidone to be only active against SARS-CoV-2 infection in Huh-7 cells but not in Vero-E6 cells, Caco-2 E6 cells, or iAEC2 cells<sup>10</sup>. Drugs exerting efficacy through direct interactions with viral proteins should be less dependent on the host cell system, and especially entry inhibitors that act in the extracellular environment. Remdesivir, the preclinical drug Z-FA-FMK, an irreversible inhibitor of cysteine proteases, amiodarone, ipratropium bromide, lomitapide, and lactoferrin, were active across multiple cell lines<sup>10</sup> indicating a stronger potential towards direct viral binding. Remdesivir was identified as a medium confidence, not high confidence hit in our *in silico* screen, presumably because it is a pro-drug and requires conversion to its active metabolite (GS-441524 triphosphate) to inhibit viral RNA polymerase.

A recent meta-analysis of FDA-approved drugs for experimental anti-SARS-CoV-2 activity listed none of our final six candidates, although calcipotriene, like calcipotriol, a synthetic vitamin D analog, is listed<sup>55</sup>. Notably, vitamin D deficiency corresponds to a greater risk of testing positive for COVID-19<sup>56,57</sup>. Nonetheless, it is feasible that drug combinations could inhibit several viral targets, for example, domperidone at Nsp3 together with avanafil at Nsp15, or even a single viral target in different ways, for instance, amprenavir, levomefolic acid, and calcipotriol at the spike protein could provide sufficient antiviral efficacy.

Independently, recent publications on SARS-CoV-2 have confirmed several of our predicted drug sites, for example allosteric sites at the Spike protein<sup>58</sup>, validating our approach. Furthermore, the present study provides a resource to aid the deconvolution of interactions of viral protein sites from phenotypic screens when the actual target is unknown. Moreover, this method and our results can assist in scaffold identification for further development using structure-activity-relationship studies to identify novel drugs against a wide variety of potential targets in SARS-CoV-2. Finally, the “two-way” virtual docking screen we implemented supplies a framework for future emergencies requiring rapidly available clinical drugs for treating diseases where a moderate number of targets are known.

## Methods

**Protein structures and homology models.** The SARS-CoV-2 predicted protein sequences were obtained from the NCBI.org database. The putative protein sequences for each protein were then analyzed using a BLAST to determine sequences with an E score < 0.001. If a SARS-CoV-2 structure was available, that was prioritized. If multiple SARS-CoV-2 structures were available, the highest resolution structure was chosen (Fig. S1). A homology model was built in cases where different domains were available from SARS-CoV-1, MERS, or Mouse Hepatitis A59. SARS-CoV-2 sequences from NCBI were aligned with template structure sequences (Table S1). A homology model using MOE (Molecular Operating Environment)<sup>40</sup> was built for each domain or protein and scored via the GBVI/WSA dG (Generalized-Born Volume Integral/Weighted Surface area) force field function in MOE with a refinement value of 0.5 kcal/mol. Refinements were applied to both intermediate and final models. For each protein, 15 different structures were generated with two different sidechain orientations for a total of 30 models each, using the default settings in MOE. The best scoring model was inspected using the “protein geometry” stereochemical quality evaluation tools in MOE (Table S1).

**Site identification and protein preparation.** All structures were prepared using the MOE “Quick Prep” to determine protonation state, calculate partial charges, and relax strain. Using the structures indicated in Tables 1 and S2<sup>28,59–77,79,80</sup>, sites were chosen using computational and structural analysis (Fig. S1). First, the MOE site finder function was used to identify potential sites. Sites were selected if there was a PLB (Propensity for Ligand Binding) index of > 3 or > 1.5 if no other sites were identified. Based on manual inspection of the

crystal structures, putative catalytic, nucleotide-binding, or drug binding sites were included, if not otherwise incorporated. Next protein–protein interaction sites were chosen based on co-crystallized SARS-CoV-2 protein complexes. Lastly, SARS-CoV-2 structures were overlaid with homologous structures to identify sites of co-crystallized ligands in the homologous structures. In cases of oligomers, only non-redundant sites were selected. Lastly, if solved structures contained small molecules (MW > 150 Da), their position on the protein was considered a potential site.

**Construction of virtual drug library.** Drug structures were obtained from DrugBank (version 5.1.5, 2020-01-0' release, '3D structures.sdf')<sup>81</sup> and imported into a Molecular Operating Environment (MOE, version 2019.0101) database. The pipeline for constructing our virtual drug library is shown in Fig. S2A. Drugs under the label 'approved' corresponding to FDA-approved drugs were isolated from the full set. Afterward, chemical structures were subjected to the 'wash' protocol in MOE to remove salts, generate 3D conformations, and protonated based on the predicted dominant species at pH 7.4. Lastly, a conformational search using LowModeMD with default settings predicted the lowest-energy conformer for each drug. Our final virtual library consisted of 1268 compounds based on the following criteria: (1) presence in DrugBank, (2) not a protein or peptide, (3) molecular weight between 100 and 700 Da, and 4) contain fewer than 15 rotatable bonds.

**Docking-based, "two-way" virtual screen.** GOLD (version 5.8.1; Cambridge Crystallographic Data Center or CCDC) was used for molecular docking of the virtual library against the 47 predicted binding sites at 21 SARS-CoV-2 proteins or domains (Table 1)<sup>82</sup>. From the MOE Site Finder predictions, residues within 4.5 Å of the alpha spheres were used to define the cavity. For each drug, 50 independent genetic algorithm runs were performed using the highest conformational search efficiency setting ('very flexible') and the default diverse solutions options. Ring conformations were explored using the 'flip ring conformers' and 'match conformations' options. Poses were evaluated with the ChemPLP scoring function, and the top-scoring pose of each drug at each site was kept for further analysis.

Since the drugs in our virtual library span between 100 and 700 Da, we normalized the docking scores based on their molecular weight by converting them into ligand efficiency (LE) values as previously reported<sup>83</sup>:

$$LE = \frac{\text{Docking Score}}{HA} \quad (1)$$

where HA is the number of heavy atoms. For similar reasons, a scaling calculation was used to account<sup>83,84</sup>:

$$LE\_Scale = 0.0715 + \frac{7.5328}{HA} + \frac{25.7079}{HA^2} - \frac{361.4722}{HA^3} \quad (2)$$

LE and LE Scales were used to calculate a fit quality (FQ) score:

$$FQ = \frac{LE}{LE\_scale} \quad (3)$$

To make each drug comparable between binding sites, their FQ scores were converted into Z-scores:

$$Z = \frac{FQ - \mu}{\sigma} \quad (4)$$

where the FQ is the FQ score of a drug at a given site.  $\mu$  and  $\sigma$  are the mean and standard deviation, respectively, of the FQ scores at the specified site, which were normally distributed at each site. Ultimately, the poses of drugs with  $Z \geq 2$  at five binding sites or less were retained and considered virtual hits. Drugs that scored well with numerous targets were not considered due to the greater possibility of being promiscuous binders. Custom Perl and R scripts were used to parse, analyze, and visualize the virtual screening results.

**Pose refinement of experimentally verified virtual hits.** The 17 drugs that matched experimental observations and possessed docked poses scoring  $Z \geq 2$  at five binding sites or less were subjected to pose refinement in MOE using the 'Induced Fit' option. Sidechains could freely rotate with a maximum of 10,000 iterations and scored using GBVI/WSA to implicitly account for solvation. Default parameters were used for the other settings.

**Molecular dynamics simulations.** All protein structures were prepared with MOE (version 2019.0101)<sup>38,40</sup> for molecular dynamics simulations. N-linked glycans were removed from the spike protein. N- and C-termini were capped with acetate and N-methyl amide, respectively. Missing loops were modeled with the "Loop Modeller" module. Hydrogen atoms were added, and the protonation state of all relevant residues was assigned at pH 7.0. Prepared structures were saved as PDB files that conformed to AMBER atom typing. Ligands were parameterized with Antechamber using the GAFF2 forcefield<sup>85</sup>, where charges were assigned using the AM1-BCC method. From AmberTools18, tleap was utilized to prepare the input files using the prepared structure with the AMBER ff14SB force field<sup>86</sup>, while disulfide bonds were manually assigned where pertinent. Each system was padded with 15 Å of TIP3P water on all sides and neutralized with an apparent number of either Na<sup>+</sup> or Cl<sup>-</sup> ions (Table S6).

The GPU implementation of PMEMD from AMBER18 was employed for conventional molecular dynamics simulations<sup>87</sup>. An initial round of minimization was performed on the systems with a 10 kcal/mol·Å<sup>2</sup> harmonic

restraint on the protein, which consisted of 2500 steps of steepest descent followed by 2500 steps of the conjugate gradient. The minimization was repeated a second time but without introducing the harmonic restraint. Subsequently, the systems were gradually heated in the NVT ensemble from 0 to 100 K for 12.5 ps, then from 100 to 310.15 K in the NPT ensemble for 125 ps at 1 bar; a 1 fs time step was used, along with a 10 kcal/mol-Å<sup>2</sup> harmonic restraint on the protein and ligand. Upon reaching the target temperature, the systems were equilibrated in the NPT ensemble at 310.15 K and 1 bar with a 2 fs time step. Starting with a 5 kcal/mol-Å<sup>2</sup> harmonic restraint on the protein and ligand, the restraint was diminished by 1 kcal/mol-Å<sup>2</sup> every 500 ps for a total of 2.5 ns, then by 0.1 kcal/mol-Å<sup>2</sup> every 500 ps for a total of 5 ns. Production simulations were run at 310.15 K (37 °C for comparison with biological studies) and 1 bar in the NPT ensemble for 100 ns with a 2 fs time step. All bond lengths involving hydrogen were constrained with the SHAKE algorithm. Particle Mesh Ewald summation was used for long-range electrostatics, while non-bonded interactions were cut off at 9.0 Å. Periodic boundary conditions were applied to all heating, equilibration, and production runs. The Langevin thermostat and Monte Carlo barostat were used where applicable. Three independent production runs were carried out for each of the systems using randomized velocities. Individual runs were performed on a single Nvidia GeForce RTX 2080 Ti GPU on a local computer cluster.

Custom CPPTRAJ scripts were used to process all trajectories of production simulations. All components were imaged to wrap all contents back to the periodic cell, and all waters and ions were stripped for analysis. Structure alignments were performed in reference to the docked proteins using the backbone atoms (C, Ca, N). Root mean square deviations of both the proteins and drugs were calculated following protein structure alignment.

**Data visualization and diagram construction.** Diagrams were created in Microsoft PowerPoint. Graphs were generated using the R statistical programming language with the open-source data visualization package ggplot2 (version 3.3.3). Molecular visualization images for protein–ligand complexes were produced using PyMOL 2.5 (Schrodinger) or MOE version 2019.0101.

Received: 17 August 2021; Accepted: 2 March 2022

Published online: 29 March 2022

## References

1. Pushpakom, S. *et al.* Drug repurposing: progress, challenges and recommendations. *Nat Rev Drug Discov* **18**, 41–58 (2019).
2. Tomazini, B. M. *et al.* Effect of dexamethasone on days alive and ventilator-free in patients with moderate or severe acute respiratory distress syndrome and COVID-19: the CoDEX randomized clinical trial. *JAMA* **324**, 1307–1316 (2020).
3. Grein, J. *et al.* Compassionate use of remdesivir for patients with severe Covid-19. *N. Engl. J. Med.* **382**, 2327–2336 (2020).
4. Cao, B. *et al.* A trial of lopinavir–ritonavir in adults hospitalized with severe Covid-19. *N. Engl. J. Med.* **382**, 1787–1799 (2020).
5. Geleris, J. *et al.* Observational study of hydroxychloroquine in hospitalized patients with Covid-19. *N. Engl. J. Med.* **382**, 2411–2418 (2020).
6. Galindez, G. *et al.* Lessons from the COVID-19 pandemic for advancing computational drug repurposing. *Nat. Comput. Sci.* **1**, 33–41 (2021).
7. Pushpakom, S. *et al.* Drug repurposing: progress, challenges and recommendations. *Nat. Rev. Drug Discov.* **18**, 41–58 (2019).
8. Singh, T. U. *et al.* Drug repurposing approach to fight COVID-19. *Pharmacol. Rep.* **72**, 1479–1508 (2020).
9. Trivedi, J., Mohan, M. & Byrareddy, S. N. Drug repurposing approaches to combating viral infections. *J. Clin. Med.* **9**, 3777 (2020).
10. Mirabelli, C. *et al.* Morphological cell profiling of SARS-CoV-2 infection identifies drug repurposing candidates for COVID-19. *Proc. Natl. Acad. Sci. USA* **118**, e2105815118 (2021).
11. Riva, L. *et al.* Discovery of SARS-CoV-2 antiviral drugs through large-scale compound repurposing. *Nature* **586**, 113–119 (2020).
12. Parvathaneni, V. & Gupta, V. Utilizing drug repurposing against COVID-19 - Efficacy, limitations, and challenges. *Life Sci.* **259**, 118275 (2020).
13. Bellera, C. L. *et al.* Can drug repurposing strategies be the solution to the COVID-19 crisis?. *Exp. Opin. Drug Discov.* **16**, 605–612 (2021).
14. Pan, H. *et al.* Repurposed antiviral drugs for COVID-19; interim WHO SOLIDARITY trial results. *N. Engl. J. Med.* **384**, 497–511 (2021).
15. Wang, L. Y. *et al.* Remdesivir and COVID-19. *Lancet* **396**, 953–954 (2020).
16. Cusinato, J., Cau, Y., Calvani, A. M. & Mori, M. Repurposing drugs for the management of COVID-19. *Exp. Opin. Ther. Pat.* **31**, 295–307 (2021).
17. Dinesh, D. C. *et al.* Structural basis of RNA recognition by the SARS-CoV-2 nucleocapsid phosphoprotein. *PLoS Pathog* **16**, e1009100. <https://doi.org/10.1371/journal.ppat.1009100> (2020).
18. Ye, Q., West, A. M. V., Silletti, S. & Corbett, K. D. Architecture and self-assembly of the SARS-CoV-2 nucleocapsid protein. *Protein Sci.* **29**, 1890–1901 (2020).
19. Kang, S. *et al.* Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites. *Acta Pharm. Sin. B* **10**, 1228–1238 (2020).
20. Smith, M. & Smith, J. C. Repurposing therapeutics for COVID-19: supercomputer-based docking to the SARS-CoV-2 viral spike protein and viral spike protein-human ACE2 interface. *ChemRxiv* <https://doi.org/10.26434/chemrxiv.11871402.v4> (2020).
21. Hartenian, E. *et al.* The molecular virology of coronaviruses. *J. Biol. Chem.* **295**, 12910–12934 (2020).
22. Martin, W. R. & Cheng, F. Repurposing of FDA-Approved Toremifene to Treat COVID-19 by blocking the spike glycoprotein and NSP14 of SARS-CoV-2. *J. Proteome Res.* **19**, 4670–4677 (2020).
23. Shah, B., Modi, P. & Sagar, S. R. In silico studies on therapeutic agents for COVID-19: Drug repurposing approach. *Life Sci.* **252**, 117652 (2020).
24. Gil, C. *et al.* COVID-19: drug targets and potential treatments. *J. Med. Chem.* **63**, 12359–12386 (2020).
25. Kalathiya, U. *et al.* Highly conserved homotrimer cavity formed by the SARS-CoV-2 spike glycoprotein: a novel binding site. *J. Clin. Med.* **9**, 1473 (2020).
26. Alhammad, Y. M. O. *et al.* The SARS-CoV-2 conserved macrodomain is a highly efficient ADP-ribosylhydrolase enzyme. *bioRxiv* <https://doi.org/10.1101/2020.05.11.089375> (2020).

27. Alhammad, Y. M. O. *et al.* The SARS-CoV-2 conserved macrodomain is a mono-ADP-ribosylhydrolase. *J. Virol.* **95**, e01969–20. <https://doi.org/10.1128/JVI.01969-20> (2020).
28. Kim, Y. *et al.* Crystal structure of Nsp15 endoribonuclease NendoU from SARS-CoV-2. *Protein Sci.* **29**, 1596–1605 (2020).
29. Sorzano, C. O. S., Crisman, E., Carazo, J. M. & León, R. Multitarget virtual screening for drug repurposing in COVID19. *ChemRxiv* <https://doi.org/10.26434/chemrxiv.12652997.v2> (2020).
30. Zhou, G., Stewart, L., Reggiano, G. & DiMaio, F. Computational drug repurposing studies on SARS-CoV-2 protein targets. *ChemRxiv* <https://doi.org/10.26434/chemrxiv.12315437.v1> (2020).
31. Bhardwaj, V. K. *et al.* Identification of bioactive molecules from tea plant as SARS-CoV-2 main protease inhibitors. *J. Biomol. Struct. Dyn.* **39**, 3449–3458 (2021).
32. Singh, R., Bhardwaj, V. K., Das, P. & Purohit, R. A computational approach for rational discovery of inhibitors for non-structural protein 1 of SARS-CoV-2. *Comput. Biol. Med.* **135**, 104555 (2021).
33. Bhardwaj, V. K. *et al.* Bioactive molecules of tea as potential inhibitors for RNA-dependent RNA polymerase of SARS-CoV-2. *Front. Med. (Lausanne)* **8**, 684020 (2021).
34. Singh, R., Bhardwaj, V. K., Sharma, J., Kumar, D. & Purohit, R. Identification of potential plant bioactive as SARS-CoV-2 Spike protein and human ACE2 fusion inhibitors. *Comput. Biol. Med.* **136**, 104631 (2021).
35. Sharma, J. *et al.* An in-silico evaluation of different bioactive molecules of tea for their inhibition potency against non structural protein-15 of SARS-CoV-2. *Food Chem.* **346**, 128933 (2021).
36. Singh, R., Bhardwaj, V. K., Sharma, J., Purohit, R. & Kumar, S. In-silico evaluation of bioactive compounds from tea as potential SARS-CoV-2 nonstructural protein 16 inhibitors. *J. Tradit. Complement Med.* <https://doi.org/10.1016/j.jtcme.2021.05.005> (2021).
37. Pansar, T. & Poso, A. Binding affinity via docking: fact and fiction. *Molecules* **23**, 1899 (2018).
38. Yang, L., Luo, H., Chen, J., Xing, Q. & He, L. SePreSA: a server for the prediction of populations susceptible to serious adverse drug reactions implementing the methodology of a chemical–protein interactome. *Nucl. Acids Res.* **37**, W406–W412 (2009).
39. Allen, J. A. & Roth, B. L. Strategies to discover unexpected targets for drugs active at G protein-coupled receptors. *Ann. Rev. Pharmacol. Toxicol.* **51**, 117–144 (2011).
40. Molecular Operating Environment (MOE) (Chemical Computing Group Montreal, QC, Canada, 2019.01).
41. Soga, S., Shirai, H., Kobori, M. & Hirayama, N. Identification of the druggable concavity in homology models using the PLB index. *J. Chem. Inf. Model* **47**, 2287–2292 (2007).
42. Soga, S., Shirai, H., Kobori, M. & Hirayama, N. Use of amino acid composition to predict ligand-binding sites. *J. Chem. Inf. Model* **47**, 400–406 (2007).
43. Zhang, S. *et al.* Pyrimidone inhibitors targeting Chikungunya Virus nsP3 macrodomain by fragment-based drug design. *PLoS ONE* **16**, e0245013. <https://doi.org/10.1371/journal.pone.0245013> (2021).
44. McPherson, R. L. *et al.* ADP-ribosylhydrolase activity of Chikungunya virus macrodomain is critical for virus replication and virulence. *Proc. Natl. Acad. Sci. U S A* **114**, 1666–1671 (2017).
45. Brooke, G. N. & Prisch, F. Structural and functional modelling of SARS-CoV-2 entry in animal models. *Sci. Rep.* **10**, 15917 (2020).
46. Chakraborty, H. & Bhattacharjya, S. Mechanistic insights of host cell fusion of SARS-CoV-1 and SARS-CoV-2 from atomic resolution structure and membrane dynamics. *Biophys. Chem.* **265**, 106438. <https://doi.org/10.1016/j.bpc.2020.106438> (2020).
47. Rouillard, A. D. *et al.* The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. *Database (Oxford)* <https://doi.org/10.1093/database/baw100> (2016).
48. Toelzer, C. *et al.* Free fatty acid binding pocket in the locked structure of SARS-CoV-2 spike protein. *Science* **370**, 725–730 (2020).
49. Scior, T. *et al.* Recognizing pitfalls in virtual screening: a critical review. *J. Chem. Inf. Model* **52**, 867–881 (2012).
50. Kubota, K., Funabashi, M. & Ogura, Y. Target deconvolution from phenotype-based drug discovery by using chemical proteomics approaches. *Biochim. Biophys. Acta Proteins Proteom.* **1867**, 22–27 (2019).
51. Deng, X. *et al.* Coronavirus nonstructural protein 15 mediates evasion of dsRNA sensors and limits apoptosis in macrophages. *Proc. Natl. Acad. Sci.* **114**, E4251–E4260 (2017).
52. Liu, X. *et al.* Porcine deltacoronavirus nsp15 antagonizes interferon- $\beta$  production independently of its endoribonuclease activity. *Mol. Immunol.* **114**, 100–107 (2019).
53. Lin, S. M. *et al.* Structure-based stabilization of non-native protein-protein interactions of coronavirus nucleocapsid proteins in antiviral drug design. *J. Med. Chem.* **63**, 3131–3141 (2020).
54. Yamamoto, N., Matsuyama, S., Hoshino, T. & Yamamoto, N. Nelfinavir inhibits replication of severe acute respiratory syndrome coronavirus 2 in vitro. *BioRxiv* <https://doi.org/10.1101/2020.04.06.026476> (2020).
55. Kuleshov, M. V. *et al.* The COVID-19 drug and gene set library. *Patterns (N. Y.)* **1**, 100090. <https://doi.org/10.1016/j.patter.2020.100090> (2020).
56. Meltzer, D. O. *et al.* Association of vitamin D status and other clinical characteristics with COVID-19 test results. *JAMA Netw. Open* **3**, e2019722–e2019722 (2020).
57. Teymoori-Rad, M., Shokri, F., Salimi, V. & Marashi, S. M. The interplay between vitamin D and viral infections. *Rev. Med. Virol.* **29**, e2032 (2019).
58. Günther, S. *et al.* X-ray screening identifies active site and allosteric inhibitors of SARS-CoV-2 main protease. *Science* **372**, 642–646 (2021).
59. Almeida, M. S., Johnson, M. A., Herrmann, T., Gerecht, M. & Wüthrich, K. Novel beta-barrel fold in the nuclear magnetic resonance structure of the replicase nonstructural protein 1 from the severe acute respiratory syndrome coronavirus. *J. Virol.* **81**, 3151–3161 (2007).
60. Serrano, P. *et al.* Nuclear magnetic resonance structure of the N-terminal domain of nonstructural protein 3 from the severe acute respiratory syndrome coronavirus. *J. Virol.* **81**, 12049–12060 (2007).
61. Michalska, K. *et al.* Crystal structures of SARS-CoV-2 ADP-ribose phosphatase: from the apo form to ligand complexes. *IUCr* **7**, 814–824 (2020).
62. Tan, J. *et al.* The SARS-unique domain (SUD) of SARS coronavirus contains two macrodomains that bind G-quadruplexes. *PLoS Pathog.* **5**, e1000428. <https://doi.org/10.1371/journal.ppat.1000428> (2009).
63. Ghosh, A. K. *et al.* Severe acute respiratory syndrome coronavirus papain-like novel protease inhibitors: design, synthesis, protein-ligand X-ray structure and biological evaluation. *J. Med. Chem.* **53**, 4968–4979 (2010).
64. Báez-Santos, Y. M. *et al.* X-ray structural and biological evaluation of a series of potent and highly selective inhibitors of human coronavirus papain-like proteases. *J. Med. Chem.* **57**, 2393–2412 (2014).
65. Serrano, P. *et al.* Nuclear magnetic resonance structure of the nucleic acid-binding domain of severe acute respiratory syndrome coronavirus nonstructural protein 3. *J. Virol.* **83**, 12998–13008 (2009).
66. Doungamath, A. *et al.* Crystallographic and electrophilic fragment screening of the SARS-CoV-2 main protease. *Nat. Commun.* **11**, 5047 (2020).
67. Gao, Y. *et al.* Structure of the RNA-dependent RNA polymerase from COVID-19 virus. *Science* **368**, 779–782 (2020).
68. Rosas-Lemus, M. *et al.* High-resolution structures of the SARS-CoV-2 2<sup>nd</sup> domain. *Sci Signal* **13**, doi:<https://doi.org/10.1126/scisignal.abe1202> (2020).
69. Jia, Z. *et al.* Delicate structural coordination of the severe acute respiratory syndrome coronavirus Nsp13 upon ATP hydrolysis. *Nucl. Acids Res.* **47**, 6538–6550 (2019).

70. Ma, Y. *et al.* Structural basis and functional analysis of the SARS coronavirus nsp14–nsp10 complex. *Proc. Natl. Acad. Sci. U S A* **112**, 9436–9441 (2015).
71. Walls, A. C. *et al.* Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **181**, 281–292 (2020).
72. Shang, J. *et al.* Cell entry mechanisms of SARS-CoV-2. *Proc Natl. Acad. Sci. U S A* **117**, 11727–11734 (2020).
73. Shang, J. *et al.* Structural basis of receptor recognition by SARS-CoV-2. *Nature* **581**, 221–224 (2020).
74. Surya, W., Li, Y. & Torres, J. Structural model of the SARS coronavirus E channel in LMPG micelles. *Biochim. Biophys. Acta Biomembr.* **1860**, 1309–1317 (2018).
75. Chen, C. Y. *et al.* Structure of the SARS coronavirus nucleocapsid protein RNA-binding dimerization domain suggests a mechanism for helical packaging of viral RNA. *J. Mol. Biol.* **368**, 1075–1086 (2007).
76. Chang, C., Michalska, K., Jedrzejczak, R., Maltseva, N., Endres, M., Godzik, A., Kim, Y., Joachimiak, A. Crystal structure of RNA binding domain of nucleocapsid phosphoprotein from SARS coronavirus 2, PDB ID 6VYO, <https://doi.org/10.2210/pdb6VYO/pdb> (2020).
77. Tan, K., Kim, Y., Jedrzejczak, R., Maltseva, N., Endres, M., Michalska, K., Joachimiak, A. The crystal structure of Nsp9 RNA binding protein of SARS CoV-2, PDB ID 6W4B, <https://doi.org/10.2210/pdb6W4B/pdb> (2020).
78. Owen, C. D., Lukacik, P., Strain-Damerell, C.M., Douangamath, A., Powell, A.J., Fearon, D., Brandao-Neto, J., Crawshaw, A.D., Aragao, D., Williams, M., Flaig, R., Hall, D.R., McAuley, K.E., Mazzorana, M., Stuart, D.I., von Delft, F., Walsh, M.A. SARS-CoV-2 main protease with unliganded active site (2019). PDB ID 6YB7, <https://doi.org/10.2210/pdb6YB7/pdb> (2020).
79. Osipiuk, J. *et al.* The crystal structure of papain-like protease of SARS CoV-2. *Nat. Commun.* **12**, 743 (2021).
80. Zhang, L. *et al.* Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved  $\alpha$ -ketoamide inhibitors. *Science* **368**, 409–412 (2020).
81. Wishart, D. S. *et al.* DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucl. Acids Res.* **46**, D1074–D1082 (2018).
82. Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W. & Taylor, R. D. Improved protein-ligand docking using GOLD. *Proteins* **52**, 609–623 (2003).
83. Reynolds, C. H., Bembenek, S. D. & Tounge, B. A. The role of molecular size in ligand efficiency. *Bioorg. Med. Chem. Lett.* **17**, 4258–4261 (2007).
84. Reynolds, C. H., Tounge, B. A. & Bembenek, S. D. Ligand binding efficiency: trends, physical basis, and implications. *J. Med. Chem.* **51**, 2432–2438 (2008).
85. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **25**, 1157–1174 (2004).
86. Maier, J. A. *et al.* ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **11**, 3696–3713 (2015).
87. Case, D. A. *et al.* AMBER 2018 (University of California, 2018).

## Acknowledgements

This work was supported by NIH Grants R37 DA39997, R01 DA035316 and T32 007268 to JRT.

## Author contributions

W.K.B.C. and K.M.O. performed the experiments. J.W.W. and J.Z.S. provided data on the phenotypic assays for inhibition of SARS-CoV-2 infection. W.K.B.C., K.M.O. and H.A.C. analyzed the data. W.K.B.C., K.M.O. and J.R.T. wrote the manuscript. J.R.T. and H.A.C. edited the manuscript. J.R.T. supervised the research. All authors reviewed and approved the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-08320-y>.

**Correspondence** and requests for materials should be addressed to J.R.T.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022