High throughput virtual screening to discover inhibitors of the main protease of the coronavirus SARS-CoV-2

Olujide O. Olubiyi[1,2*], Maryam Olagunju[1], Monika Keutmann[1], Jennifer Loschwitz[1,3], and Birgit Strodel[1,3*]

1 Institute of Biological Information Processing: Structural Biochemistry, Forschungszentrum Jülich, Jülich, Germany
2 Department of Pharmaceutical Chemistry, Faculty of Pharmacy, Obafemi Awolowo University, Ile-Ife, Nigeria
3 Institute of Theoretical and Computational Chemistry, Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany

* Corresponding authors: olubiyioo@oauife.edu.ng, b.strodel@fz-juelich.de

ABSTRACT

We use state-of-the-art computer-aided drug design (CADD) techniques to identify prospective inhibitors of the main protease enzyme, 3CL^{pro} of the SARS-CoV-2 virus causing COVID-19. From our screening of over one million compounds including approved drugs, investigational drugs, natural products, and organic compounds, and a rescreening protocol incorporating enzyme dynamics via ensemble docking, we have been able to identify a range of prospective 3CL^{pro} inhibitors. Importantly, some of the identified compounds had previously been reported to exhibit inhibitory activities against the 3CL^{pro} enzyme of the closely related SARS-CoV virus. The top-ranking compounds are characterized by the presence of multiple bi- and monocyclic rings, many of them being heterocycles and aromatic, which are flexibly linked allowing the ligands to adapt to the geometry of the 3CL^{pro} substrate site and involve a high amount of functional groups enabling hydrogen bond formation with surrounding amino acid residues, including the catalytic dyad residues H41 and C145. Among the top binding compounds we identified several tyrosine kinase inhibitors, which include a bioflavonoid, the group of natural products that binds best to 3CL^{pro}. Another class of compounds that decently binds to the SARS-CoV-2 main protease are steroid hormones, which thus may be endogenous inhibitors and might provide an explanation for the age-dependent severity of COVID-19. Many of the compounds identified by our work show a considerably stronger binding than found for reference compounds with *in vitro* demonstrated 3CL^{pro} inhibition and anticoronavirus activity. The compounds determined in this work thus represent a good starting point for the design of inhibitors of SARS-CoV-2 replication.

Keywords COVID-19 · docking · drug repurposing · natural products · in silico drug design · viral replication inhibition

1   Introduction

The recently identified COVID-19 (coronavirus disease 2019) causing virus is a coronavirus belonging to a much diverse group of the family enveloped RNA viruses of *Coronaviridae* [1, 2]. Coronaviruses have been reported in different animal hosts and have been implicated in various respiratory and enteric infections of epidemic and pandemic proportion [1, 3, 4]. One of them, the SARS-CoV, was identified as the cause of the 2003 severe acute respiratory syndrome (SARS), an epidemic of pneumonia that resulted in more than 800 deaths worldwide [5]. In 2013, another member of the coronavirus group was found being responsible for the Middle East respiratory syndrome coronavirus (MERS-CoV), an infection characterized by acute pneumonia and renal failure and with a fifty percent mortality rate recorded in admitted patients [6, 7]. HCoV-229E, HCoV-OC43, HCoV-NL63, and HCoV-HKU1 are other identified human coronaviruses whose effect on the respiratory system results in milder forms of common colds [8, 9].

In late 2019, a previously unknown member of the *Coronaviridae* family was identified and implicated in a global epidemic of respiratory systems. On March 11th 2020, the World Health Organization (WHO) declared the outbreak a pandemic. As of May 28th 2020, there are almost 6 million confirmed cases globally [10], and the infection fatality rate is reported to be around 0.4 [11]. The virus causing COVID-19 has been named SARS-CoV-2, because its RNA genome is about 82% identical to SARS-CoV [12]. Upon infection, COVID-19 affects first the upper respiratory tract with symptoms ranging from dry non-productive cough to sore throat and fever. Subsequently the lower trees of the respiratory tract are affected. However, the illness can also cause malaise, confusion, dizziness, headaches, digestive issues, and a loss of smell and taste. It has been suggested that these neurological signs may result from the ability of the virus to invade the central nervous system [13]. With its highly effective mode of transmission, COVID-19, in spite of its relatively low fatality rate [11], represents one of the greatest public health challenges in recent times.

Unfortunately, there are currently no antiviral drugs or vaccines approved for COVID-19 or any other human coronavirus infections [9]. The genome of SARS-CoV-2 encodes for different proteins, including the 3-chymotrypsin-like protease (3CL^{pro}), also called main protease (M^{pro}), papain-like protease, helicase, and RNA-dependent RNA polymerase [14, 15]. Since the main protease 3CL^{pro} is crucial for viral replication and well conserved across the *Coronaviridae* family, it represents a viable target for drug design [12]. 3CL^{pro} cleaves the large polyprotein 1ab (replicase 1ab, $\sim$790 kDa) at eleven or even more cleavage sites involving, in most cases, the recognition sequence L-N*(S,A,G) (* marks the cleavage site), yielding functional proteins that are then packed into

the virion. Another advantage of targeting 3CL^{pro} is that although the mutagenesis rate is high in viruses, this does not apply to this protein since any mutation here can be fatal for the virus. Furthermore, since no human proteases with a similar cleavage specificity are known, it should be possible to identify inhibitors of no or low toxicity.

Since the outbreak, several SARS-CoV-2 protein structures have been solved using either X-ray diffraction or cryo-electron microscopy. One of these structures, employed in this work, is the crystal structure of the SARS-CoV-2 3CL^{pro} enzyme in complex with a synthetic peptidomimetic inhibitor called N3 (PDB code 6LU7, Figure 1) [16]. While the catalytically active form of 3CL^{pro} is a dimer, the two protomers most likely act independently from each other as the two active sites are solvent-exposed and symmetrically located at opposite edges of the cleft between the two protomers [17]. The proteolytic process in the active site of 3CL^{pro} is enabled by the catalytic C145-H41 dyad with the cysteine thiol group acting as the nucleophile (Figure 1C) [18]. The initial drug discovery efforts after the SARS outbreak in 2003 aimed at electrophilic attack to the cysteine residue of the catalytic dyad via covalent Michael inhibitors [19]. While this was considered to be safe due to the different proteolytic cleavage specificities between SARS-CoV and human proteases, electrophiles are usually no good drug candidates as they often cause adverse effects such as allergies, tissue destruction, or carcinogenesis [20]. After 2005–2006, many of the initial efforts of developing small-molecule compounds with anticoronavirus activity were discontinued due to a sharp decline in funding of coronavirus research as it was erroneously assumed that another zoonotic coronavirus transmission was extremely unlikely to happen again. Thus, none of these attempts resulted in an anticoronavirus drug, not even the clinical stage was reached.

This drastically changed after the COVID-19 outbreak. Less than four months after the first cases were reported in Wuhan in China, several studies aiming at designing and developing treatment for the disease have already been published [16, 21–27]. In one of these studies, the mechanism-based inhibitor N3 was designed (Figure 1) using computer-aided drug design (CADD) techniques [16]. Seven further compounds were identified in that study through a combination of structure-based virtual and high-throughput screening of over 10,000 compounds, including approved drugs, drug candidates in clinical trials, and other pharmacologically active compounds, as inhibitors of 3CL^{pro} with half maximal inhibitory concentrations (IC$_{50}$) ranging from 0.67 to 21.4 μM. The strongest antiviral activity in cell-based assays was found for ebselen, a synthetic organoselenium drug molecule with anti-inflammatory, anti-oxidant and cytoprotective activity which is currently in a clinical trial as a potential treatment for bipolar disorder [28].

Here, we use high-throughput virtual screening to discover potential inhibitors of 3CL^{pro} of SARS-CoV-2. We screened over 1 million compounds covering multiple compound libraries most of which are available through the ZINC database [31–33], which is a curated collection of more
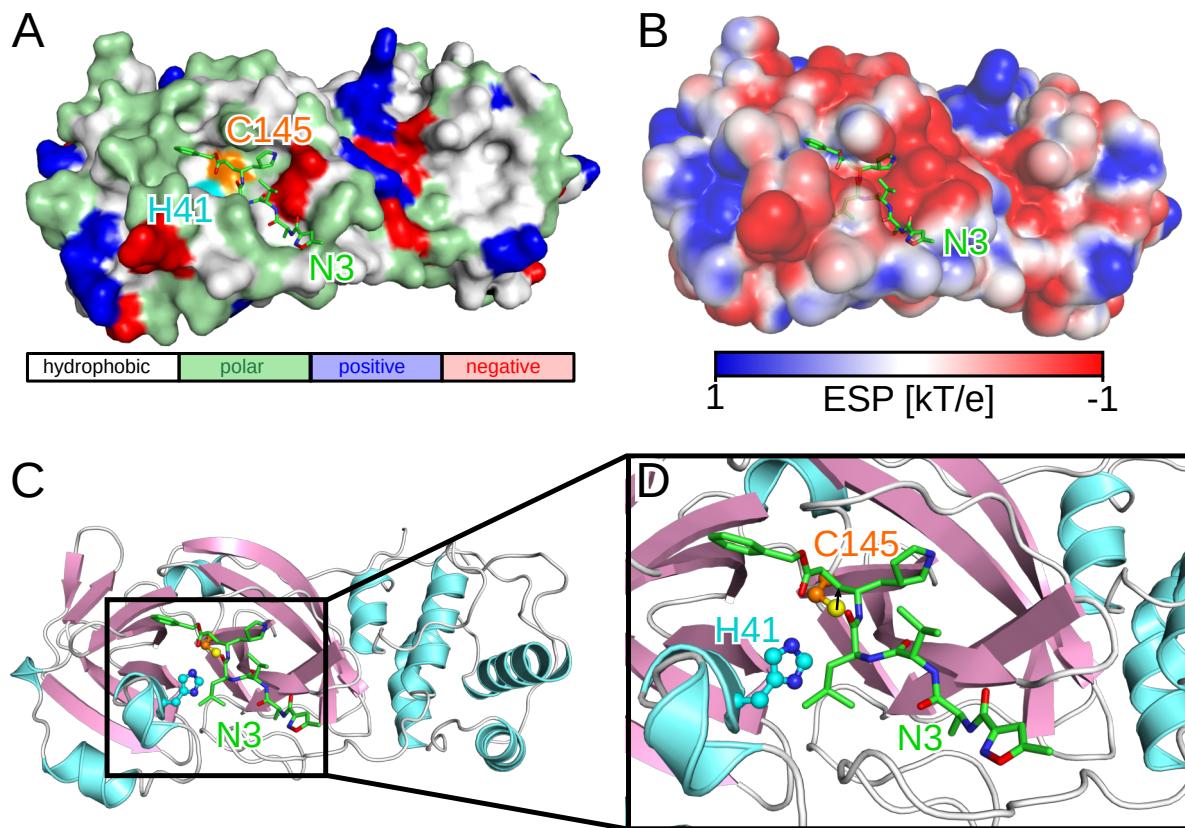
Figure 1: Structure and binding site of 3CL^{pro} including the ligand N3 (PDB code 6LU7).
A. The surface of 3CL^{pro} is shown and colored according to residue type using the color code
given at the bottom. The catalytic dyad residues H41 and C145 are colored in cyan and orange,
respectively. The stick representation is used for N3, which is colored in green with red and
blue for the O and N atoms, respectively. **B.** Another surface representation of 3CL^{pro} but now
colored according to the electrostatic potential using the color code given at the bottom (in kT with
k being the Boltzmann constant, created with the Adaptive Poisson–Boltzmann Solver (APBS)
software [29, 30]). **C.** Cartoon representation of 3CL^{pro} with β-sheets shown in lilac and α-helices
in light blue. The sidechains of H41 and C145 are shown in ball-and-stick representation in cyan
and orange, respectively, but using blue for N atoms, red for O atoms, and yellow for the S atom of
C145. **D.** Zoom into the substrate binding site of 3CL^{pro}. In panels B–D the same representation for
N3 as in panel A is used.

than 230 million commercially available chemical compounds prepared for virtual screening. The
screened compounds include drugs approved by the U.S. Food and Drug Administration (FDA) or
other authorities, investigational drugs, natural products, as well as organic compounds that do not
necessarily fall into the other mentioned libraries [34, 35]. In addition, we also scanned our in-house
database with almost 3,200 natural compounds isolated from African plants. We employed a virtual
screening against the crystallographic 3CL^{pro} structure as well as ensembles of the SARS-CoV-2
enzyme generated from an explicit-solvent molecular dynamics (MD) simulation. We present here

our analyses and important findings relating to the inhibition of 3CL^{pro}, which we believe can guide the search for a COVID-19 treatment.

2 Results and Discussion

In this work, we have employed CADD approaches in search of compounds that are capable of forming thermodynamically feasible binary complexes with the SARS-CoV-2 main protease enzyme 3CL^{pro}. We adopted a protocol that is slightly different from previous works conducted by other groups in the search for potential inhibitors of this critical enzyme [16, 21]. First, we have not limited our virtual screening to previously known inhibitors of viral proteases but instead embarked on what would count as an unbiased screening of a fairly large compound library. We believe this approach reduces the possibility of omitting potentially useful compounds, especially considering that the known inhibitors were tested on different viral species. Second, our screening procedure incorporated protein dynamics in order to model as accurately as possible the conformational preferences available *in vitro* and *in vivo* for substrate recognition. The immediate objective of our research work is the identification of small-molecule inhibitors of 3CL^{pro}, and in particular of approved drugs which should reduce the amount of time to clinical readiness. This also holds true for natural products present in plants that are available as herbal, nutraceutical or food products.

Another objective is to determine the structural basis of inhibitor interaction with 3CL^{pro} and identify chemical structures capable of serving as templates for the design of potent inhibitors. We expect that the chances of finding suitable inhibitors will be enhanced by screening chemically diverse libraries. To this end, we screened a library of 1,227,186 compounds derived from the ZINC library [31–33] as well as our in-house library of natural products of Nigerian origin (unpublished) against the crystallographic structure of the functional form of the SARS-CoV-2 3CL^{pro} enzyme (PDB code 6LU7 [16]). We also included in our screening eight recently reported inhibitors of the SARS-CoV2 3CL^{pro} enzmyme, which are the peptidomimetic N3 present in the crystal structure used in this work as well as ebselen, disulfiram, carmofur, cinanserin, shikonin, tideglusib, and PX-12 reported alongside this crystal structure [16]. All eight reference compounds were able to bind to the active site of the crystal structure of the enzyme. The best binding for these reference compounds was found for tideglusib with a binding free energy value (ΔG) of -6.64 kcal/mol, while PX-12, disulfiram and shikonin were indicated in our calculations as possessing only weak binding affinities with a ΔG values above -4.0 kcal/mol. Interestingly, the experimental IC$_{50}$ values obtained for these compounds also suggested them to be the least binding out of the eight experimentally validated 3CL^{pro} inhibitors [16].

In the sections that follow, we discuss the 3CL^{pro} substrate-site binding of the best predictions. In order to understand the 3CL^{pro}–compound interactions, we plotted the protein–ligand interactions

using LigPlot+ [36,37]. While all front runner compounds exhibit strong favorable interactions with the substrate binding site of 3CL$^{\text{pro}}$, we have decided to accord special attention to those compounds that additionally interact with at least one of the two catalytic dyad residues, as well as to certain other compounds identified in the present work, such as steroid hormones that are of special interest in the ongoing global epidemic. The binding site because of its peculiar topology enables the binding of various chemical groups at different subsites within the rather large accommodating substrate site. We expect that the overall usefulness and predictive power of computational efforts should factor both the test compounds ability to bind strongly within the active site (indicated by the computed ΔG) as well as the ability to make specific contacts with amino acids within the active site, with special emphasis given to the catalytic dyad residues H41 and C145 (indicated by the computed distance to the 3CL$^{\text{pro}}$ catalytic dyad, d_{dyad}).

2.1 Screening of the synthetic compounds library

We split the entire data set of 1,227,186 screened compounds into screening group A (sGrA) and screening group B (sGrB). The set sGrA is made up of 1,068,161 synthetic compounds that we first screened against the crystallographic structure of the 3CL$^{\text{pro}}$ enzyme. Following this screening, we exerted a purely energetic cutoff (ΔG of -8.0 kcal/mol) to select the most promising ligands, yielding a total of 9,515 synthetic compounds from the sGrA library. The choice of the cutoff was to improve the chances of attaining compounds with binding strengths significantly superior to those obtained for the reference inhibitors, for which the best binding was found for tideglusib with a ΔG of -6.64 kcal/mol.

We then analyzed the physicochemical properties of the 9,515 virtual hits. They belong to several distinct chemical groups with molecular weights (MW) ranging from 200 to about 1,000 g/mol. Since the intention at this time is to identify any compounds with the potential to bind to the active site of 3CL$^{\text{pro}}$ regardless of their pharmacokinetic attributes, we have decided to deprioritize filtering based on physicochemical attributes usually employed in predicting compatibility with the oral route of administration. We analyzed the dominant ligand chemical fragments to find out if any of them featured more disproportionately than others. Figure S1 shows that the preponderant chemical fragments are mono- and bicyclic rings, many of them being heterocycles, which are chemical motifs generally associated with known drug molecules.

To account for the influence of protein flexibility in ligand binding, we performed a 100 ns MD simulation of the 3CL$^{\text{pro}}$-N3 complex in water and clustered the active site conformations sampled during that simulation. The five most dominant conformations after removing the bound N3 were then employed for a rescreening of the 9,515 compounds selected from the initial screening against the protein crystal structure. The ΔG values and distances between the bound ligand and the
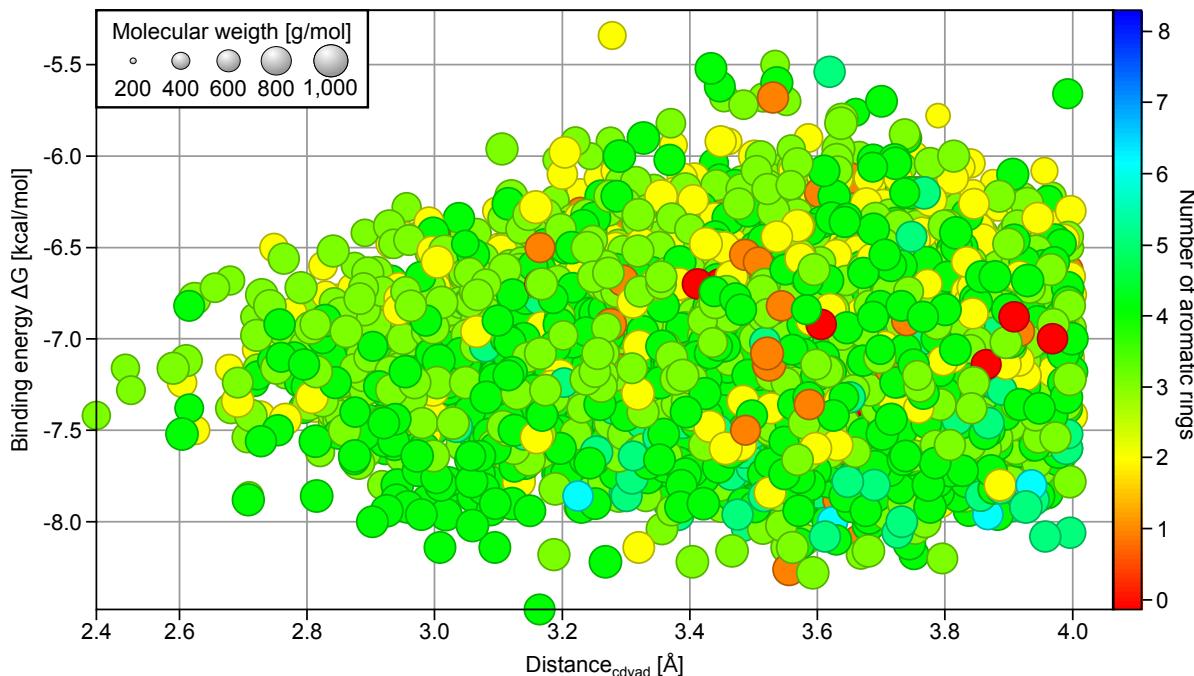
Figure 2: Distribution of the best binding synthetic compounds in terms of d_{dyad} (x-axis), ΔG (y-axis), number of aromatic rings (color), and molecular weight (circle size). The ΔG and d_{dyad} values are averages obtained from ensemble docking. Only compounds with $\Delta G < -5.0$ kcal/mol and $d_{dyad} \leq 4$ Å are shown.

3CL^{pro} catalytic dyad, d_{dyad}, obtained for each compound were averaged over the five protease conformations. The resulting two values per compound were then employed in ranking the screened library and selecting the best compounds for further analysis. Introduction of receptor dynamics was observed to reduce the ability of many of the compounds to strongly interact with 3CL^{pro}. While the docking against the crystal structure did not allow to discriminate between good and bad binders based on MW, after ensemble docking no compound with a MW higher than 600 g/mol returned computed affinities equal to or less than -7.0 kcal/mol. Moreover, only a few such high MW compounds were found to interact with the catalytic dyad within a 3.5 Å binding distance. On the other hand, compounds with small MW also turned out to be poor binders, dominating the compounds with binding free energy values above -5.0 kcal/mol.

A 4D plot showing the relationship between ΔG, d_{dyad}, MW, and the number of aromatic rings present (Figure 2, only compounds with $\Delta G < -5.0$ kcal/mol and $d_{dyad} < 4$ Å are shown) reveals that compounds with fewer than two aromatic rings (orange and red) are seldomly found in the top binder subset ($\Delta G \leq -7.5$ kcal/mol); instead they are clustered around -6.5 kcal/mol. On the other hand, many of the compounds with more than four aromatic rings in their chemical structure (cyan and blue) have ΔG values below -7.5 kcal/mol. The number of aromatic rings also plays a role in determining the ability of a ligand to interact with H41 and C145 of the catalytic dyad,

where fewer than five rings appears to be advantageous using $d_{\mathrm{dyad}} \leq 3.2$ Å as cutoff. Analysis of the structural fragments present in the good protease binders with $\Delta G \leq -7.5$ kcal/mol and $d_{\mathrm{dyad}} \leq 3.5$ Å (Figure S2) reveals an increased preference for bicyclic aromatic ring structures. Most of them are heterocycles with the majority having nitrogen atoms as heteroatoms, followed by oxygen atoms. This suggests a structural selection by the SARS-CoV-2 3CL[pro] binding site for three or four aromatic, often hetero- and bicyclic rings and a MW between 400 and 600 g/mol.

To further analyze the binding of the ligands to the substrate site of 3CL[pro], we evaluated the six compounds with the strongest computed affinities ($\Delta G \leq -8.2$ kcal/mol) in detail. They are characterized by an essentially rigid core featuring polycyclic systems with three (compounds **1**, **4**, and **6** in Figure 3), four (compounds **2** and **5**), and seven (compound **3**) fused rings. The presence of these relatively large hydrophobic units encourages contacts with multiple amino acid residues within the 3CL[pro] substrate site. However, none of them directly binds to the catalytic dyad and the respective distances are all above 3 Å. The very low ΔG values recorded for them appears to result from majorly hydrophobic interactions, especially for compound **3** which is a pure hydrocarbon (no functionalization) and thus should serve no more than as a useful probe for exploring the limits of shape and size compatible with the substrate site. Comparison with the electrostatic potential (ESP) of 3CL[pro] in Figure 1B further supports the observation that for such compounds the binding is mainly driven by hydrophobic interactions since at the entry to the active site, where most of them are bound, the ESP is close to zero, whereas it is partly negative in the active site defined by the catalytic dyad.

To identify the compounds that most intimately interact with the catalytic dyad, all ligands with an average $\Delta G \leq -7.0$ kcal/mol were ranked based on distance from the catalytic dyad. The top six compounds that bind directly to the catalytic dyad (d_{dyad} ranging from 2.47 Å to 2.60 Å) are shown in Figure 4. Comparing these compounds with those with the highest calculated affinities, three structural patterns immediately emerge. First, the direct catalytic dyad binders feature smaller fused ring systems with a maximum of three fused rings (compounds **8** and **10** in Figure 4) and the majority having two fused rings (compounds **7**, **9**, **10**, and **12**). Second, all six compounds in Figure 4 have flexible substructures that allow them to adopt to the conformation needed to interact with the catalytic H41 and C145. In compounds **7**, **8**, and **9**, the superimposition of the three bound compounds (not shown) revealed a perfect alignment of their urea groups which makes it possible to establish hydrogen-bonding contacts with the catalytic dyad. Third, five of the six ligands feature an amide group and all of them have at least one NH group, allowing for hydrogen bonding with the substrate site of 3CL[pro].

The knowledge of the binding poses adopted by the rigid binders in Figure 3 and the flexible binders in Figure 4 along with the analysis of the physicochemical properties of good binders in

1

(-8.45 kcal/mol, 3.16 Å)

2

(-8.28 kcal/mol, 3.59 Å)

3

(-8.26 kcal/mol, 3.56 Å)

4

(-8.22 kcal/mol, 3.27 Å)

5

(-8.22 kcal/mol, 3.40 Å)

6

(-8.20 kcal/mol, 3.80 Å)
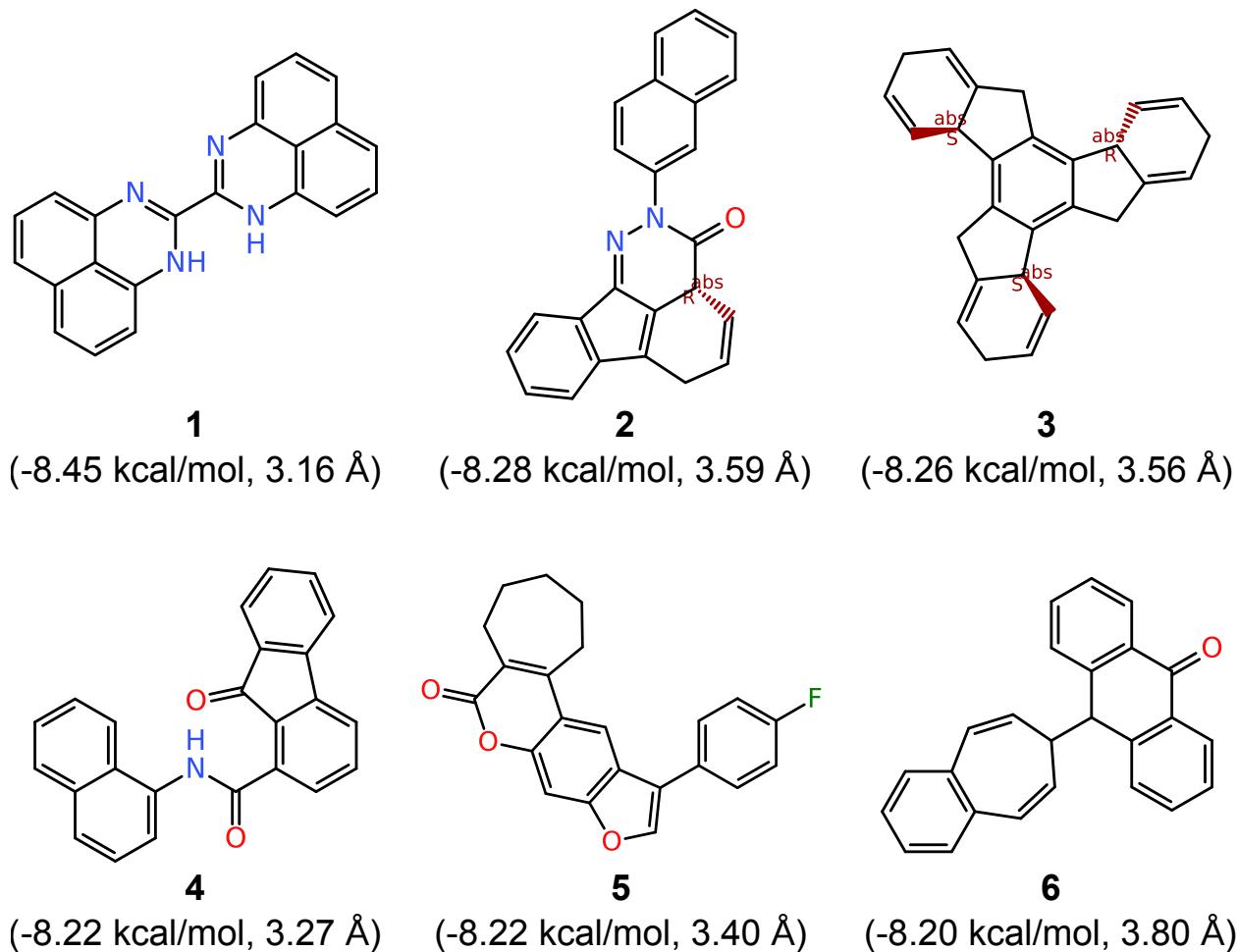
Figure 3: Synthetic compounds with the highest computed affinities for the substrate site of the SARS-CoV-2 3CL$^{\text{pro}}$ enzyme obtained from ensemble docking. The average ΔG and d_{dyad} values are given in parantheses.

Figure 2 and the dissection of their chemical fragments in Figure S2 can help facilitate the design of 3CL$^{\text{pro}}$ specific inhibitors.

2.2   Interaction of approved and investigational drugs with 3CL$^{\text{pro}}$

A total of 10,090 drugs were screened against both the crystallographic structure and the five MD-generated conformations of the SARS-CoV2 3CL$^{\text{pro}}$ enzyme. The screened compounds comprised drugs approved by the United States of America's regulatory agency, FDA and those approved by other countries. Investigational drugs being trialed for various clinical indications were also screened. After the ensemble docking phase, we ranked the drug molecules based on their average ΔG and d_{dyad} values. The resulting top binding drug molecules are listed in Table S1, where FDA-approved and other drugs are separately listed. In the top ten based on cluster-

7
(-7.42 kcal/mol, 2.47 Å)

8
(-7.16 kcal/mol, 2.52 Å)

9
(-7.28 kcal/mol, 2.52 Å)

10
(-7.16 kcal/mol, 2.59 Å)

11
(-7.52 kcal/mol, 2.60 Å)

12
(-7.24 kcal/mol, 2.60 Å)
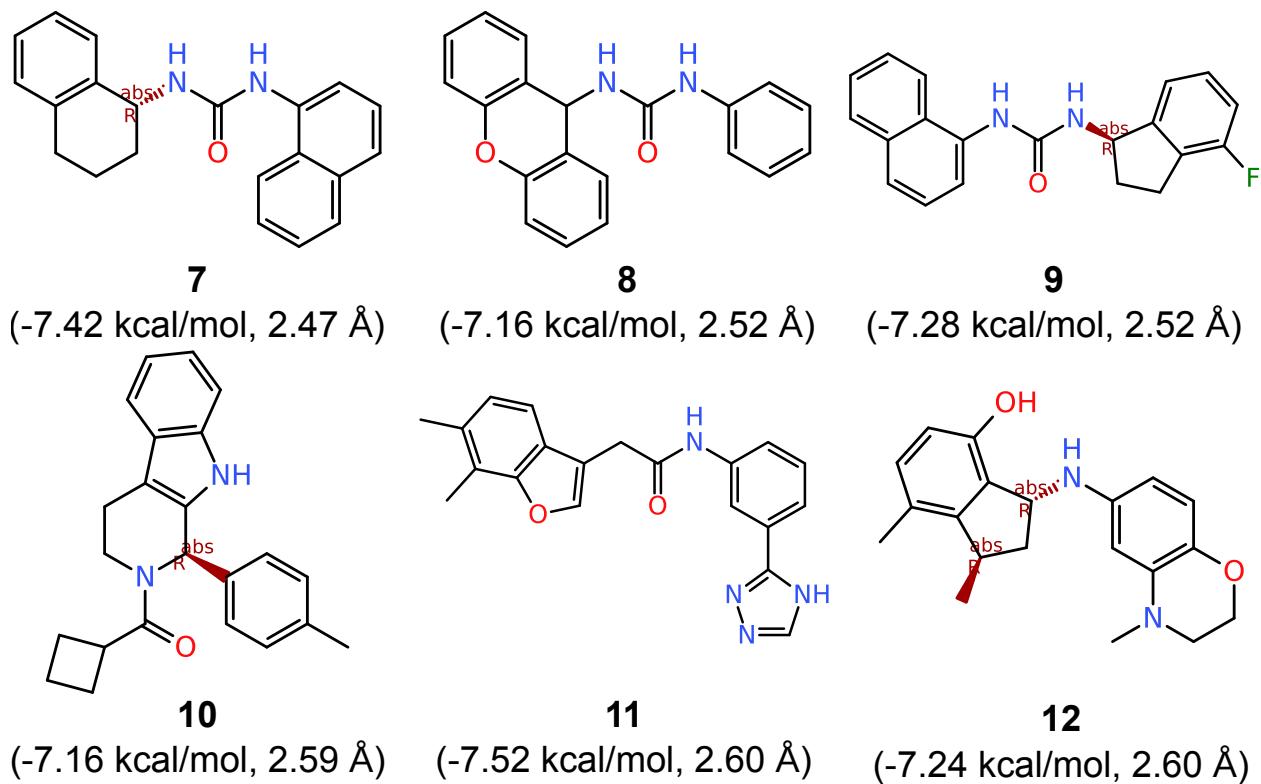
Figure 4: Synthetic compounds with the most intimate interaction with the catalytic dyad of the SARS-CoV-2 3CL$^{\text{pro}}$ enzyme obtained from ensemble docking. The average ΔG and d_{dyad} values are given in parantheses.

averaged ΔG values there is only one FDA-approved drug, which is nilotinib (Table S1, entry 1), a tyrosine kinase inhibitor, with $\Delta G = -8.66$ kcal/mol (Figure 5). Interestingly, the compound found to display the strongest 3CL$^{\text{pro}}$ binding with $\Delta G = -10.46$ kcal/mol is phthalocyanine (Table S1, entry 62), a non-toxic dye (IC$_{50}$ of 10 g/kg) with biomedical applications including use as photosensitizer for non-invasive cancer therapy. The planar macrocyclic tetrapyrrole structure of phthalocyanine, coupled with the relatively extensive nature of the substrate binding site allows the therapeutic dye adopt two alternate binding poses (Figure S3). The first arrangement involves a purely hydrophobic contact with C145, with the ligand's planar axis perpendicular to the H41–C145 axis. The alternate binding arrangement aligns the planar axis of the ligand parallel to the catalytic dyad axis. In this arrangement, the two polar hydrogen atoms at the center of the tetrapyrrole structure of phthalocyanine can form a hydrogen bond with H41. A similar binding pattern as for phthalocyanine was obtained for the second best binder, hypericin with $\Delta G = -9.12$ kcal/mol (Table S1, entry 63). Hypericin also possesses a rigid and planar naphthodianthrone structure and in its 3CL$^{\text{pro}}$ bound poses we observed it to alternate between a parallel arrangement enabling hydrophobic contact with the catalytic dyad and a perpendicular arrangement that makes hydrogen bond connection with H41 possible (Figure S3). It is of interest to point out that hypericin has

been reported to possess a various spectrum of antiviral properties including inhibition of the replication of a coronavirus, the infectious bronchitis virus as reported in December 2019 by a Chinese group [38]. Islam et al. employed computational methods directed at forty known antiviral phytochemicals – hypericin is one of the principal active constituents of Saint John's wort – and also identified hypericin as a good candidate for inhibiting the 3CL^{pro} enzyme of SARS-CoV-2 [39]. Other non-FDA drugs that outperform the best FDA-approved 3CL^{pro} ligand, i.e., nilotinib in terms of ΔG include adozelesin ($\Delta G = -8.84$ kcal/mol), telomestatin ($\Delta G = -8.80$ kcal/mol), MK-3207 ($\Delta G = -8.74$ kcal/mol), and radotinib ($\Delta G = -8.68$ kcal/mol) (Table S1, entries 64–67).
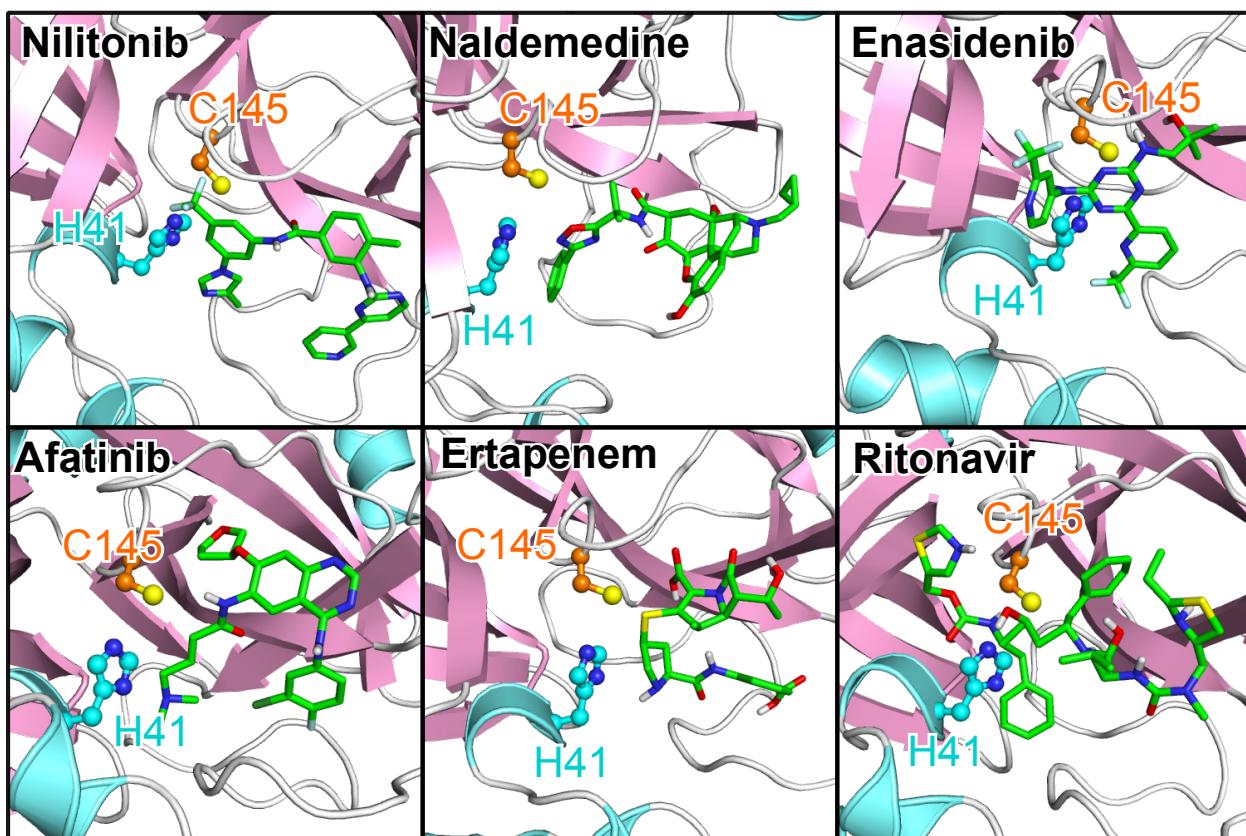
Figure 5: The poses of the five best FDA-approved drugs and ritonavir. The top five compounds bind well to the substrate site in terms of both ΔG and closeness to the catalytic dyad of 3CL^{pro}. The same protein and ligand representation as well as color scheme as in Figure 1 are used.

Of the top-performing FDA-approved drugs four that directly interact with the 3CL^{pro} catalytic dyad are of special interest (Figure 5), and these include the opioid receptor antagonist naldemedine ($\Delta G = -8.06$ kcal/mol, $d_{\mathrm{dyad}} = 2.98$ Å), enasidenib indicated for the treatment of acute myeloid leukemia ($\Delta G = -7.76$ kcal/mol, $d_{\mathrm{dyad}} = 2.89$ Å), afatinib, an orally active tyrosine kinase inhibitor used in the treatment of non-small cell lung carcinoma ($\Delta G = -7.44$ kcal/mol, $d_{\mathrm{dyad}} =$
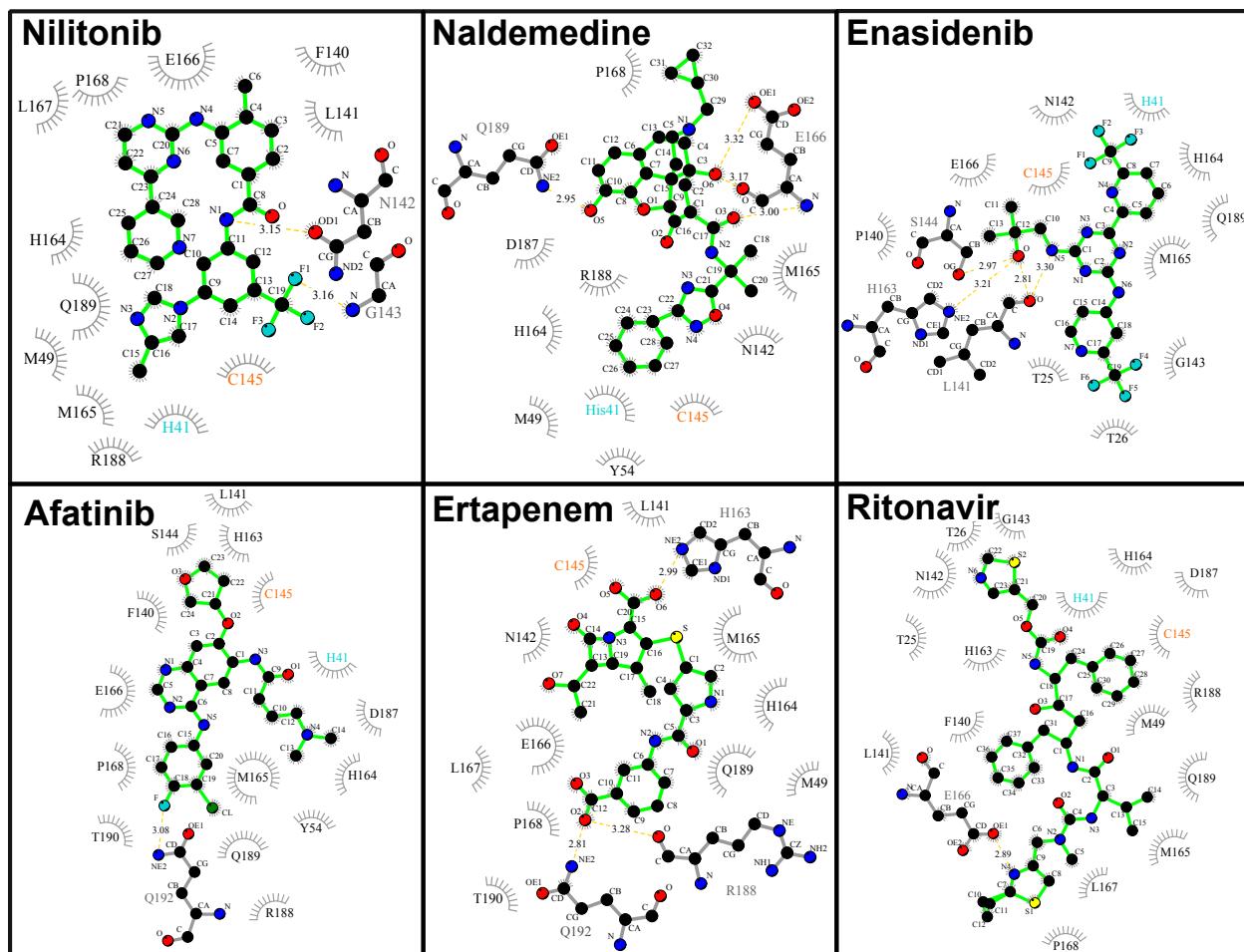
Figure 6: The 3CL^pro–compound interactions for the five best FDA-approved drugs and ritonavir. The interactions were analyzed and plotted with LigPlot+ [36, 37]. Hydrogen bonds are indicated by orange dashed lines between the atoms involved and the donor–acceptor distance is given in Å, while hydrophobic contacts are represented by gray arcs with spokes radiating towards the ligand atoms they contact. The contacted atoms are shown with spokes radiating back.

3.04 Å), and ertapenem, an antibiotic ($\Delta G = -7.24$ kcal/mol, $d_{\text{dyad}} = 2.95$ Å). Apart from these four and nilotinib with the lowest energy value, ritonavir ($\Delta G = -6.70$ kcal/mol, $d_{\text{dyad}} = 3.52$ Å), an antiviral agent currently being investigated for COVID-19 treatment [40], also desires special mention (Figure 5). Both nilotinib and afatinib are tyrosine kinase inhibitors, and interestingly this particular class of drugs has been shown to possess inhibitory activities against related coronaviruses [41]. Analysis of the interaction of the two tyrosine kinase inhibitors (Figure 6) shows effective catalytic dyad contacts between the trifluoromethyl group of nilotinib and C145, while its imidazole ring makes another contact with H41. While afatinib also makes contacts with both catalytic dyad residues, the fewer contacts it makes with other substrate site residues accounts for its slightly lower affinity. All the remaining four drugs presented in Figures 5 and 6 make vital contacts

with the two catalytic dyad residues with the exception of ertapenem which makes contact with only C145. Ritonavir as well as other agents like imatinib (entry 36 in Table S1) and sildernafil (entry 43 in Table S1) are currently going through clinical evaluation for the possible treatment of COVID-19 [42, 43].

The screened non-FDA approved drugs produced four drug compounds directly interacting with the catalytic dyad and these include UK-432,097, a selective adenosine A_{2A} receptor agonist with urea linker approved for the treatment of chronic obstructive pulmonary disease; R-343, an investigational agent with tyrosine kinase inhibitory activities; and amrubicin used in the treatment of lung cancer. The binding poses of these drug compounds in the substrate site of 3CL^{pro} and their interactions are shown in Figure S4. This figure also includes remdesivir, which is a broad-spectrum antiviral medication and is currently being tested as a specific treatment for COVID-19 [40].

2.3   Interaction of natural products with 3CL^{pro}

The natural products that were screened were made up of the over 140,000 natural products retrieved from the ZINC library as well as over 3,100 natural products of Nigerian plants origin. For consistent referencing we have quoted ZINC database [31–33] or PubChem (`https://pubchem.ncbi.nlm.nih.gov/`) references for all reported compounds. The natural products were screened against both the SARS-CoV-2 3CL^{pro} crystal structure and the MD-generated conformations of the enzyme. In Table S1 we have presented a total of 38 natural products demonstrating strong binding with both the crystal structure and the five ensemble structures. A significant percentage of the identified natural products is represented by flavonoids, some of which had previously been reported as possessing inhibitory activities against 3CL^{pro} of other coronavirus species. Figure 7 shows the binding poses of the top natural products in terms of both ΔG and d_{dyad}, which are theacitrin A, corilagin, theaflavin, amentoflavone, epigallocatechin gallate (EGCG), and neodiosmin. It is pertinent to note that all six natural products, together with rhoifolin, daidzein, ginkgetin, proanthocyanidins and epitaraxerol, or their closely related derivatives and plant preparations have been reported to possess antiviral, and in some cases anti-coronavirus activities [44–53].

Analysis of the catalytic dyad interaction (Figure 8) revealed theacitrin A, theaflavin, and EGCG as being able to establish specific contacts with both H41 and C145, while corilagin only interacts with C145 and neodiosmin only with H41. Amentoflavone is the only one of the six selected compounds that did not form direct interactions with the catalytic dyad with respect to the 3CL^{pro} crystal structure ($d_{\text{dyad}} > 6$Å). This however changed when 3CL^{pro} dynamics was factored in, the binding distance averaged over five distinct 3CL^{pro} conformations diminished to 3.48 Å. In fact, after incorporating receptor flexibility none of the 38 top-performing natural products demonstrated catalytic dyad distances greater than 3.94 Å (Table S1).
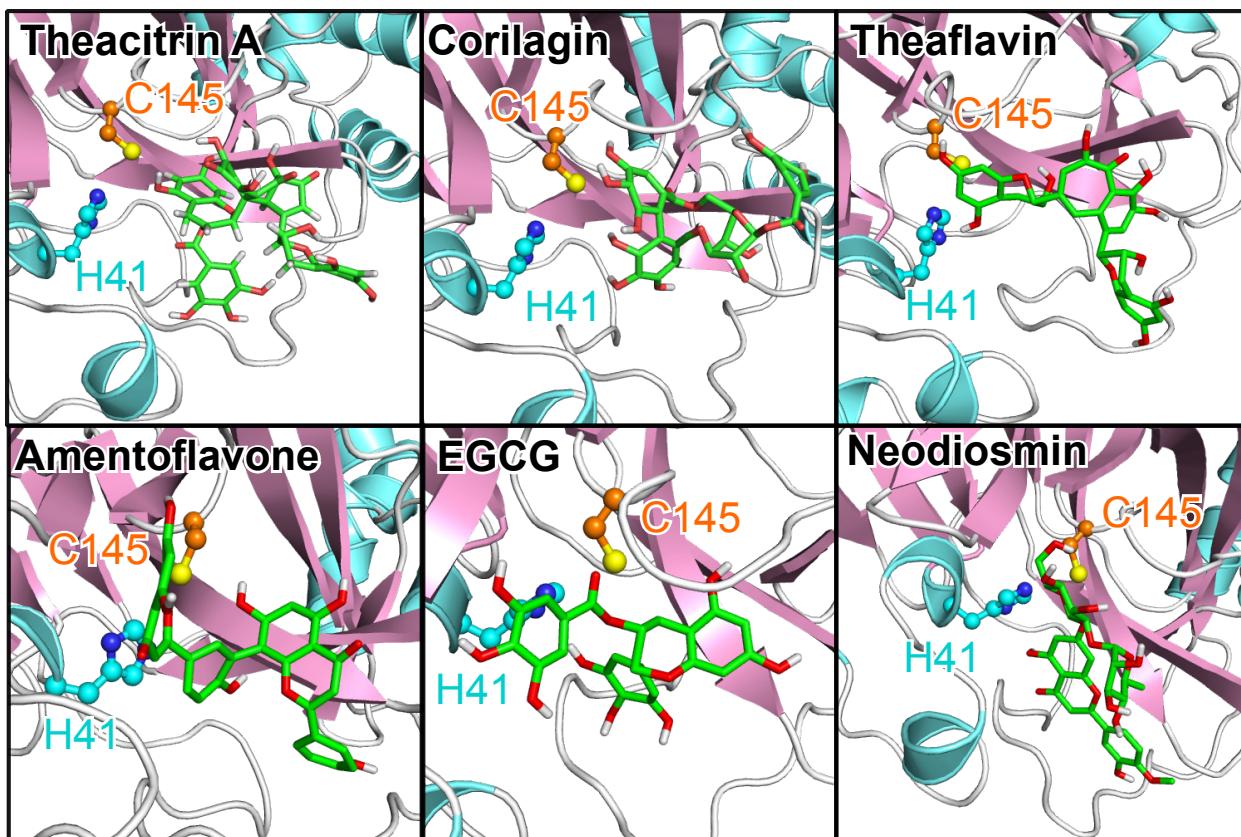
Figure 7: The poses of the six best natural drugs. The top six compounds bind well to the substrate site in terms of both ΔG and closeness to the catalytic dyad of 3CL^{pro}. The same protein and ligand representation as well as color scheme as in Figure 1 are used.

Amentoflavone is a biflavonoid present in multiple plant sources such as *Ginkgo biloba* and St. John's Wort; the latter one also harbors hypericin (see above). Amentoflavone represents one of the top-ranking candidates without an amide functional within its structure; in its place, it employs one of its phenolic groups in forming a hydrogen bond with T26, while establishing another hydrogen bond with H164. In addition, multiple hydrophobic contacts account for the strong binding to the 3CL^{pro} substrate site. Interestingly, amentoflavone, amidst its myriad pharmacological effects, was reported already in 2010 to possess micromolar range inhibitory activity against the main protease enzyme of SARS-CoV [50]. In the study, FRET analysis was employed in a bioactivity-guided isolation of four bioflavonoids and eight diterpenoids from the plant *Torreya nucifera*. Out of all isolated natural products, amentoflavone was found to demonstrate the most potent inhibition of the enzyme with an IC$_{50}$ of 8.3 μM, while 280.8 μM inhibition was obtained for the monoflavonoids apigenin. This agreement with previously reported experimental observations specifically regarding the main protease target of a virus with a high similarity with SARS-CoV-2 provides an important validation for the computational results presented in this work. Amentoflavone has similarly been
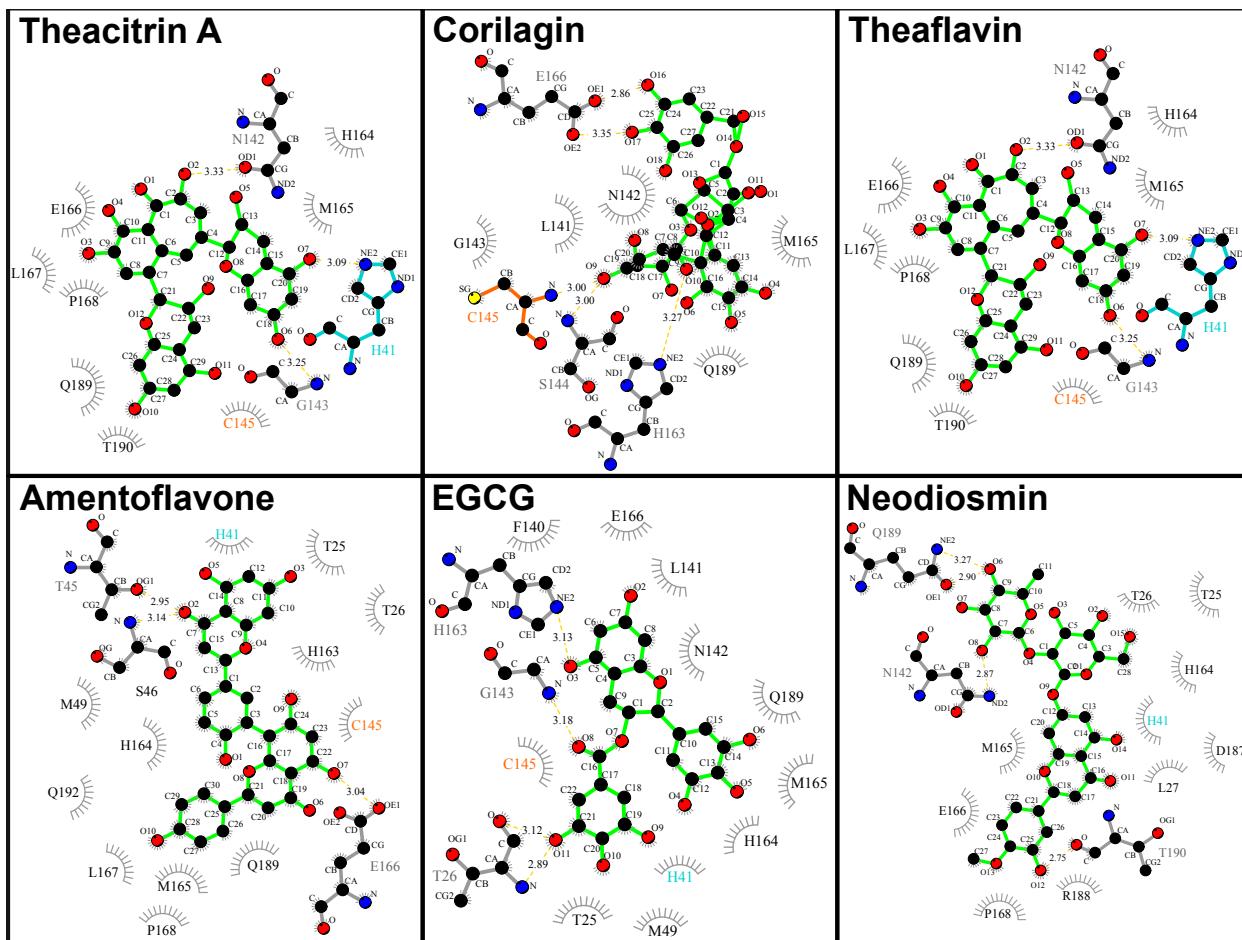
Figure 8: The 3CL^pro–compound interactions for the six best natural drugs. The interactions were analyzed and plotted with LigPlot+ [36, 37]. Hydrogen bonds are indicated by orange dashed lines between the atoms involved and the donor–acceptor distance is given in Å, while hydrophobic contacts are represented by gray arcs with spokes radiating towards the ligand atoms they contact. The contacted atoms are shown with spokes radiating back.

reported with antiviral activities mediated by interaction with other viral targets in viral infections including Dengue virus [54], Coxsackievirus B3 [55], the human immunodeficiency virus [56], and respiratory syncytial virus [57].

Glabrolide and zeylanone are other interesting natural products identified in our computational screening. Glabrolid (Table S1, entry 136) is found in herbs and spices such as *Glycyrrhiza glabra* (licorice) and *Glycyrrhiza uralensis* (Chinese licorice). While possessing structural similarity to the steroid structure, it differs from it by the replacement of the ring D with a six-membered ring and the possession of an extra lactone ring. In interacting with SARS-CoV-2 3CL^pro, it establishes an hydrogen bond with T26 and several hydrophobic contacts, including a critical catalytic dyad contact with C145 (Figure S5). Zeylanone (Table S1, entry 117), on the other hand, employs

its characteristic kink in its five-ring system to nestle comfortably within the hydrophobic cavity, forming a contact with C145, while at the same time establishing hydrogen bonds with the back bone atoms of F140 and E166 (Figure S5). Extracts of the originating plant, *Plumbago zeylanica*, were shown to possess inhibitory activities against Coxsackie virus B3, influenza A virus, and herpes simplex virus type 1 Kupka [58]. The pure zeylanone epoxide was reported to have inhibited influenza virus replication *in vitro* [59].

2.4   Interactions of steroidal hormones with 3CL^{pro}

We profiled the interaction for the principal sex hormones to understand any possible modulatory role played by the steroids on 3CL^{pro}-mediated SARS-CoV-2 viral replication. Not surprisingly given their high structural similarity, all investigated steroids demonstrated comparable binding affinities in interacting with the 3CL^{pro} enzyme (Table S1, entries 140–149). The endogenous steoridal hormones may thus act as natural inhibitors of 3CL^{pro}. This includes cortisol, which was also among the best 3CL^{pro} binders in the FDA-approved drugs category. While estradiol with an averaged computed ΔG of -6.74 kcal/mol could be inferred as possessing a marginally superior binding strength compared to testosterone ($\Delta G = -6.52$ kcal/mol), the observation that testosterone forms direct contacts with the two catalytic dyad residues (estradiol interacts with only C145) suggests that testosterone could have a stronger inhibitory effect on 3CL^{pro} (Figure 9).

However, it is unlikely that the 3CL^{pro} binding properties of the different sex hormones explain the gender-related differences that have been observed in vulnerability to COVID-19, but the possibility exists that the modulatory effects of sex hormones on viral proteins could be part of the complex phenomenon. This appears reasonable considering that age and gender-related differences characterizing human sex hormone levels have also been observed in susceptibility to the SARS-CoV2 infectivity [60]. Other factors likely to contribute will include how other SARS-CoV-2 targets interact with the different hormones, the roles played by sex hormones in modulating immunological processes (especially testosterone), as well as the differential presence of structures like the testis that serves as additional reservoir for angiotensin-converting enzyme 2 (ACE2) needed for viral entry. It is possible that the presence of significantly higher titers of testosterone in younger men contribute to a relative inhibitory resistance against SARS-CoV-2 replication compared with elderly men with the characteristic functionally diminished testosterone levels, while the presence of additional ACE2-related SARS-CoV-2 entry sites in males' testis perhaps explains the higher susceptibility relative to females. The male testis and adipose tissues represent sites in which ACE2 expression levels are reportedly highest in the human body [61]. In fact, the male testicles as well as obesity have been implicated in the higher vulnerability of male patients and obese individuals compared with female patients and non-obese people, respectively [62].
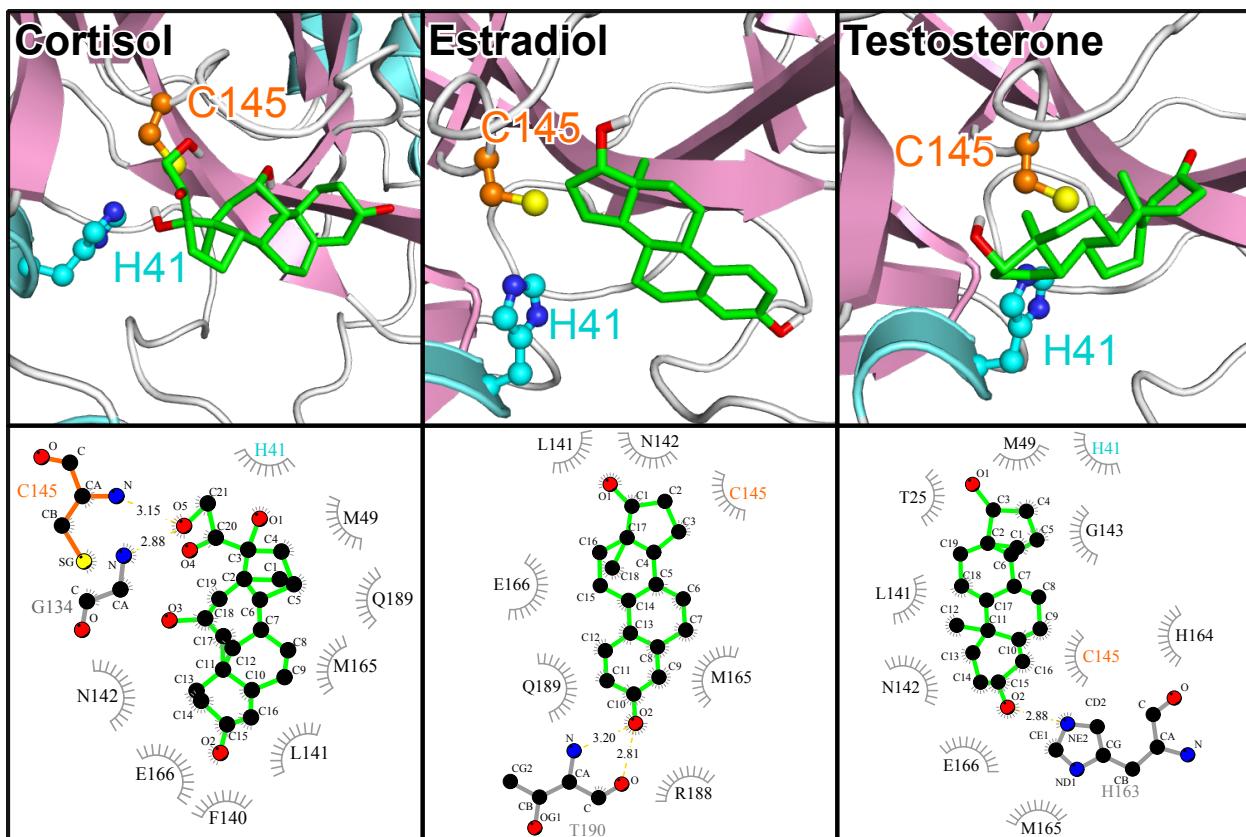
Figure 9: The poses and 3CL^pro–compound interactions for three selected natural steroids.
(Top) The binding poses of cortisol, estradiol, and testosterone are shown, using the same protein
and ligand representations as well as color scheme as in Figure 1. **(Bottom)** The interactions were
analyzed and plotted with LigPlot+ [36, 37]. Hydrogen bonds are indicated by orange dashed lines
between the atoms involved and the donor–acceptor distance is given in Å, while hydrophobic
contacts are represented by gray arcs with spokes radiating towards the ligand atoms they contact.
The contacted atoms are shown with spokes radiating back.

3 Conclusions

Using a computational drug design approach, we have identified a number of compounds whose
ability to establish contacts with critical substrate site residues of 3CL^pro suggests them as good
starting point for the design of inhibitors of SARS-CoV-2 replication. While most of the top-
performing compounds were able to interact with impressive energetics, we have particularly
focused on those compounds that form direct contacts with residues C145 and H41 that make up
the catalytic dyad. The top-ranking compounds belong to diverse chemical classes and multiple
structural features appear to underlie their ability to establish energetically favorable binding with
the 3CL^pro substrate site. In order to gain a better understanding of the prerequisites for good 3CL^pro
ligands, we dissected the top-ranked synthetic compounds with no current application as drugs into

their fragments. This analysis revealed a clear preference for bicyclic, mostly aromatic rings with N and O atoms replacing some of the C atoms. Two or more of these rings are linked by flexible groups, which often bear NH groups and allow the ligands to optimally adopt to the geometry of the substrate site while hydrogen bonds can be formed via the various NH groups, N and O atoms. This knowledge can be used if one wishes to follow the rational drug design route for creating novel SARS-CoV-2 3CL^{pro} inhibitors.

A similar preference for multiple and multicyclic, mostly aromatic rings combined with flexible linker groups is also observed from our screening of existing drugs and natural products. Most of the best binding drugs are characterized by the presence of heterocycles where one or more C atoms are replaced with N (preferred) or O atoms. In addition, several of the rings bear functional groups such as carboxyl or trifluormethyl groups, while the linker groups often involve amide bonds. These groups allow the formation of hydrogen bonds with the amino acids in the substrate site of 3CL^{pro}, while the amide bonds are peptidomimetics, resembling the natural substrate of 3CL^{pro}. The best binding natural products are mostly bioflavonoids, which are also characterized by the presence of mono- and bicyclic aromatic rings which can change their orientation with respect to each other and thus adapt to the geometric requirements of the substrate binding site. However, unlike to the synthetic and drug compounds, the rings and functional groups of the bioflavonoids do not include nitrogen in any form, but a considerable amount of O atoms, either replacing C atoms in the ring structures or as part of functional hydroxyl groups. Especially, the latter lead to hydrogen bonds between the ligands and the substrate site of 3CL^{pro}.

A similar finding as for the bioflavonoids was made for hypericin, which we identified as a good binding drug but does not involve any N atom. It is an anthraquinone derivative that is found in Saint John's wort and is thus also a natural product. Hypericin is considered for the treatment of depressions and is believed to act as an antibiotic, antiviral and non-specific kinase inhibitor [63]. Interestingly, with amentoflavone we identified another natural product found in Saint John's wort which may act as SARS-CoV-2 3CL^{pro} inhibitor. It should be noted that amentoflavone was reported a decade ago with micromolar range inhibitory effect on the closely related main protease enzyme of SARS-CoV [50]. This finding provides a validation for the virtual screening method incorporating protein dynamics employed in this work. Another interesting observation that we made is that several tyrosine kinase inhibitors were identified by our screening approach against 3CL^{pro}. In addition to hypericin, this includes nilotinib, afatinib and R-343. Given that an earlier study reported another tyrosine kinase inhibitor, imatinib as inhibitor of SARS-CoV and MERS-CoV viral RNA expression [41], this class of inhibitors deserves further attention for finding a therapeutic for SARS-CoV-2. Finally, it should be mentioned that we identified many compounds that bind better to the main protease of SARS-CoV-2 than the eight reference compounds whose 3CL^{pro} inhibition

capabilities were already demonstrated *in vitro* [16]. Ritonavir and remdesivir, which are currently tested as a treatment of COVID-19 [40], also bind to SARS-CoV-2. Interestingly, remdesivir which is thought to inhibit RNA-dependent RNA polymerase from SARS-CoV-2 binds better to the catalytic dyad of 3CL[pro] than ritonavir, which is considered to be a protease inhibitor.

After we had identified cortisol among the drugs that exhibit good binding properties towards 3CL[pro], we decided to screen all endogenous steroidal hormones and found that most of them bind well to the substrate site of 3CL[pro]. These hormones may thus act as natural inhibitors of the SARS-CoV-3 main protease, which would explain the age divide in terms of COVID-19 severity. This hypothesis is supported by the fact that progesterone was identified to be antiviral against SARS-CoV-2 [64]. Moreover, low testosterone levels in men are a risk factor for being more severely affected by COVID-19 than men with normal or high testosterone levels and women in general [60]. The question why postmenopausal women do not equally strongly suffer from COVID-19 might be answered by the different expression levels of ACE2, the receptor that SARS-CoV-2 uses for host entry and which in considerably amounts is expressed in testis [62]. A similar argument may account for the fact that COVID-19 is less prevalent in children [65], who have, like older people, low levels of sex hormones and may thus not rely on their possible protection from 3CL[pro] inhibition. However, it was found that ACE2 is present in considerably lower quantities in children compared to adults above 18 years in age [66].

In the next state of our search for potent and clinically useful inhibitors of SARS-CoV-2 infection, we plan to employ further docking analysis involving other druggable COVID-19 targets, MD simulations to further assess the best docking predictions, as well as *in vitro* and cell-based assays to validate the outcomes of our the CADD data presented here.

Methods and Materials

Virtual screening protocol. We assembled a total of 1,227,186 ligand structural models principally from the ZINC database [31–33] which hosts 3D models of ligand molecules from other databases including the DrugBank library [34, 35]. The screened ligand dataset includes existing drugs approved both by the United States Food and Drugs Administration (FDA) as well as by other countries regulatory authorities. In this work, we make clear distinction between FDA and non-FDA-approved drugs. Drugs still being investigated for various clinical indications represent another subset of our ligand library. A small but distinct subclass of our library consists of drugs being currently investigated for their activities (both validated and perceived) as anti-SARS-CoV-2 agents. More than 140,000 of the downloaded models are of natural products origin, while the largest part of the ZINC database-downloaded models contain synthetic chemical compounds. Apart from the ZINC database models, almost 3,200 models of natural products from Nigerian plants

were obtained from an in-house database. While we did not deliberately characterize the screening virtual library based on chemical classes, we assumed that inclusion of compounds from multiple sources would achieve a decent coverage of sufficiently diverse chemical classes. This coverage is especially crucial for appropriately sampling the chemical scaffolds that are most fitting for interacting with the enzyme binding site. For the same reason and the need to look beyond the present state of understanding of the 3CL^{pro} enzyme, we have not selected a screening protocol focused solely on known protease inhibitors. Other compounds, including the peptidomimetic inhibitor called N3, tideglusib, shikonin, cinanserin, ebselen, carmofur, disulfiram and PX-12 that had been reported to inhibit 3CL^{pro} of SARS-CoV-2 [16] were also included in our virtual screening runs as references. Random visual inspection of the ligand models obtained from the different sources indicated *prima facie* structural correctness, and the isolated instances of inappropriate model (e.g., replacing esterified candesartan currently in clinical trials with candesartan) were treated on individual bases.

We retrieved the X-ray crystallographic structure of the SARS-CoV-2 (3CL^{pro}) enzyme (PDB code 6LU7 [16]) from the RCSB website [67]. The structure represents the functional form of 3CL^{pro} in complex with the inhibitor N3 at a 2.16 Å resolution. In the structure, the peptidomimetic inhibitor could be seen making contacts with critical substrate binding site residues including the C145–H41 catalytic dyad. A docking grid allowing a 3.0 Å buffer region around the bound position of N3 in the 3CL^{pro} substrate site was generated with AutoDock Tool [68, 69] and centered at x, y, z-position of -10.745 Å, 12.33 Å, 68.84 Å. The buffer region serves to accommodate different molecular sizes covered in the screening library. After adding hydrogen atoms to 3CL^{pro}, Gasteiger partial charges were computed and added to 3CL^{pro} using AutoDock Tool [68, 69] and the resulting protein model was saved in the PDBQT format. PDBQT files were also generated for each of the ligands in the screening library, following which the library was subjected to virtual screening against the 3CL^{pro} molecule using AutoDock Vina [70], which treats the ligand as fully flexible while keeping the receptor rigid. Different cycles of screening were performed in a way that balances the complexity of the ligand libraries with computational resources on one hand, with the need to assure reliability of our computational model for the biochemical process (i.e., enzyme inhibition) being investigated. In the initial round, the screening group A (sGrA) composed of 1,068,161 million synthetic compounds from the ZINC database was screened against the 3CL^{pro} crystal structure. From sGrA, 9,515 virtual hits with computed binding free energy ΔG values less than or equal to -8.0 kcal/mol were selected for rescreening using ensemble docking [71]. The second screening group B (sGrB) that is composed of FDA-approved drugs (2,099) and other approved and investigational drugs (7,991), as well as natural products from ZINC library (145,753) and from our in-house library of secondary metabolites from Nigerian plants (3,182), and reference compounds

was similarly screened against the 3CL^{pro} crystal structure and as complete set submitted to the second round of screening with ensemble docking.

Virtual rescreening protocol – incorporating protein dynamics.   To account for protein flexibility in the second docking round, we performed ensemble docking [71] on both the hit list generated from sGrA ($N = 9,515$) and the entire sGrB containing a total of 159,025 compounds. To this end, we performed a 100 ns MD simulation of the 3CL^{pro}–N3 complex in solution using the MD software Gromacs 2018 [72]. As protein force field AMBER14SB [73] with Parmbsc1 parameters [74] was used and combined with the TIP3P water model [75] to explicitly simulate water. For N3, the generalized AMBER force field parameters (GAFF) [73] were used. To derive the parameters, we conducted quantum mechanics optimization at the HF6-31G* level of theory with Gaussian 09 [76] and used the restrained electrostatic potential (RESP) method to calculate partial charges [77, 78] via Antechamber [79, 80] available in AmberTool 19 [81]. To generate the GROMACS input files for N3, we used ACPYPE tool [82]. As starting structure, the 3CL^{pro}–N3 crystal structure was taken (PDB code 6LU7) [16], which was centered in a cubic box of size $80 \times 80 \times 80$ Å^3, solvated, and NaCl was added at a concentration of 150 mM while at the same time neutralizing the system, resulting in a system size of 50,966 atoms in total. The system was first energy-minimized using a steepest descent algorithm, equilibrated in the *NVT* ensemble (i.e., with a constant number of molecules, volume, and temperature) for 0.1 ns and then for 1 ns in the *NpT* ensemble at $T = 310$ K (Nosé-Hoover thermostat [83, 84]) and $p = 1.0$ bar (Parrinello-Rahman barostat [85]). The production MD simulation was performed for 100 ns with the same parameters as in the *NpT* equilibration. Electrostatic interactions were treated with the particle-mesh Ewald method [86, 87] in conjunction with periodic boundary conditions and a real-space cutoff of 12 Å. The Lennard-Jones interactions were also cut at 12 Å. A leapfrog stochastic dynamics integrator was used for the integration of the equations of motion using a time step of 2 fs. The LINCS algorithm [88] was used to constrain all bond lengths during the MD simulation. The resulting trajectory was subjected to geometric clustering using the Daura algorithm [89] based on the conformations sampled by 3CL^{pro} substrate binding site residues (within 10 Å of N3). With a clustering cutoff of 1.5 Å, 15 clusters were identified, from which the representative structures of the five most populated clusters representing 88.1% of the dynamics were selected for ensemble docking. As for the docking against the 3CL^{pro} crystal structure, a docking grid was generated for each of the five MD-generated 3CL^{pro} conformations with a 3.0 Å buffer around the bound position of N3. The selected hits from the sGrA as well as the entire sGrB group were then subjected to another round of screening against the five MD-generated 3CL^{pro} conformers using AutoDock Vina. The results were subsequently analyzed to understand critical structural and physicochemical aspects of 3CL^{pro} enzyme inhibition that may benefit rational design of inhibitors.

Visualization.   All 3D structures were created with PyMol [90], while Figure 2 as well as the 2D structures were made with DataWarrior [91]. The interactions between protein and ligands were analyzed and plotted with LigPlot+ [36, 37].

Acknowledgments

The authors gratefully acknowledge the computing time granted through JARA-HPC (project COVID19MD) on the supercomputer JURECA at Forschungszentrum Jülich [92], the hybrid computer cluster purchased from funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) project number INST 208/704-1 FUGG, and the Zentrum für Informations- und Medientechnologie (ZIM) at Heinrich-Heine University. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

[1] S.R. Weiss and S. Navas-Martin. Coronavirus Pathogenesis and the Emerging Pathogen Severe Acute Respiratory Syndrome Coronavirus. *Microbiology and Molecular Biology Reviews*, 69(4):635–664, 2005.

[2] P.S. Masters. The Molecular Biology of Coronaviruses. volume 66 of *Advances in Virus Research*, pages 193–292. Academic Press, 2006.

[3] S. Siddell, H. Wege, and V. Ter Meulen. The biology of coronaviruses. *Journal of General Virology*, 64(4):761–776, 1983.

[4] S.R. Weiss and J.L. Leibowitz. Chapter 4 – Coronavirus Pathogenesis. volume 81 of *Advances in Virus Research*, pages 85–164. Academic Press, 2011.

[5] R.L. Graham, E.F. Donaldson, and R.S. Baric. A decade after SARS: strategies for controlling emerging coronaviruses. *Nature Reviews Microbiology*, 11(12):836–848, 2013.

[6] Z.A. Memish, A.I. Zumla, R.F. Al-Hakeem, A.A. Al-Rabeeah, and G.M. Stephens. Family cluster of Middle East respiratory syndrome coronavirus infections. *New England Journal of Medicine*, 368(26):2487–2494, 2013.

[7] A.M. Zaki, S. Van Boheemen, T.M. Bestebroer, A.D.M.E. Osterhaus, and R.A.M. Fouchier. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *New England Journal of Medicine*, 367(19):1814–1820, 2012.

[8] L. van der Hoek, K. Pyrc, M.F. Jebbink, W. Vermeulen-Oost, R.J.M. Berkhout, K.C. Wolthers, P.M.E. Wertheim-van Dillen, J. Kaandorp, J. Spaargaren, and B. Berkhout. Identification of a new human coronavirus. *Nature medicine*, 10(4):368–373, 2004.

[9] P.C.Y. Woo, S.K.P. Lau, C. Chu, K. Chan, H. Tsoi, Y. Huang, B.H.L. Wong, R.W.S. Poon, J.J. Cai, W. Luk, L.L.M. Poon, S.S.Y. Wong, Y. Guan, J.S.M. Peiris, and K. Yuen. Characterization and Complete Genome Sequence of a Novel Coronavirus, Coronavirus HKU1, from Patients with Pneumonia. *Journal of Virology*, 79(2):884–895, 2005.

[10] JHU CSSE. Coronavirus COVID-19 Global Cases by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU), 2020.

[11] H. Streeck, B. Schulte, B. Kuemmerer, E. Richter, T. Hoeller, C. Fuhrmann, E. Bartok, R. Dolscheid, M. Berger, L. Wessendorf, M. Eschbach-Bludau, A. Kellings, A. Schwaiger, M. Coenen, P. Hoffmann, M. Noethen, A.M. Eis-Huebinger, M. Exner, R. Schmithausen, M. Schmid, and B. Kuemmerer. Infection fatality rate of SARS-CoV-2 infection in a German community with a super-spreading event. *medRxiv*, 2020.

[12] L. Zhang, D. Lin, X. Sun, U. Curth, C. Drosten, L. Sauerhering, S. Becker, K. Rox, and R. Hilgenfeld.

[13] Y.C. Li, W.Z. Bai, and T. Hashikawa. The neuroinvasive potential of SARS-CoV2 may play a role in the respiratory failure of COVID-19 patients. *Journal of Medical Virology*, 92(6):552–555.

[14] G. Li and E. De Clercq. Therapeutic options for the 2019 novel coronavirus (2019-nCoV), 2020.

[15] C.I. Paules, H.D. Marston, and A.S. Fauci. Coronavirus Infections – More Than Just the Common Cold. *JAMA*, 323(8):707–708, 02 2020.

[16] Z. Jin, X. Du, Y. Xu, Y. Deng, M. Liu, Y. Zhao, B. Zhang, X. Li, L. Zhang, C. Peng, Y. Duan, J. Yu, L. Wang, K. Yang, F. Liu, R. Jiang, X. Yang, T. You, X. Liu, X. Yang, F. Bai, H. Liu, X. Liu, L. W. Guddat, W. Xu, G. Xiao, C. Qin, Z. Shi, H. Jiang, Z. Rao, and H. Yang. Structure of M$^{\text{pro}}$ from COVID-19 virus and discovery of its inhibitors. *Nature*, 2020.

[17] T. Pillaiyar, M. Manickam, V. Namasivayam, Y. Hayashi, and S.H. Jung. An overview of severe acute respiratory syndrome–coronavirus (SARS-CoV) 3CL protease inhibitors: Peptidomimetics and small molecule chemotherapy. *Journal of Medicinal Chemistry*, 59(14):6595–6628, 2016.

[18] E. de Wit, N. van Doremalen, D. Falzarano, and V.J. Munster. SARS and MERS: recent insights into emerging coronaviruses. *Nature Reviews Microbiology*, 14(8):523, 2016.

[19] X. Xue, H. Yang, W. Shen, Q. Zhao, J. Li, K. Yang, C. Chen, Y. Jin, M. Bartlam, and Z. Rao. Production of Authentic SARS-CoV M$^{\text{pro}}$ with Enhanced Activity: Application as a Novel Tag-cleavage Endopeptidase for Protein Overproduction. *Journal of Molecular Biology*, 366(3):965–975, 2007.

[20] Q.M. Wang, R.B. Johnson, G.A. Cox, E.C. Villarreal, and J. Richard Loncharich. A Continuous Colorimetric Assay for Rhinovirus-14 3C Protease Using Peptidep-Nitroanilides as Substrates. *Analytical Biochemistry*, 252(2):238–245, 1997.

[21] S. Durdagi, B. Aksoydan, B. Dogan, K. Sahin, and A. Shahraki. Screening of Clinically Approved and Investigation Drugs as Potential Inhibitors of COVID-19 Main Protease: A Virtual Drug Repurposing Study. 3 2020.

[22] P. Schlagenhauf, M.P. Grobusch, J.D. Maier, and P. Gautret. Repurposing antimalarials and other drugs for covid-19, 2020.

[23] J. Meng, G. Xiao, J. Zhang, X. He, M. Ou, J. Bi, R. Yang, W. Di, Z. Wang, Z. Li, H. Gao, L. Liu, and G. Zhang. Renin-angiotensin system inhibitors improve the clinical outcomes of COVID-19 patients with hypertension. *Emerging Microbes & Infections*, 9(1):757–760, 2020.

[24] F. Bennardo, C. Buffone, and A. Giudice. New therapeutic opportunities for COVID-19 patients with Tocilizumab: Possible correlation of interleukin-6 receptor inhibitors with osteonecrosis of the jaws. *Oral Oncology*, page 104659, 2020.

[25] Y.W. Chen, C.P.B. Yiu, and K.Y. Wong. Prediction of the SARS-CoV-2 (2019-nCoV) 3c-like protease (3CL pro) structure: virtual screening reveals velpatasvir, ledipasvir, and other drug repurposing candidates. *F1000Research*, 9, 2020.

[26] A.T. Ton, F. Gentile, M. Hsing, F. Ban, and A. Cherkasov. Rapid Identification of Potential Inhibitors of SARS-CoV-2 Main Protease by Deep Docking of 1.3 Billion Compounds. *Molecular Informatics*, n/a(n/a), 2020.

[27] A.A. Elfiky. Anti-HCV, nucleotide inhibitors, repurposing against COVID-19. *Life Sciences*, 248:117477, 2020.

[28] N. Singh, A.C. Halliday, J.M. Thomas, O.V. Kuznetsova, R. Baldwin, E.C.Y. Woon, P.K. Aley, I. Antoniadou, T. Sharp, S.R. Vasudevan, and G.C. Churchill. A safe lithium mimetic for bipolar disorder. *Nature communications*, 4(1):1–7, 2013.

[29] N.A. Baker, D. Sept, S. Joseph, M.J. Holst, and J.A. McCammon. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences*, 98(18):10037–10041, 2001.

[30] E. Jurrus, D. Engel, K. Star, K. Monson, J. Brandi, L.E. Felberg, D.H. Brookes, L. Wilson, J. Chen, K. Liles, M. Chun, P. Li, D.W. Gohara, T. Dolinsky, R. Konecny, D.R. Koes, J.E. Nielsen, T. Head-Gordon, W. Geng, R. Krasny, G.W. Wei, M.J. Holst, J.A. McCammon, and N.A. Baker. Improvements to the APBS biomolecular solvation software suite. *Protein Science*, 27(1):112–128, 2018.

[31] J.J. Irwin and B.K. Shoichet. ZINC–A Free Database of Commercially Available Compounds for Virtual Screening. *Journal of Chemical Information and Modeling*, 45(1):177–182, 2005.

[32] J.J. Irwin, T. Sterling, M.M. Mysinger, E.S. Bolstad, and R.G. Coleman. ZINC: A Free Tool to Discover Chemistry for Biology. *Journal of Chemical Information and Modeling*, 52(7):1757–1768, 2012.

[33] T. Sterling and J.J. Irwin. Zinc 15 – ligand discovery for everyone. *Journal of Chemical Information and Modeling*, 55(11):2324–2337, 2015.

[34] D.S. Wishart, C. Knox, A.C. Guo, S. Shrivastava, M. Hassanali, P. Stothard, Z. Chang, and J. Woolsey. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Research*, 34.

[35] D.S. Wishart, Y.D. Feunang, A.C. Guo, E.J. Lo, A. Marcu, J.R. Grant, T. Sajed, D. Johnson, C. Li, Z. Sayeeda, N. Assempour, I. Iynkkaran, Y. Liu, A. Maciejewski, N. Gale, L. Wilson, A.and Chin, R. Cummings, D. Le, A. Pon, C. Knox, and M. Wilson. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Research*, 46(D1):D1074–D1082, 11 2017.

[36] A.C. Wallace, R.A. Laskowski, and J.M. Thornton. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Engineering, Design and Selection*, 8(2):127–134, 02 1995.

[37] R.A. Laskowski and M.B. Swindells. LigPlot+: Multiple Ligand–Protein Interaction Diagrams for Drug Discovery. *Journal of Chemical Information and Modeling*, 51(10):2778–2786, 2011.

[38] H. Chen, R. Feng, I. Muhammad, G. Abbas, Y. Zhang, Y. Ren, X. Huang, R. Zhang, L. Diao, X. Wang, and G. Li. Protective effects of hypericin against infectious bronchitis virus induced apoptosis and reactive oxygen species in chicken embryo kidney cells. *Poultry Science*, 98(12):6367–6377, 2019.

[39] I. Rajib, P. Md. Rimon, P.A. Sundar, U. Nizam, R. Md. Sajjadur, M. Abdulla Al, H. Md. Nayeem, A. Md. Ackas, and A. H. Mohammad. A molecular modeling approach to identify effective antiviral phytochemicals against the main protease of SARS-CoV-2. *Journal of Biomolecular Structure and Dynamics*, 0(0):1–12, 2020.

[40] I.F.N. Hung, K.C. Lung, E.Y.K. Tso, R. Liu, T.W.H. Chung, M.Y. Chu, Y.Y. Ng, J. Lo, J. Chan, A.R. Tam, H.P. Shum, V. Chan, A.K.L. Wu, K.M. Sin, W.S. Leung, W.L Law, D.C. Lung, S. Sin, P. Yeung, C.Y.C. Yip, R.R. Zhang, A.Y.F Fung, E.Y.W Yan, K.H. Leung, J.D. Ip, A.W.H. Chu, W.M. Chan, A.C.K. Ng, R. Lee, K. Fung, A. Yeung, T.C. Wu, J.W.M Chan, W.W. Yan, W.M. Chan, J.F.W Chan, A.K.W. Lie, Q.T.K. Tsang, V.C.C. Cheng, T.L. Que, C.S. Lau, K.H.

Chan, K.K.W. To, and K.Y. Yue. Triple combination of interferon beta-1b, lopinavir–ritonavir, and ribavirin in the treatment of patients admitted to hospital with COVID-19: an open-label, randomised, phase 2 trial. *The Lancet*, 395(10238):1695–1704, 2020.

[41] C.M. Coleman, J.M. Sisk, R.M. Mingo, E.A. Nelson, J.M. White, and M.B. Frieman. ABL kinase inhibitors are potent inhibitors of SARS-CoV and MERS-CoV fusion. *Journal of Virology*, 2016.

[42] P. Rousselot. A randomized non-comparative phase 2 pilot study testing the value of imatinib mesylate as an early treatment of covid-19 disease in aged hospitalized patients.

[43] Q. Ning. A pilot study of sildenafi in the treatment of covid-19.

[44] M. Ohba, T. Oka, T. Ando, S. Arahata, A. Ikegaya, H. Takagi, N. Ogo, C. Zhu, K. Owada, F. Kawamori, Q. Wang, L.J. Saif, and A. Asai. Antiviral effect of theaflavins against calicivir-uses. *The Journal of Antibiotics*, 70:443–447, 2016.

[45] Y. Guo, T. Luo, F. Wu, H. Liu, H. Li, Y. Mei, S. Zhang, J. Tao, J. Dong, Y. Fang, and L. Zhao. Corilagin protects against HSV_1 encephalitis through inhibiting the TLR_1 signaling pathways in vivo and in vitro. *Molecular Neurobiology*, 52:1547–1560, 2014.

[46] S. Yeo, J.H. Song, E. Hong, B. Lee, Y. Kwon, S. Chang, S. Kim, S. Lee, J.H. Park, and H.J. Antiviral effects of *Phyllanthus urinaria* containing corilagin against human enterovirus 71 and Coxsackievirus A16 in vitro. *Archives of Pharmacal Research*, 38:193–202, 2014.

[47] F. Li, X. Song, G. Su, Y. Wang, Z. Wang, J. Jia, S. Qing, L. Huang, Y. Wang, K. Zheng, and Y. Wang. Amentoflavone Inhibits HSV-1 and ACV-Resistant Strain Infection by Suppressing Viral Early Infection. *Viruses*, 11(5):466, 2019.

[48] Y.M. Lin, M.T. Flavin, Schure R., F.C. Chen, R. Sidwell, D.L. Barnard, J.H. Huffman, and E.R. Kern. Antiviral activities of biflavonoids. *Planta Medica*, 65(2):120–125, 1999.

[49] S. Jo, S. Kim, D.H. Shin, and M.S. Kim. Inhibition of SARS-CoV 3CL protease by flavonoids. *Journal of Enzyme Inhibition and Medicinal Chemistry*, 35(1):145–151, 2020.

[50] Y.B. Ryu, H.J. Jeong, J.H. Kim, Y.M. Kim, J.Y. Park, D. Kim, T.T.H. Nguyen, S.J. Park, J.S. Chang, K.H. Park, M.C. Rho, and W.S. Lee. Biflavonoids from *Torreya* nucifera displaying SARS-CoV 3CL^{pro} inhibition. *Bioorganic & medicinal chemistry*, 18(22):7940–7947, 2010.

[51] C.N. Chen, C.P. Lin, K.K. Huang, W.C. Chen, P.H Hsieh, H.P. and.Liang, and J.T.A. Hsu. Inhibition of SARS-CoV 3C-like protease activity by theaflavin-3,3'-digallate TF3. *Evidence-Based Complementary and Alternative Medicine*, 2(2):209–215, 2005.

[52] K. Miki, T. Nagai, K. Suzuki, R. Tsujimura, K. Koyama, K. Kinoshita, K. Furuhata, H. Yamada, and K. Takahashi. Anti-influenza virus activity of biflavonoids. *Bioorganic & Medicinal Chemistry Letters*, 17(3):772–775, 2007.

[53] Katayama, H., Tokutake, S., Yamamoto, N., and Yoshinaka, Y. Anti-SARS coronavirus agent. *KIKKOMAN CORP*, JP2005314316A.

[54] P. Coulerie, M. Nour, A. Maciuk, C. Eydoux, J.C. Guillemot, N. Lebouvier, E. Hnawia, K. Leblanc, G. Lewin, B. Canard, and . Figadere. Structure-Activity Relationship Study of Biflavonoids on the Dengue Virus Polymerase DENV-NS5 RdRp. *Planta Medica*, 79:1313–1318, 08 2013.

[55] S. Wilsky, K. Sobotta, N. Wiesener, J. Pilas, N. Althof, T. Munder, P. Wutzler, and A. Henke. Inhibition of fatty acid synthase by amentoflavone reduces coxsackievirus B3 replication. *Archives of virology*, 157:259–69, 11 2011.

[56] Y.M. Lin, H. Anderson, M.T. Flavin, Y.H. Pai, E. Mata-Greenwood, T. Pengsuparp, J.M. Pezzuto, R.F. Schinazi, S.H. Hughes, and F.C. Chen. In vitro anti-HIV activity of biflavonoids isolated from *Rhus succedanea* and *Garcinia multiflora*. *Journal of natural products*, 60(9):884–888, 1997.

[57] S.C. Ma, P.P. But, V.E. Ooi, Y.H. He, S.H. Lee, S.F. Lee, and R.C. Lin. Antiviral amentoflavone from *Selaginella sinensis*. *Biological & pharmaceutical bulletin*, 24(3):311–312, 2001.

[58] T. Gebre-Mariam, R. Neubert, P.C. Schmidt, P. Wutzler, and M. Schmidtke. Antiviral activities of some ethiopian medicinal plants used for the treatment of dermatological disorders. *Journal of ethnopharmacology*, 104(1-2):182–187, 2006.

[59] L. Cetina-Montejo, G. Ayora-Talavera, and R. Borges-Argáez. Zeylanone epoxide isolated from *Diospyros anisandra* stem bark inhibits influenza virus *in vitro*. *Archives of Virology*, 164:1543–1552, 03 2019.

[60] M. Schroeder, B. Tuku, D. Jarczak, A. Nierhaus, T. Bai, H. Jacobsen, M. Zickler, Z. Mueller, S. Stanelle-Bertram, A. Meinhardt, J. Aberle, S. Kluge, and G. Gabriel. The majority of male patients with COVID-19 present low testosterone levels on admission to intensive care in hamburg, germany: a retrospective cohort study. *medRxiv*, 2020.

[61] M. Li, L. Li, Y. Zhang, and X. Wang. Expression of the SARS-CoV-2 cell receptor gene *ace2* in a wide variety of human tissues. *Infectious Diseases of Poverty*, 9(45), 2020.

[62] A. Shastri, J. Wheat, S. Agrawal, N. Chaterjee, K. Pradhan, M. Goldfinger, N. Kornblum, U. Steidl, A. Verma, and J. Shastri. Delayed clearance of SARS-CoV2 in male compared to female patients: High ACE2 expression in testes suggests possible existence of gender-specific viral reservoirs. *medRxiv*, 2020.

[63] A. DerMarderosian and J.A. Beutler. *The review of natural products*. Lippincott Williams & Wilkins, 7 edition, 2012.

[64] D.E. Gordon, G.M. Jang, M. Bouhaddou, J. Xu, K. Obernier, K.M. White, M.J. O'Meara, V.V. Rezelj, J.Z. Guo, D.L. Swaney, T.A. Tummino, R. Huettenhain, R.M. Kaake, A.L. Richards, B. Tutuncuoglu, H. Foussard, J. Batra, K. Haas, M. Modak, M. Kim, P. Haas, B.J. Polacco, H. Braberg, J.M. Fabius, M. Eckhardt, M. Soucheray, M.J. Bennett, M. Cakir, M.J. McGregor, Q. Li, B. Meyer, F. Roesch, T. Vallet, A. Mac Kain, L. Miorin, E. Moreno, Z.Z. Naing, Y. Zhou, S. Peng, Y. Shi, Z. Zhang, W. Shen, I.T. Kirby, J.E. Melnyk, J.S. Chorba, K. Lou, S.A. Dai, I. Barrio-Hernandez, D. Memon, C. Hernandez-Armenta, J. Lyu, C.J.P. Mathy, T. Perica, K.B. Pilla, S.J. Ganesan, D.J. Saltzberg, R. Rakesh, X. Liu, S.B. Rosenthal, L. Calviello, S. Venkataramanan, J. Liboy-Lugo, Y. Lin, X.P. Huang, Y.F. Liu, S.A. Wankowicz, M. Bohn, M. Safari, F.s. Ugur, C. Koh, N.S. Savar, Q.D. Tran, D. Shengjuler, S.J. Fletcher, M.C. O'Neal, Y. Cai, J.C.J. Chang, D.J. Broadhurst, S. Klippsten, P.P. Sharp, N.A. Wenzell, D. Kuzuoglu, H.Y. Wang, R. Trenker, J.M. Young, D.A. Cavero, J. Hiatt, T.L. Roth, U. Rathore, A. Subramanian, J. Noack, M. Hubert, R.M. Stroud, A.D. Frankel, O.S. Rosenberg, K.A. Verba, D.A. Agard, M. Ott, M. Emerman, N. Jura, M. von Zastrow, E. Verdin, A. Ashworth, O. Schwartz, C. d'Enfert, S. Mukherjee, M. Jacobson, H.S. Malik, D.G. Fujimori, T. Ideker, C.S. Craik, S.N. Floor, J.S. Fraser, J.D. Gross, A. Sali, B.L. Roth, D. Ruggero, J. Taunton, T. Kortemme, P. Beltrao, M. Vignuzzi, A. García-Sastre, K.M. Shokat, B.K. Shoichet, and N.J. Krogan. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature*, pages 1–13, 2020.

[65] Z. Wu and J.M. McGoogan. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: Summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. *JAMA*, 323(13):1239–1242, 2020.

[66] A.B. Patel and A. Verma. Nasal ACE2 Levels and COVID-19 in Children. *JAMA*, 2020.

[67] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, and P.E. Bourne. The Protein Data Bank. *Nucleic Acids Research*, 28(1):235–242, 01 2000.

[68] D.S. Goodsell, G.M. Morris, and A.J. Olson. Automated docking of flexible ligands: Applications of AutoDock. *Journal of Molecular Recognition*, 9(1):1–5, 1996.

[69] D. Santos-Martins, S. Forli, M.J. Ramos, and A.J. Olson. AutoDock4$_{Zn}$: An Improved AutoDock Force Field for Small-Molecule Docking to Zinc Metalloproteins. *Journal of Chemical Information and Modeling*, 54(8):2371–2379, 2014. PMID: 24931227.

[70] O. Trott and A.J. Olson. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, 31(2):455–461, 2010.

[71] R.E. Amaro, J. Baudry, J. Chodera, Ö. Demir, J.A. McCammon, Y. Miao, and J.C. Smith. Ensemble Docking in Drug Discovery. *Biophysical Journal*, 114(10):2271–2278, 2018.

[72] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *Journal of Chemical Theory and Computation*, 4(3):435–447, 2008. PMID: 26620784.

[73] C.J. Dickson, L. Rosso, R.M. Betz, R.C. Walker, and I.R. Gould. GAFFlipid: a General Amber Force Field for the accurate molecular dynamics simulation of phospholipid. *Soft Matter*, 8:9617–9627, 2012.

[74] I. Ivani, P. D. Dans, A. Noy, A. Pérez, I. Faustino, A. Hospital, J. Walther, P. Andrio, R. Goñi, A. Balaceanu, G. Portella, F. Battistini, J.L. Gelpí, C. González, M. Vendruscolo, C.A. Laughton, S.A. Harris, D.A. Case, and M. Orozco. Parmbsc1: a refined force field for DNA simulations. *Nature methods*, 13(1):55, 2016.

[75] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, and M.L. Klein. Comparison of simple potential functions for simulating liquid water. *Journal of Chemical Theory and Computation*, 79(2):926–935, 1983.

[76] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G.A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H.P. Hratchian, A.F. Izmaylov, J. Bloino, G. Zheng, J.L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J.A. Montgomery, J.E. Peralta, F. Ogliaro, M. Bearpark, J.J. Heyd, E. Brothers, K.N. Kudin, V.N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J.M. Millam, M. Klene, J.E. Knox, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, R.L. Martin, K. Morokuma, V.G. Zakrzewski, G.A. Voth, P. Salvador, J.J. Dannenberg, S. Dapprich, A.D. Daniels, Ö. Farkas, J.B. Foresman, J.V. Ortiz, J. Cioslowski, and D.J. Fox. Gaussian 09 Revision E.01, 2009. Gaussian Inc. Wallingford CT.

[77] C.I. Bayly, P. Cieplak, W.D. Cornell, and P.A. Kollman. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *Journal of Physical Chemistry*, 97(2):10269–10280, 1993.

[78] W.D. Cornell, P. Cieplak, C.I. Bayly, and P.A. Kollman. Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation. *Journal of Physical Chemistry*, 115(21):9620–9631, 1993.

[79] J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, and D.A. Case. Development and testing of a general amber force field. *Journal of Computational Chemistry*, 25:1157–1174, 2004.

[80] J. Wang, W. Wang, P.A. Kollman, and D.A. Case. Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of Molecular Graphics and Modelling*, 25(2):247–260, 2006.

[81] D.A. Case, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, V.W.D. Cruzeiro, III, T.A. Darden, R.E. Duke, D. Ghoreishi, G. Giambasu, T. Giese, M.K. Gilson, H. Gohlke, A.W. Goetz, D. Greene, R. Harris, N. Homeyer, Y. Huang, S. Izadi, A. Kovalenko, R. Krasny, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, V. Man, D.J. Mermelstein, K.M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, A. Onufriev, F. Pan, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, J. Smith, J. Swails, R.C. Walker, J. Wang, H. Wei, L. Wilson, R.M. Wolf, X. Wu, L. Xiao, Y. Xiong, D.M. York, and P.A. Kollman. Amber 2019, 2019. University of California, San Francisco.

[82] A.W. Sousa da Silva and W.F. Vranken. ACPYPE – AnteChamber PYthon Parser interfacE. *BMC Research Notes*, 5(1):367, 2012.

[83] S. Nosé. Molecular-Dynamics Method for Simulations in the Canonical Ensemble. *Molecular Physics*, 52(2):255–268, 1984.

[84] W.G. Hoover. Canonical Dynamics – Equilibrium Phase-Space Distributions. *Physical Review A*, 31(3):1695–1697, 1985.

[85] M. Parrinello and A. Rahman. Polymorphic Transitions in Single-Crystals - a New Molecular-Dynamics Method. *Molecular Physics*, 52(12):7182–7190, 1981.

[86] T. Darden, D. York, and L. Pedersen. Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. *Journal of Chemical Physics*, 98(12):10089–10092, 1993.

[87] U. Essmann, L. Perera, and M.L. Berkowitz. A Smooth Particle Mesh Ewald Method. *Journal of Chemical Physics*, 103(19):8577–8593, 1995.

[88] B. Hess, H. Bekker, H.J.C. Berendsen, and J. Fraaije. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, 18(12):1463–1472, 1997.

[89] X. Daura, K. Gademann, B. Jaun, D. Seebach, W.F. van Gunsteren, and A. E. Mark. Peptide Folding: When Simulation Meets Experiment. *Angewandte Chemie International Edition*, 38(1–2):236–240, 1999.

[90] The PyMOL Molecular Graphics System, Version 1.8, 2015. Schrödinger, LLC.

[91] T. Sander, J. Freyss, M. von Korff, and C. Rufener. DataWarrior: An Open-Source Program For Chemistry Aware Data Visualization And Analysis. *Journal of Chemical Information and Modeling*, 55(2):460–473, 2015. PMID: 25558886.

[92] D. Krause and P. Thörnig. JURECA: Modular supercomputer at Jülich Supercomputing Centre. *Journal of large-scale research facilities*, 4:A132, 2018.