

Novel inhibitors of the main protease of SARS-CoV-2 identified via a molecular dynamics simulation-guided *in vitro* assay

Jennifer Loschwitz^{a,b,*}, Anna Jäckering^{a,*}, Monika Keutmann^{a,b}, Maryam Olagunju^a, Raphael J. Eberle^{a,c}, Monika Aparecida Coronado^a, Olujide O. Olubiyi^{a,d,**} and Birgit Strodel^{a,b,**}

^aInstitute of Biological Information Processing: Structural Biochemistry (IBI-7), Forschungszentrum Jülich, 52428 Jülich, Germany

^bInstitute of Theoretical and Computational Chemistry, Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany

^cMultiuuser Center for Biomolecular Innovation, IBILCE, Universidade Estadual Paulista (UNESP), São Jose do Rio Preto-SP, Brazil

^dDepartment of Pharmaceutical Chemistry, Faculty of Pharmacy, Obafemi Awolowo University, Ile-Ife, Nigeria

ARTICLE INFO

Keywords:

COVID-19

3CL^{PRO}

viral replication inhibition

MD simulations

enzyme inhibition assay

natural products

drug repurposing

ABSTRACT

For the COVID-19 pandemic caused by SARS-CoV-2, there are currently no effective drugs or vaccines to treat this coronavirus infection. In this study, we focus on the main protease enzyme of SARS-CoV-2, 3CL^{PRO}, which is critical for viral replication. We employ explicit solvent molecular dynamics simulations of about 150 compounds docked into 3CL^{PRO}'s binding site and that had emerged as good main protease ligands from our previous *in silico* screening of over 1.2 million compounds. By incorporating protein dynamics and applying a range of structural descriptors, such as the ability to form specific contacts with the catalytic dyad residues of 3CL^{PRO} and the structural fluctuations of the ligands in the binding site, we are able to further refine our compound selection. Fourteen compounds including estradiol shown to be the most promising based on our calculations were procured and screened against recombinant 3CL^{PRO} in a fluorescence assay. Eight of these compounds have significant activity in inhibiting the SARS-CoV-2 main protease. Among these are corilagin, a gallotannin, and lurasidone, an antipsychotic drug, which emerged as the most promising natural product and drug, respectively, and might thus be candidates for drug repurposing for the treatment of COVID-19. In addition, we also tested the inhibitory activity of testosterone, and our results reveal testosterone as possessing moderate inhibitory potency against the 3CL^{PRO} enzyme, which may thus provide an explanation why older men are more severely affected by COVID-19.

1. Introduction

COVID-19 infection begins with the exposure of a human host to the recently discovered SARS-CoV-2, a positive-sense RNA virus belonging to the *coronaviridae* family. The first critical stage in the development of a COVID-19 diseased state involves viral entry whereby the viral pathogen employs its surface spike glycoprotein in binding to the host's angiotensin converting enzyme 2 (ACE2) [1, 2, 3, 4]. Following invasion of the host cells, the next critical stage that is central to a successful colonization of host cellular processes and establishment of the infection is viral replication. SARS-CoV-2 employs a multiplex array of independently functional enzymatic units called the replication-transcription complex (RTC) for its replication and transcription functions. The host's ribosome is hijacked and appropriated for translating the infecting SARS-CoV-2's mRNA from which crucial viral enzymes, structurally important protein units, as well as new viral genomes needed to create functional viral units, are produced.

SARS-CoV-2's genome in its full assembly involves a replication complex featuring 14 open reading frames (ORFs) of which ORF1a encoding for polyprotein 1a (PP1a) and ORF1ab that encodes for PP1ab are integral. The genome assembly runs from an ORF1a segment at the 5' end, fol-

lowed immediately by an ORF1b section, and terminating with a segment of about 8 kb units that contains genetic codes for critical structural proteins and accessory factors at the 3' end [5]. The ORF1ab is thus a linear arrangement of the separate ORF1a and ORF1b units. Important structural proteins, including the spike protein employed in viral entry, the membrane and envelope proteins, as well as the nucleocapsid are encoded at the 3' end. At the 5' end of the genome, non-structural proteins (NSPs) involved in various enzymatic functions are encoded. A total of sixteen such NSPs (named nsp1 through nsp16) have been identified in SARS-CoV-2 and known to be responsible for different specific catalytic functions, which are crucial for a successful viral replication and transcription. The autocatalytic activity of the two important RTC proteases, the papain-like protease PL^{PRO} and the main protease 3CL^{PRO}, on the two large polyproteins is responsible for the generation of functional enzymes without which the entire replication architecture becomes nonfunctional. PL^{PRO} cleaves the polyprotein PP1ab at three positions producing nsp1, nsp2, and nsp3, while 3CL^{PRO} processes PP1ab and PP1a at eleven points to generate nsp4 to nsp16.

This makes the main protease the most important enzyme in the RTC and therefore it represents an attractive target for the development of therapeutics for the treatment of COVID-19. And indeed, since the outbreak of the disease in late 2019 and the recognition of its pandemic status, hundreds of research articles have been published exploring the possibility of evolving small molecules inhibitors of 3CL^{PRO}. Computational design involving ligand docking protocols is

*These authors contributed equally.

**Corresponding authors

✉ olubiyi00@oauife.edu.ng (O.O. Olubiyi);

b.strodel@fz-juelich.de (B. Strodel)

ORCID(s):

a significant percentage of these publications. In our first work, we conducted large-scale virtual screening of diverse compound libraries against the three-dimensional structure of 3CL^{PRO} [6]. In total we screened over 1.2 million compounds, where we incorporated protein dynamics via so-called ensemble docking [7]. Using this approach we identified important structural factors within the enzyme substrate site as well as encoded in the docked compounds that we believe are crucial for substrate recognition.

In the present work we use all-atom, explicit-solvent molecular dynamics (MD) simulations of the 147 best binding compounds docked to 3CL^{PRO} to further delineate good 3CL^{PRO} binders from poor ones regardless of initially computed affinities. As the present pandemic mandates fast action, we decided to focus our current attention on existing drugs with possible drug repurposing for the treatment of COVID-19 in mind as well as natural products, which exhibit a wide range of pharmacophores and a high degree of stereochemistry creating a great source of possible hits. A detailed analysis of the MD simulations involving geometric and energetic arguments identified 34 ligands with binding to 3CL^{PRO} predicted to be superior compared to the other compounds investigated here. From this list, 14 compounds were chosen and, for comparison with estradiol, augmented by testosterone, which were then tested for their 3CL^{PRO} inhibition capabilities using an *in vitro* assay. This step revealed eight novel non-covalent inhibitors comprising five existing drugs, four of them being approved by the FDA (U.S. Food and Drug Administration) and one being under investigation, and three natural products. Our analysis provides important information about the inhibition of the SARS-CoV-2 main protease as well as new leads for the development of treatment options against COVID-19.

2. Results and Discussion

2.1. Selection procedure

This study addresses the dynamical aspects of prospective inhibitors of the 3CL^{PRO} enzyme of SARS-CoV-2, previously identified using ensemble docking [6]. A scheme of our selection procedure employed in our previous and current work is given in Figure 1. The current work starts with 147 independent 20 ns MD simulations for ligand-main protease complexes involving the top-performing compounds identified by our previous virtual screening approach. Of this number 61 are FDA-approved drugs, 38 are drugs approved by other countries' national regulatory agencies (non-FDA) and investigational drugs (INV), 39 are natural products (NP), while the remaining nine compounds are steroids (ST), most of which also have FDA approval status. The steroids were included because some of them were identified as good binders by the initial study [6]. In consideration of the observed gender-related differences in the clinical presentation of COVID-19 [8] we additionally aim to understand differences (if any) at the steroids' level of interaction with the virus' main protease. Apart from these, we also simulated different reference compounds (REF) involving previously

identified inhibitors of the SARS-CoV-2 enzyme, including eight ligands from Jin et al. [9], and four recently identified inhibitory α -ketoamides [10]. In total, 160 compounds were considered in this work, which are listed in Table S1 and for which 20 ns MD simulations were performed. To progress the predicted inhibitors to the next simulation phase, we employed the following three selection criteria:

1. the root mean square deviation (RMSD) of the ligand (mean $\text{RMSD}_{\text{ligand}} \leq 6 \text{ \AA}$),
2. the distance to the binding site (mean and maximal $d_{\text{BS}} \leq 4 \text{ \AA}$),
3. the distance to the catalytic dyad residues H41 and C145 (mean and maximal $d_{\text{dyad}} \leq 4 \text{ \AA}$).

These selection parameters were computed for the last 5 ns of the 20 ns MD simulations to make allowance for structural adaption in the first part of the trajectories. Ninety-five of the 147 selections from our previous work as well as three of the 12 reference compounds satisfied the filtering parameters, leading to 98 compounds for which the MD simulations were subsequently extended to 100 ns. In addition, testosterone, which did not meet the selection criteria, was also simulated for 100 ns to have a representative of the male hormone system for comparison to estradiol, the only steroid satisfying the filtering parameters. The last 25 ns of each 100 ns trajectory was subjected to the same filtering criteria 1.–3. listed above, which were augmented with two additional criteria: the requirement of maximal $\text{RMSD}_{\text{ligand}} \leq 6 \text{ \AA}$ and an energy parameter that reflects the interaction between the catalytic dyad residues (H41 and C145) and the ligand requiring $E_{\text{int}} \leq -2.4 \text{ kcal/mol}$. For the initial 20 ns MD simulations, only the average $\text{RMSD}_{\text{ligand}}$ was required to be lower than 6 \AA to allow the ligands to adapt to a more energetically favorable binding pose, a phenomenon expected to temporarily increase the $\text{RMSD}_{\text{ligand}}$ values. However, after 75 ns of MD simulations we expect the strong binders to have finally adopted a stable binding pose, which is why a maximal $\text{RMSD}_{\text{ligand}}$ of 6 \AA was additionally employed as a criterion in the second set of MD simulations. Applying these five selection criteria yielded 34 compounds of the selection library as well as one reference compound meeting the filtering conditions, which were grouped based on the number of interactions with the catalytic dyad residues. More precisely, ligands interacting with both H41 and C145 were considered the preferred binders as opposed to those interacting with just one catalytic dyad residue. Finally, the total binding free energy ΔG_{bind} was computed and 14 compounds with lower ΔG_{bind} values as well as testosterone were selected for *in vitro* testing using a 3CL^{PRO} inhibition assay, which revealed eight ligands with strong inhibitory activity.

2.2. Filtering based on 20 ns MD simulations

Ligand detachment

Docking studies allow a good estimation if a compound fits well into the catalytic site of a potential drug target and therefore might be a good binder. To properly differentiate between good and poor binders using computational methods,

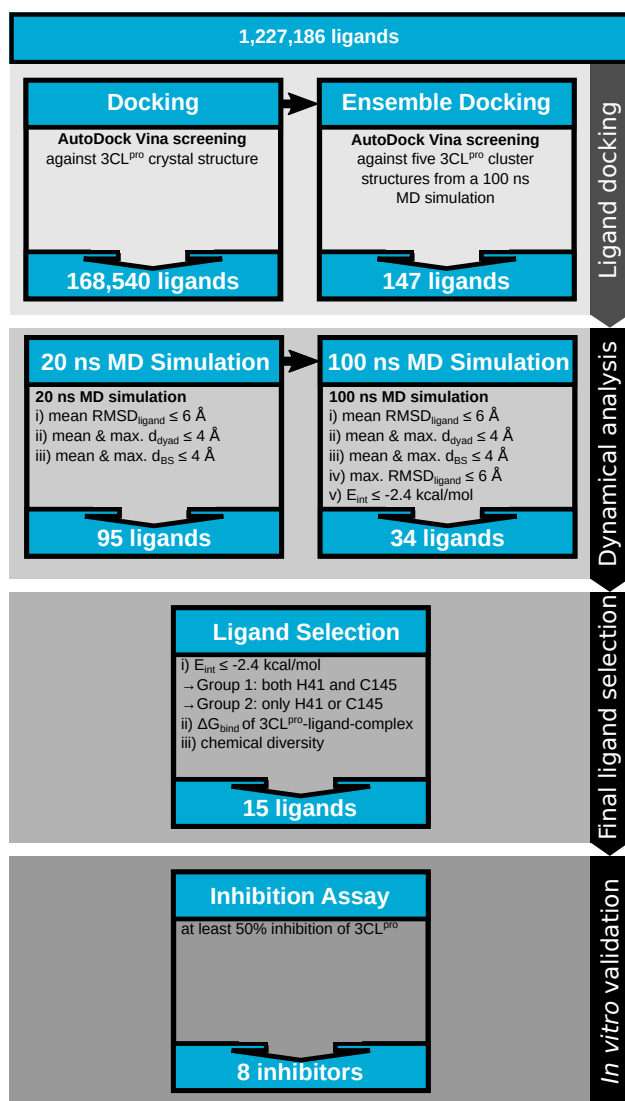


Figure 1: Scheme of the selection procedure.

not only a static analysis but also a dynamic investigation like MD simulations should be conducted [11]. It is possible that only weak interactions between the ligand and the binding site exist, which cause a ligand to leave the binding site under dynamical force. Indeed, during the 20 ns MD simulations some of the ligands display very high distances to the catalytic dyad, which raises the question of whether these ligands remain bound to the binding site. To answer this question, we calculated the distance between the center of mass (COM) of the ligand with respect to the COM of the binding site residues, where the binding site of 3CL^{pro} was defined as all amino acids within 10 Å of the covalently bound ligand N3 in the structure of the 3CL^{pro}-N3 complex [9], yielding a total of 72 residues. Based on this quantity denoted d_{COM} , the following detachment criterion was formulated: a ligand is detached when it resides at a d_{COM} above 15 Å for at least 2 ns. This criterion is applied to the whole trajectories, instead of focusing on only the last 5 ns. This analysis reveals three ligands of different classes as detaching

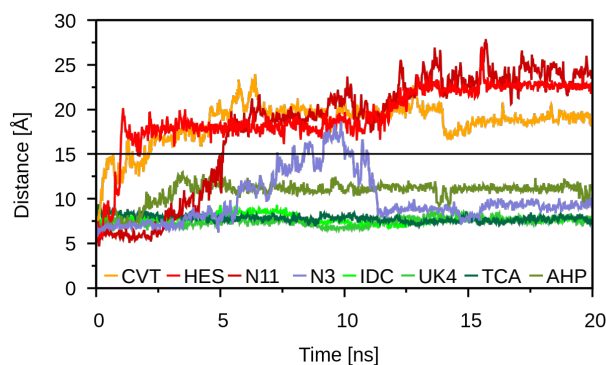


Figure 2: The distance between the COMs of the ligand and the binding site residues, d_{COM} , of selected ligands during 20 ns MD simulations. The distance is shown for the detaching ligands conivaptan (CVT, orange), 2-hydroxyestrone (HES, red) and ZINC000008764269 (N11, dark red). The cutoff distance at 15 Å is indicated by a black line. The reference compound N3 (blue) does not detach despite transient motions away from the binding site. For comparison, the distances of the ligands displaying the smallest d_{COM} in each ligand class are shown (from dark to light green): indocyanine (IDC, FDA), UK-432,097 (UK4, INV), theacitrin A (TCA, NP) and 17- α -hydroxypregnanolone (AHP, ST).

from the binding site, for which d_{COM} is plotted in Figure 2. Conivaptan (FDA) and 2-hydroxyestrone (ST) are leaving the binding site already after around 1 ns, while the d_{COM} of the natural product with the ID ZINC000008764269 remains below 15 Å for the first 5 ns before increasing beyond the cutoff value. Once detached, these compounds did not reenter the binding site during the 20 ns MD simulations. Interestingly, also the reference ligand N3 moves away from the binding site, even though not for a sufficiently long duration to have been considered a detachment event. It is important to note that N3 is a Michael acceptor inhibitor, which inactivates 3CL^{pro} irreversibly by forming a covalent bond with C145 [9]. In contrast, N3 was not covalently bound in our MD simulation setup and for this reason the nonbonded interactions seem not to be strong enough resulting in reorientation of the ligand.

For each compound class the ligand exhibiting the smallest d_{COM} values is also shown in Figure 2. Interestingly, UK-432,097 (INV), theacitrin A (NP), and 17- α -hydroxypregnanolone (ST) already appeared among the ligands with the smallest distances to the catalytic dyad residues in their respective class in our previous docking study [6]. Only indocyanine (FDA) did not stand out in the previous study, where it even displayed a higher distance to the catalytic dyad than conivaptan which detached from the binding site within 5 ns of the MD simulation.

Ligand flexibility

To further quantify the reorientation and flexibility of the ligands that remained in the binding site within the 20 ns MD simulations, we computed the RMSD_{ligand} values. High RMSD values indicate unsatisfactory binding with low affin-

ity towards the binding site. An RMSD value over 2 Å suggests a pose that is significantly different from the reference structure, which in our case was set to the starting structure of each MD production run. However, we recognize that subject to dynamical forces, there may be need for readjustment of the docking-generated binding poses to adapt to conformational changes in the binding site of 3CL^{PRO}; because of this we chose a relatively high and slightly forgiving cutoff of 6 Å for the mean RMSD_{ligand}. The values obtained for the selection library range from 1.3 Å to 25.2 Å, and from 3.3 Å to 9.7 Å for the reference ligands. Nine ligands display RMSD_{ligand} mean values over 10 Å, which include the three ligands that detached, namely conivaptan (FDA, 19.5 Å), ZINC000008764269 (NP, 25.2 Å) and 2-hydroxyestrone (ST, 21.7 Å), as well as dolutegravir (FDA, 11.0 Å), tucatinib (INV, 11.2 Å), telomestatin (INV, 11.5 Å), bavacoumestan-A (NP, 11.1 Å), allopregnanolone (ST, 10.8 Å), and androstenedione (ST, 11.2 Å). It is important to point out that all these compounds feature a rigid chemical structure with few or no rotatable bonds present. This feature most likely renders the compounds incapable of employing intramolecular readjustment for adapting to the dynamically adjusting 3CL^{PRO} binding site. Instead, the entire ligand molecules have to be involved in adjusting to the enzyme binding site with consequential increase in RMSD_{ligand}.

A significant number of the steroids (three out of nine) also show high fluctuations at the binding site, most likely due to weaker interactions compared to steroids with low RMSD_{ligand} such as estradiol (3.0 Å) or estetrol (3.4 Å) (Figure 3). The main difference between these steroids lies in their moieties at the D ring. Estradiol and estetrol contain one or three hydroxy groups at the D ring, respectively. It is possible that binding site interactions involving the D ring moieties might be impaired in androstenedione and 2-hydroxyestrone harboring carbonyl moieties or allopregnanolone having an acetyl group in their D ring. Surprisingly, cortisol, which was identified as one of the best binders among the FDA-approved compounds previously [6], is not falling below the catchment RMSD cutoff with an RMSD_{ligand} of 7.7 Å. Similar to the steroids with extremely high RMSD, the D ring of cortisol contains not only a hydroxy group but additionally an hydroxyacetone group which might hinder interactions under dynamical forces.

Besides cortisol, there are various ligands that showed great binding affinity in the initial virtual screening [6] but poor dynamical characteristics with respect to their flexibility at the binding site. These include naldemedine (6.3 Å) and enasidenib (7.8 Å) that ranked among the top binding FDA-approved drugs as well as ritonavir (6.6 Å), which is currently under investigation for a potential treatment of COVID-19 [14]. Furthermore, several non-FDA and INV ligands, which were able to surpass the best-binding FDA drug nilotinib (5.2 Å) in the docking studies [6], display here an RMSD_{ligand} above the cutoff, especially the previously mentioned INV drug telomestatin (11.5 Å), a compound with an inflexible macrocyclic core. These findings highlight the importance of fully incorporating structural dynamics while searching for

prospective inhibitors of 3CL^{PRO}. We also observed that more than half of the reference compounds surpass the RMSD_{ligand} cutoff, namely carmofur, N3, PX-12, shikonin, tideglusib and two α -ketoamides, as well as testosterone (8.5 Å).

Distance between ligand and binding site

Inhibition of 3CL^{PRO} requires interaction of the ligand with the enzyme's binding site as a *sine qua non* regardless of the nature of inhibition, competitive or irreversible. Preferably, such interactions should involve direct contacts with the catalytic dyad residues H41 and C145. To identify the compounds in close binding proximity, we set the cutoff for the mean and the maximum distance to both the binding site (d_{BS}) and the dyad (d_{dyad}) at 4 Å. These distances are defined as minimum distances between ligand and the respective group of residues over the last 5 ns of the MD simulations, yielding a maximum and a mean for them. All compounds that were not already identified as detaching before, remain associated with the enzyme with d_{BS} mean values ranging between 1.6 Å and 3.1 Å. In contrast, several of the ligands violate the averaged and/or maximal d_{dyad} cutoff values. These include the antibiotic ertapenem (FDA, maximal d_{dyad} = 4.7 Å), epigallocatechin gallate (EGCG, NP, maximal d_{dyad} = 5.1 Å), and adozelesin (INV, maximal d_{dyad} = 7.5 Å) that had represented the top drugs in their respective categories in the previous docking study [6]. This again underpins the importance of analyzing dynamical features beyond docking.

2.3. Filtering based on 100 ns MD simulations

Ligand detachment

After applying the various cutoffs using the last 5 ns of the initial 20 ns MD simulations, the remaining 95 ligands of our compound selection, 3 remaining reference compounds, as well as testosterone for comparison with estradiol and estetrol were simulated for 100 ns to detect the most strongly binding ligands. During these longer simulations another ligand (the FDA approved drug crizotinib) detached from the 3CL^{PRO} binding site after around 85 ns. The dichlorfluorophenyl moiety of this compound was highly mobile and flexible in the binding site during the MD simulation. Being solvent-exposed it was observed to leave the binding site first, following which the entire molecule also detached. For this reason, crizotinib was excluded in subsequent analysis. The fact that ligand detachment could be observed even after around 85 ns highlights the importance of a more detailed screening of the most promising ligands by extending the initial MD simulations to at least 100 ns.

Ligand flexibility and distance to the catalytic dyad

To identify the best binding ligands based on the 100 ns MD simulations, the same selection criteria as before were applied using the last 25 ns of these longer simulations. In addition, for the RMSD_{ligand} also the maximal values are now considered. In the following, the statistics provided is in relation to the total amount of ligands in the second MD selection round without the reference ligands (i.e., 100% = 95 ligands). The mean RMSD_{ligand} ranges between 1.4 Å and 12.0 Å, d_{BS}

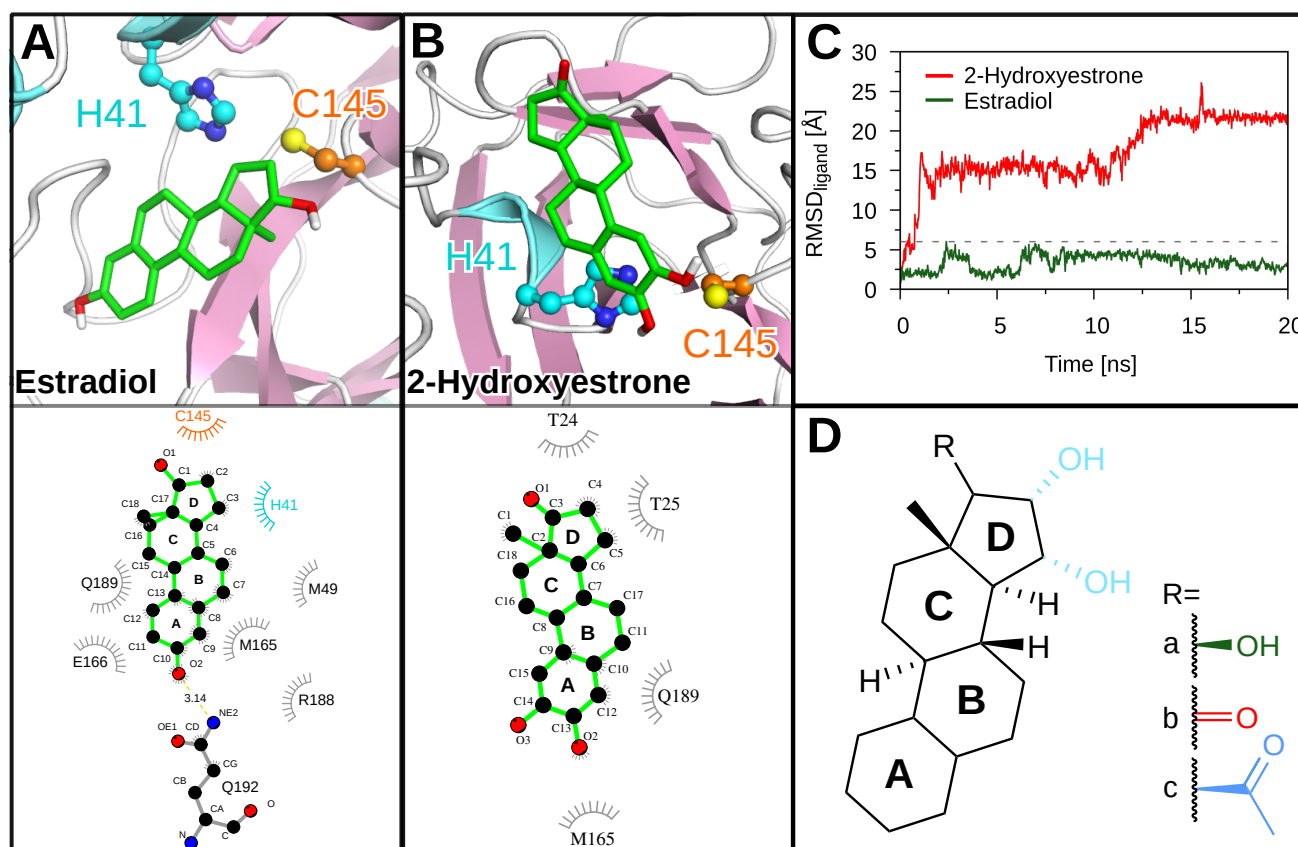


Figure 3: The binding pose and 3CL^{pro}-steroid interactions for estradiol and 2-hydroxyestrone. Cartoon representation of the binding site of 3CL^{pro} (β -sheets in lilac, α -helices in light blue) with (A, top) estradiol and (B, top) 2-hydroxyestrone bound to it. The ligands are shown as green sticks, the side chains of H41 and C145 are shown in ball-and-stick representation in cyan and orange, respectively, with the N atoms (blue), O atoms (red), and the S atom of C145 (yellow) being highlighted. (A/B, bottom) The interactions of the two ligands with the binding site were analyzed and plotted with LigPlot+ [12, 13]. Hydrogen bonds are indicated by yellow 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. (C) The RMSD_{ligand} of estradiol (green) and 2-hydroxyestrone (red) is shown over simulation time. The cutoff value at 6 Å used to distinguish between good (e.g., estradiol) and poor (e.g., 2-hydroxyestrone) binders is marked as a gray dotted line. (D) The steroid scaffold is shown with functional moieties at the D ring of some of the steroids analyzed in this study highlighted: estradiol (R=a, green), 2-hydroxyestrone and androstenedione (R=b, red), estrol (R=a plus hydroxy groups shown in light-blue), and allopregnanolone (R=c, blue).

between 1.7 Å and 2.1 Å, and d_{dyad} between 1.9 Å and 8.8 Å. All ligands stay close to the binding site but four ligands (corresponding to 4% of 95 ligands) – olsalazine (FDA, 8.2 Å), entospletinib (INV, 13.3 Å), amrubicin (non-FDA, 8.5 Å), and estrol (ST, 12.7 Å) – seem to loose contact with the catalytic dyad as they display maximal d_{dyad} values above 8 Å. The criterion of a maximum d_{dyad} below 4 Å is not met by 22 ligands (23%), most of them belonging to the FDA-approved drugs (12 ligands, 13%). Sixty ligands (63%), including 21 of the 22 ligands surpassing the maximum d_{dyad} cutoff, are very flexible as indicated by maximal RMSD_{ligand} values above 6 Å. Only deferasirox (FDA) displays a maximal d_{dyad} above the cutoff (4.6 Å) without simultaneously exceeding the maximum RMSD_{ligand} cutoff (5.6 Å). Deferasirox harbors four rotatable bonds, which enable rotation of the three aromatic rings at the central triazole and therewith adjust-

ing the position of the hydroxy and carboxy moieties at the rings. However, the overall structure is quite rigid, so that the movement of the compound away from the catalytic dyad is mainly accompanied by overall translation.

The criterion for the maximal RMSD_{ligand} cutoff is accomplished by 36 (38%) ligands. This leads to 35 ligands (37%) that meet all three criteria, while out of the 3 reference drugs and testosterone only disulfiram possesses RMSD_{ligand}, d_{BS} and d_{dyad} values below the respective cutoff. Interestingly, the two FDA-approved drugs nilotinib (maximal RMSD_{ligand} = 13.3 Å) and afatinib (maximal RMSD_{ligand} = 10.3 Å) as well as the non-FDA drug R-343 (maximal RMSD_{ligand} = 9.4 Å), which belong to the class of tyrosine kinase inhibitors that were shown to be able to inhibit related coronaviruses [15], display a maximal RMSD_{ligand} above 6 Å. These three, among others, were considered as top-performance ligands due to low binding free energies obtained from docking in the pre-

vious study [6].

Interaction of ligands with catalytic dyad residues

For further selection of the top ligands, the interaction energy E_{int} composed of Coulomb and Lennard-Jones (LJ) interactions between ligand and either catalytic dyad residue was calculated and used as an additional criterion. The requirement is that a ligand needs to interact with one or both catalytic dyad residues, H41 and C145, using the criterion $E_{\text{int}} \leq -2.4$ kcal/mol to be selected for further consideration. This removes tubocurarine (non-FDA) from the list since it does not interact strong enough with the dyad residues. Although it stays closely buried in the binding site with its benzyloquinoline ring facing the dyad residues, the methyl ether and the tertiary nitrogen are unable to constitute strong interaction partners for H41 and C145, respectively. From the remaining 34 compounds, 15 ligands interact with only one of the two catalytic dyad residues and the other 19 ligands form contacts with both H41 and C145. The only reference compound left, disulfiram, interacts with H41 and is therefore also included in the final ligand selection. To get a better view on the class composition of the final ligand selection, statistics is provided in relation to the total amount of ligands in the respective class and to the total amount of ligands subjected to MD simulation (100% = 160 ligands, including steroids and reference compounds). Accordingly, 34 ligands (21%) are identified as top binders, including 10 out of 61 FDA-approved drugs (16%, 6%), 8 out of 38 non-FDA-approved drugs (21%, 5%), 15 out of 39 natural products (38%, 9%), 1 out of 9 steroids (11%, 1%). The binding poses of these ligands in the active site of 3CL^{PRO} are shown in Figures S1–S3.

Best performing 35 compounds

After narrowing down the tremendous amount of potential SARS-CoV-2 main protease 3CL^{PRO} ligands to a sufficiently small library of 35 compounds (34 from our selection list [6] plus disulfiram as the only remaining reference compound), we analyzed the top binders further based on binding free energies, ΔG_{bind} , calculated with the molecular mechanics Poisson-Boltzmann surface area (MM/PBSA) method [16]. Here, we divided the ligands into two groups, those interacting with both catalytic dyad residues, which are considered better binders, and those interacting with only H41 or C145. Each group was then ranked using the ΔG_{bind} values, which in comparison to E_{int} not only consider the Coulomb- and LJ interactions, but also include solvent effects. For ligands interacting with both dyad residues, ΔG_{bind} ranges from -4.20 kcal/mol to -49.93 kcal/mol, while values between -12.60 kcal/mol to -42.38 kcal/mol are obtained for the compounds interacting with only one of the dyad residues (Table 1).

Statistics is generated in relation to the total number of ligands in each class as well as for all ligands selected in the top binders list (100% = 35 ligands). 14 ligands interact with only H41, while epitaraxerol and fenoverine interact with C145 only. These include five FDA compounds (8%, 14%), two non-FDA and INV ligands (5%, 6%), seven NP (18%,

20%), one ST (11%, 3%), and one REF (8%, 3%). The group of ligands binding both dyad residues includes 19 ligands in total, among these are five FDA compounds (8%, 14%), six non-FDA and INV ligands (16%, 17%), and eight NP (21%, 23%). It is particularly noticeable that many NP compounds rank among the top binders, even six exist among the top ten (Table 1). They resemble each other structurally by harboring a lot of oxygen atoms mostly in hydroxy functional groups and involving multiple ring systems. These hydroxy groups were observed to mostly orient towards the dyad residues allowing the establishment of hydrogen bonds with the catalytic dyad. Furthermore, estradiol representing the female hormonal system is found to be the only steroid present among the top 35 binders, while testosterone as a prominent male hormone violates the selection criteria.

2.4. Enzyme inhibition assay

Of the 35 compounds in the selection list (Table 1), 15 were procured and a $20 \mu\text{M}$ concentration of each was employed in an enzyme-based fluorescence assay to determine their inhibitory activities against the SARS-CoV-2 main protease enzyme at $20 \mu\text{M}$ concentration in the presence of a peptide substrate. The choice which of the compounds to test *in vitro* was to a large extent based on ΔG_{bind} , i.e., ligands with smaller binding free energy values were given preference. Moreover, compounds forming contacts with both H41 and C145 were also preferentially considered, resulting in 10 compounds from this group (cilostazol, rhoifolin, apixaban, corilagin, dihydroergotoxine, telcagepant, ZINC000011865175, lurasidone, hypericin, and proanthocyanidin A2), while dasatinib, teniposide, palbociclib, and estradiol, which were also chosen, interact with only one of the catalytic dyad residues. Other selection criteria were chemical diversity among the compounds tested as well as availability. For instance, theacitrin C, which we wanted to include in the enzyme inhibition assay, is currently not commercially available. One out of the 15 compounds tested is testosterone, which we included as comparison to estradiol despite it not making it onto the final selection list that resulted from the *in silico* screening.

In Figure 4 the results from the enzyme inhibition assay show various degrees of 3CL^{PRO} inhibition for the 15 compounds. The three compounds having zero inhibitory activity at a ligand concentration of $20 \mu\text{M}$ against the proteolytic properties of the main protease (teniposide, estradiol, and palbociclib) are from the selection group that interacts with only one catalytic dyad residue. This represents 75% of those screened from this group in the enzyme inhibitory assay. The only ligand efficiently inhibiting 3CL^{PRO} out of this group is dasatinib, which is an FDA-approved drug interacting with H41 only, provoking an activity loss of 58%. Out of the ten compounds from the selection group shown to interact with both catalytic dyad residues, seven of them (corresponding to 70%) were found to inhibit greater than 50% of the enzyme function. Of these, the natural product corilagin produces an 88% loss of the SARS-CoV-2 main protease activity, followed by 83% for ZINC000011865175, another natural product. Trailing closely behind these two are three drugs:

Table 1
Characteristics of the 35 top ligands of 3CL^{pro} identified by MD simulations. The compounds selected for validation in a 3CL^{pro} inhibition assay are highlighted in bold.

| No. | Compound name | Class | d_{dyad} [Å] | d_{BS} [Å] | ΔG_{bind} [kcal/mol] |
|---|----------------------------|----------------|-----------------------|---------------------|-------------------------------------|
| Interaction with both catalytic dyad residues (H41 and C145) | | | | | |
| 1 | Cilostazol | FDA | 2.295 | 1.856 | -49.9 ± 4.8 |
| 2 | Proanthocyanidin A1 | NP | 2.151 | 1.696 | -45.4 ± 5.9 |
| 3 | Rhoifolin | NP | 2.101 | 1.745 | -44.1 ± 7.4 |
| 4 | Apixaban | FDA | 2.511 | 1.901 | -43.4 ± 5.2 |
| 5 | Theacitrin C | NP | 2.278 | 1.660 | -42.9 ± 7.1 |
| 6 | Corilagin | NP | 2.134 | 1.667 | -41.5 ± 5.1 |
| 7 | Dihydroergotoxine | non-FDA | 2.456 | 1.810 | -40.8 ± 5.8 |
| 8 | Telcagepant | INV | 2.665 | 1.931 | -39.1 ± 4.0 |
| 9 | ZINC000002114470 | NP | 2.479 | 1.817 | -37.3 ± 5.6 |
| 10 | ZINC000011865175 | NP | 2.046 | 2.012 | -36.7 ± 3.5 |
| 11 | Lurasidone | FDA | 2.153 | 1.963 | -35.6 ± 3.4 |
| 12 | ZINC000049888572 | INV | 2.109 | 1.930 | -33.5 ± 5.1 |
| 13 | Hypericin | INV | 2.410 | 1.865 | -31.8 ± 3.8 |
| 14 | Pimozide | FDA | 2.382 | 1.928 | -30.6 ± 3.5 |
| 15 | Sotrastaurin | INV | 2.143 | 1.911 | -29.9 ± 3.6 |
| 16 | Proanthocyanidin A2 | NP | 2.584 | 1.845 | -29.0 ± 5.4 |
| 17 | Enzastaurin | INV | 2.545 | 1.910 | -27.3 ± 4.0 |
| 18 | ZINC000008297065 | NP | 2.030 | 1.992 | -27.0 ± 4.4 |
| 19 | Telmisartan | FDA | 2.360 | 1.908 | -4.2 ± 6.1 |
| Interaction with one catalytic dyad residues (H41 or C145) | | | | | |
| 20 | Isocorilagin | NP | 2.250 | 1.683 | -42.4 ± 6.0 |
| 21 | Dasatinib | FDA | 2.361 | 1.881 | -37.4 ± 5.4 |
| 22 | Teniposide | FDA | 2.232 | 1.796 | -36.2 ± 6.4 |
| 23 | Palbociclib | FDA | 2.347 | 1.942 | -35.6 ± 3.4 |
| 24 | Tadalafil | FDA | 2.333 | 1.938 | -33.9 ± 3.2 |
| 25 | ZINC000012881832 | NP | 2.153 | 1.984 | -33.5 ± 6.0 |
| 26 | TMC647055 | INV | 2.336 | 2.083 | -31.8 ± 4.9 |
| 27 | Fenoverine | INV | 2.281 | 2.005 | -31.5 ± 3.3 |
| 28 | Zeylanone | NP | 2.827 | 1.954 | -29.5 ± 3.7 |
| 29 | Remdesivir | NP | 2.421 | 1.843 | -29.0 ± 5.4 |
| 30 | Estradiol | ST | 2.395 | 1.880 | -26.5 ± 3.9 |
| 31 | Disulfiram | REF | 2.317 | 2.102 | -24.1 ± 2.8 |
| 32 | Epitaraxerol | NP | 2.263 | 2.049 | -23.7 ± 3.1 |
| 33 | Theaflavin | NP | 2.482 | 1.847 | -23.5 ± 4.9 |
| 34 | Daidzein | NP | 2.504 | 1.760 | -22.9 ± 3.5 |
| 35 | Montelukast | FDA | 2.186 | 1.896 | -12.6 ± 7.5 |

lurasidone, an antipsychotic FDA-approved drug with 79% inhibition, and cilostazol and telcagepant both displaying 68% main protease inhibition. Cilostazol inhibits the phosphodiesterase 3 and platelet aggregation and has been approved for peripheral vascular disease, while telcagepant was initially being developed as a treatment for migraine. After these, rhoifolin, a natural product and apixaban, an FDA-approved drug, were found to inhibit the enzyme by 64% and 59%, respectively. Interestingly, the anticoagulant apixaban was found to reduce the mortality from COVID-19 by about 50% in a therapeutic dose (data from Mount Sinai Hospital, New York, USA) and the flavonoid rhoifolin has been previously proposed to be an efficient inhibitor of 3CL^{pro} [17]. Comparing the 3CL^{pro} activity of estradiol and testosterone, one can see that 3CL^{pro} loses about 30% of its activity in the presence of testosterone but is fully active in the presence of estradiol. On account of this, we can exclude a potential

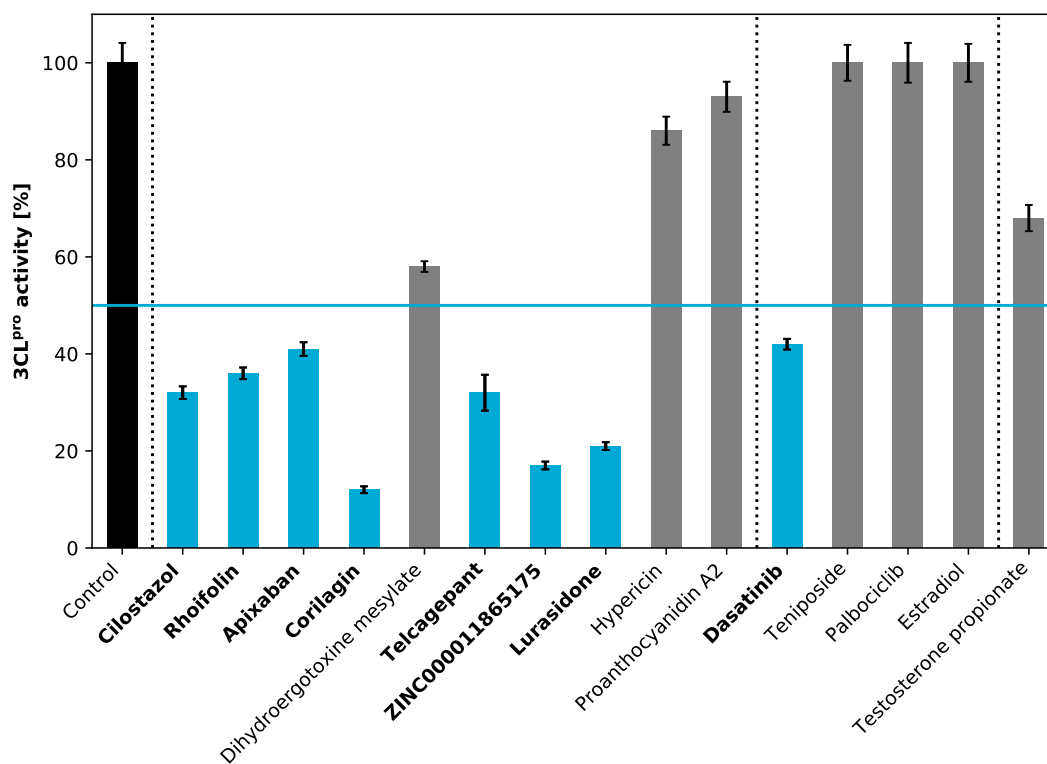


Figure 4: Activity of 3CL^{pro} in the presence of the 15 tested compounds for their inhibition capacity. The activity of 3CL^{pro} is provided relative to the control sample (black) without additional ligand. The group of compounds displaying interactions with both catalytic dyad residues is separated from the one with only one such interaction by a dashed vertical line and the compounds in each group are sorted by their ΔG_{bind} values. Testosterone not belonging to the top 35 ligands is shown separately. The bars for compounds with a residual 3CL^{pro} activity above 50% are colored in gray, while those with a residual activity below 50% are highlighted in blue with the compound names written in bold. A ligand concentration of 20 μM was used for the inhibition test, which was performed in triplicate per ligand. The standard deviation is shown as error bar.

3CL^{pro} inhibition activity of the female hormone estradiol as the cause behind the different COVID-19 progression between men and women [8]. In total, eight compounds could successfully be identified as potent 3CL^{pro} inhibitors with more than 50% inhibition at 20 μM concentration comprising five FDA-approved drugs, two natural products, and one investigational drug. The binding poses of the top eight inhibitors are shown in Figure 5, and their chemical structures can be seen in Figure S4.

2.5. Top eight inhibitors

Interactions with 3CL^{pro}

It is interesting to note that all eight compounds found to inhibit 3CL^{pro} in our *in vitro* assay passed all employed filtering criteria and were additionally found to be capable of forming specific interactions with both catalytic dyad residues as well as establishing hydrogen bonding networks with other binding site amino acids (Figure S5). A difference in the classes can be observed as the three natural products corilagin, rhoifolin and ZINC000011865175 contain many oxygen containing functional groups, while the FDA-approved drugs apixaban, dasatinib, cilostazol and lurasidone as well as the investigational compound telcagepant contain more functional moieties comprising nitrogen. We analyzed the in-

teractions with binding site residues using LigPlot+ [12, 13] and extracted all residues exhibiting contacts, accounting specifically for hydrogen bonding with one or several of the eight 3CL^{pro} inhibitors (Table 2). Here, residues interacting with several ligands and those being involved in hydrogen bonding are considered as more important for inhibition than those not fulfilling these two criteria.

The catalytic dyad residues H41 and C145 form hydrogen bonds exclusively with ZINC000011865175 but have less strong interactions with the other inhibitors (Figure S5). While C145 interacts with all ligands, no interaction can be observed between H41 and rhoifolin. Besides C145, also M165, R188, and Q189 interact with all eight ligands. The latter two form hydrogen bonds with three ligands each, but similar to the catalytic dyad mainly weaker interactions with the other ligands. No hydrogen bond formation but other contacts with all eight ligands is recorded for M165. Also G143, H164, E166, and N142 are suggested as important residues for inhibitor binding forming both hydrogen bonds and other contacts with the latter as the main interaction type. D187 establishes contacts but no hydrogen bonds with five ligands, while S144 exclusively forms hydrogen bonds with half of the ligands. Except from G143 and M165, these residues are all polar or charged and located either at the entrance to or at

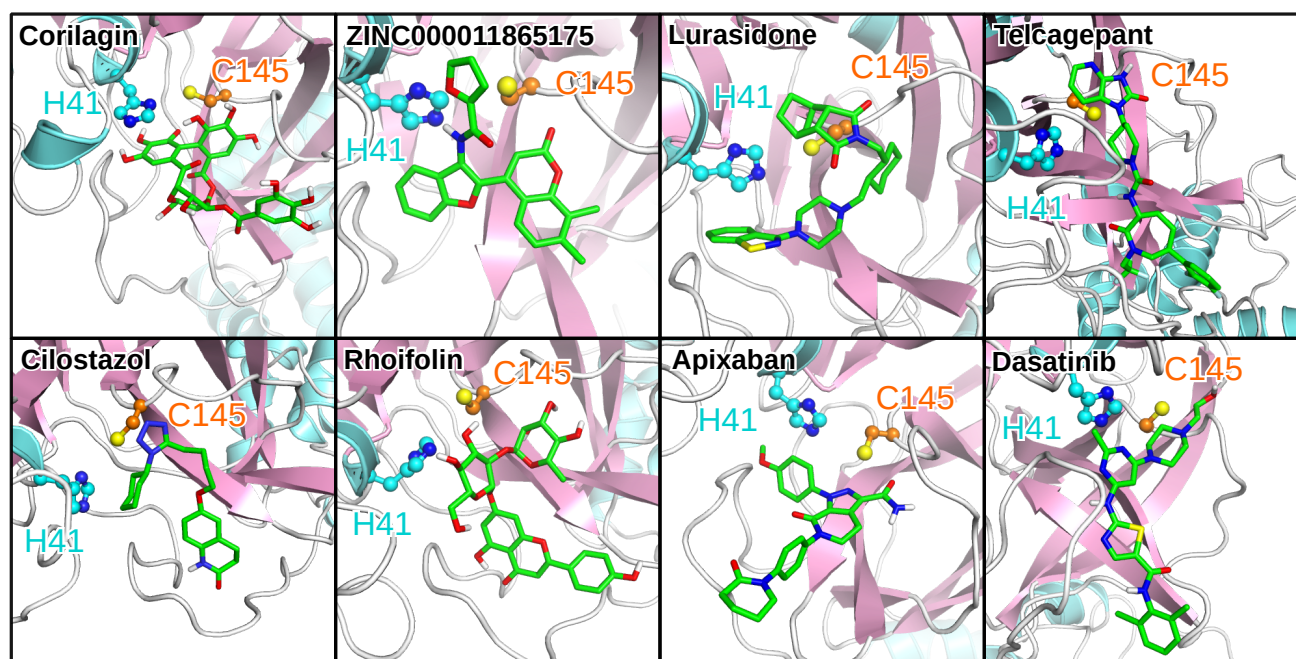


Figure 5: The binding poses of the top eight binders. The top eight compounds identified by an enzyme inhibition assay following *in silico* screening are shown in the order of decreasing inhibitory activity: corilagin (NP), ZINC000011865175 (NP), lurasidone (FDA), telcagepant (INV), cilostazol (FDA), rhoifolin (NP), apixaban (FDA), and dasatinib (FDA). The same protein and ligand representations as in Figure 3 are used.

Table 2

Binding site residues interacting with the top eight inhibitors. The number of ligands interacting with the different binding site residues via hydrogen bonds exclusively or via any kind of interactions (including hydrogen bonds) is listed.

| Number of ligands involved | Hydrogen bonds | All interactions |
|----------------------------|------------------------------------|--------------------------------------|
| 1 | H41, C145, E166, V186 | L27, S46, L50, Y54, L167, T169, V186 |
| 2 | L141, N142, G143, H164, Q189, T190 | T25, F140, L141, H172, A191 |
| 3 | R188, Q192 | T26, H163, Q192 |
| 4 | S144 | M49, S144, P168, T190 |
| 5 | — | N142, D187 |
| 6 | — | G143, H164 |
| 7 | — | H41, E166 |
| 8 | — | C145, M165, R188, Q189 |

the inside of the binding site (Figure 6). The latter might be responsible for tight inhibitor binding and the former might hinder ligand detachment due to a closure of the binding pocket, together accounting for the high $3CL^{pro}$ inhibition activity. Hydrophobic residues interacting with few of the ligands are located rather at the surface close to the binding site and might be responsible for initial recruitment of binders to the active site.

In order to better quantify the hydrogen bonding between the ligands and $3CL^{pro}$, the average number of hydrogen bonds formed with all the binding site residues during the 100 ns MD simulations were calculated for the top 35 ligands identified *in silico* (Figure S6). Except from ZINC000011865175, lurasidone, and telcagepant, the top eight inhibitors belong to the ligands developing on average the most hydrogen bonds, between 2.75 and 5.00. From the other ligands in-

teracting with both catalytic dyad residues, four ligands – proanthocyanidin A1, theacitrin C, dihydroergotoxine, and proanthocyanidin A2 – form a similar amount of hydrogen bonds, while of those interacting solely with one catalytic dyad residue only two ligands – isocorilagin and remdesivir – exceed 2.75 hydrogen bonds on average. It is noticeable that besides the top eight inhibitors, the compounds considered as the best binders as judged by ΔG_{bind} show the most pronounced hydrogen bond formation with binding site residues. Thus, hydrogen bond formation considerably contributes to the binding free energy. Ligands, which were excluded from the list of potential $3CL^{pro}$ inhibitors by the activity assay such as hypericin, teniposide, palbociclib, and estradiol are found to form less than 2.75 hydrogen bonds on average. Therefore we can conclude that a ligand being able to develop many hydrogen bonds with binding site residues

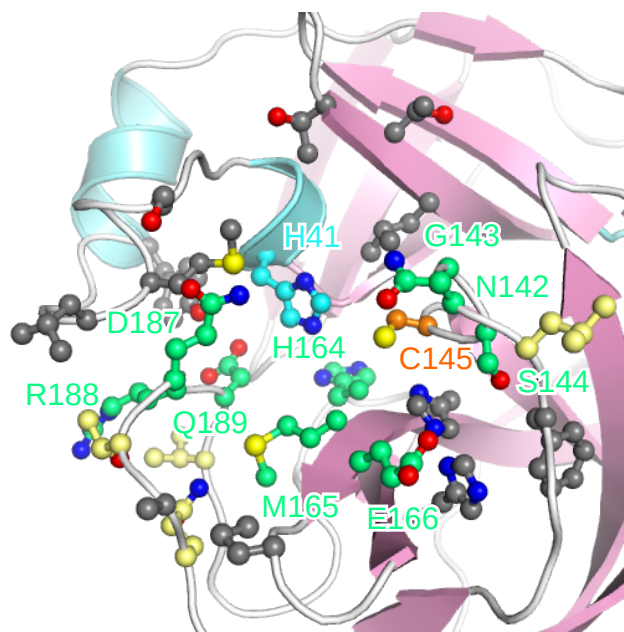


Figure 6: The binding site with the residues predominantly interacting with the top eight inhibitors highlighted. The residues showing the most pronounced interactions are colored in green and those forming hydrogen bonds are emphasized by a yellow color, while all other interacting residues are shown in gray.

is sought to be important for its strong binding and potential inhibition activity. However, some of the ligands like ZINC000011865175, lurasidone, and telcagepant as well as testosterone form on average less than 2.75 hydrogen bonds (Figure S6), yet nonetheless are able to inhibit 3CL^{pro}, in the case of the former three even strongly. Testosterone formed on average only 0.63 hydrogen bonds and also failed all other selection criteria, but surprisingly reduces the activity of 3CL^{pro} by 32%. Testosterone may thus be an endogenous inhibitor of the SARS-CoV-2 main protease and help to attenuate a COVID-19 infection. This would contribute to an explanation for the age-dependent severity of COVID-19, given the positive correlation between serum testosterone levels, disease progression and clinical outcomes in male Covid-19 patients, independent of patient age and comorbidities [18].

Distance to the catalytic dyad residues

In addition to hydrogen bonds, we analyzed various other properties of the ligands, such as the number of aromatic rings and rotatable bonds, as well as their interactions with 3CL^{pro} and plotted the resulting values against the experimentally obtained inhibition data to identify further characteristics contributing to the inhibitory activities. A fairly good correlation is obtained between experimental data and d_{dyad} suggesting higher protease inhibition at closer distances from the catalytic dyad (Figure 7). Comparing the chemical structures of the top eight inhibitors (Figure S4) and the other experimentally addressed ligands, a decent number of ring systems can be found as a common feature among the ligands. Fur-

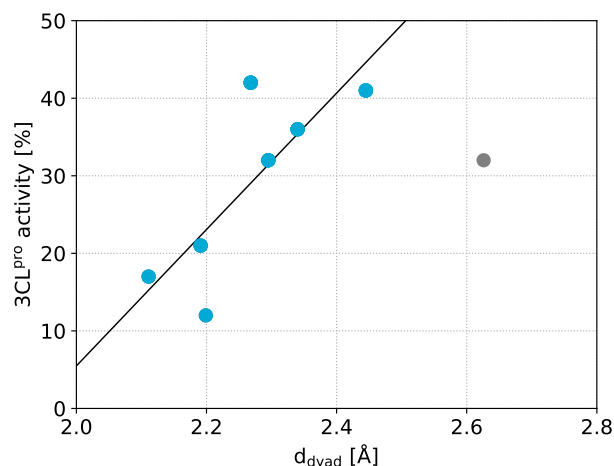


Figure 7: Correlation between the distance to the catalytic dyad (d_{dyad}) and the experimentally determined residual activity of 3CL^{pro} for the top eight inhibitors. A linear regression without the outlier telcagepant colored in gray shows that lower d_{dyad} correlates with stronger 3CL^{pro} inhibition.

thermore, good inhibitors seem to need a certain flexibility to adapt a strong binding position, which is not the case for e.g. the large and rigid hypericin and the tested steroids. It would be interesting to extend the experimental testing to ligands exhibiting descriptors found to be important, i.e. those with small d_{dyad} values and those with pronounced hydrogen bonding like proanthocyanidin A1 and theacitrin C, which were not experimentally investigated in the first testing round. This includes remdesivir, which according to our *in silico* screening could act as a 3CL^{pro} inhibitor, while based on cell assays it is suggested to interfere with the RNA polymerase, i.e., nsp12 of SARS-CoV-2 [19, 20]

3. Conclusion

By carefully combining computational investigation methods with an enzyme inhibition assays we have identified eight chemical compounds that showed inhibitory activity in the low micromolar range against the main protease enzyme central to viral replication in SARS-CoV-2. With functional groups known for covalent attack absent in the molecular structure of these eight compounds, a non-covalent inhibitory mechanism is strongly implicated. Of the eight identified inhibitors, four are approved drugs, one an investigational drug, and three are natural products, a feature that will likely shorten the time to clinical availability following successful completion of relevant testing proceedings. We expect that the findings presented in this work will be of importance in the development of novel therapeutics for managing COVID-19. In addition, we have presented important structural features that appear to underly 3CL^{pro} enzyme inhibition as well as surprising inhibition features for hormonal steroids involved in gender-based response to COVID-19 infection. While estradiol was found to be devoid of inhibitory activity against SARS-CoV-2's main protease enzyme, testosterone

was discovered to possess relatively strong inhibitory activity against the enzyme. This observation may explain, at least in parts, why obese and/or elderly men with expected reduced testosterone titres appear to be more vulnerable to the infection compared with males with higher serum testosterone levels. Further analyses are required to explore the full implications and ramifications of these findings both for the development of COVID-19 treatment and for understanding hormonal involvements in host response to SARS-CoV-2 infection.

4. Methods and Materials

4.1. Molecular dynamics simulations

Simulation flow

In our previous study [6], we docked FDA-approved drugs, non-FDA and investigational drugs as well as natural products to the main protease 3CL^{pro} binding site of the SARS-CoV-2 virus. In the first screening step, we screened 1,227,186 ligands against the 3CL^{pro} crystal structure (PDB code 6LU7) [9]. To consider protein flexibility, ensemble docking [7] was performed with the 168,540 best performing compounds from the first screening (Figure 1). To this end, five different 3CL^{pro} conformers were taken from a 500 ns MD simulation of the protein with the N3 ligand bound to it. In order to select the ligands for the MD simulations for testing their stability in the binding site, we sorted the ligands based on the binding free energies predicted by docking and using the cutoff $\Delta G_{\text{docking}} \leq -7.5$ kcal/mol and the distances to the catalytic dyad residues with cutoff $d_{\text{dyad}} < 3.5$ Å. For each of the resulting 147 compounds we performed a 20 ns MD simulation of the 3CL^{pro}-ligand complex using the crystal structure of the enzyme. Additionally, 20 ns MD simulations were performed for 13 other ligands of interest in connection with 3CL^{pro} (Table S1) [9, 10]. For 99 of these ligands the MD simulations were extended to 100 ns using the selection criteria explained in section 2.1.

Parameterization of the ligands

Before the MD simulations could be started, generalized AMBER force field (GAFF) parameters [21] for the 160 ligands considered had to be derived. To this end, quantum mechanics calculations at the HF6-31G* level were performed using Gaussian 09 [22], followed by restrained electrostatic potential (RESP) calculations for determining partial charges [23, 24] via Antechamber [21, 25] as available in Amber-Tool 19 [26]. The GROMACS input files we then generated with the ACPYPE tool [27].

Simulation details

All MD simulations were performed with GROMACS 2018 [28]. We used AMBER14SB [29] with Parmbsc1 parameters [30] as protein force field combined with the TIP3P water model [31] to explicitly simulate water. The, 3CL^{pro}-ligand complexes were centered in a cubic box of size $80 \times 80 \times 80$ Å³, solvated with water, and Na⁺ and Cl⁻ added at a concentration of 150 mM while at the same time neutralizing the system. This results in a system size of

≈51,000 atoms in total. The energy of the systems was minimized via the steepest descent algorithm [32]. Afterwards, the systems were equilibrated, first in the *NVT* ensemble (i.e., with a constant number of molecules, volume, and temperature) for 0.1 ns and second for 1 ns in the *NpT* ensemble at 310 K (37 °C, Nosé-Hoover thermostat [33, 34]) and 1.0 bar (Parrinello-Rahman barostat [35]). The production runs of 20 ns or 100 ns lengths used the same settings as the *NpT* equilibration runs. Electrostatic interactions were processed with the particle-mesh Ewald method [36, 37] in combination with periodic boundary conditions and a real-space cutoff of 12 Å. The Lennard-Jones (LJ) interactions were also cut at 12 Å. For the integration of the equations of motion, a leapfrog stochastic dynamics integrator was used with a time step of 2 fs. The LINCS algorithm [38] was applied to constrain all bond lengths during the MD simulations. The coordinates were saved every 20 ps.

Analysis

As explained in the Results and Discussion section, various quantities were calculated for each ligand to determine the stability of the compounds in the binding site of 3CL^{pro}. One of these quantities is the distance d_{dyad} , which defines the minimum distance between the catalytic dyad residues (H41 and C145) and ligand and was calculated with the GROMACS tool *gmx mindist*. Another distance that was determined is the distance d_{BS} , which is the minimum distance between the binding site and ligand. Here, the binding site was defined as the collection of 72 residues that reside within 10 Å around the ligand N3 in the crystal structure of the 3CL^{pro}-N3 complex (PDB code 6LU7) [9]. Also this distance was calculated with *gmx mindist*. To identify the ligands detaching from the binding site, we defined a third distance called d_{COM} which measures the distance of the center of mass of the binding site residues to the center of mass of the ligand in question. This distance was computed with *gmx distance*. We further determined the root mean square deviation (RMSD) of each ligand, called $\text{RMSD}_{\text{ligand}}$ here, which indicates how flexible a ligand is in the binding site. To this end, we aligned the protein structures sampled during the MD simulations to the MD starting structure (excluding the ligand during the alignment) and then calculated the RMSD for the ligand using the *gmx rms* tool.

To quantify the strength of the 3CL^{pro}-ligand interactions, their interaction energy $E_{\text{int}} = E_{\text{Coul}} + E_{\text{LJ}}$ consisting of Coulomb and LJ contributions was determined, which was accomplished by rerunning the simulation using *gmx mdrun -rerun* to obtain the energies, which were processed using *gmx energy* to calculate E_{Coul} and E_{LJ} between the ligand and the catalytic dyad residues H41 and C145. For the 35 best ligands identified *in silico*, we also computed the binding free energy ΔG_{bind} using the MM/PBSA method as implemented in *g_mmpbsa* (https://rashmikumari.github.io/g_mmpbsa/) [16]. This analysis was applied to 626 MD snapshots sampled every 40 ps between 75 ns and 100 ns of the MD simulations. Within the MM/PBSA scheme the

binding free energy is defined as

$$\Delta G_{\text{bind}} = \langle G_{\text{complex}} - G_{\text{protein}} - G_{\text{ligand}} \rangle \quad (1)$$

where $\langle \cdot \rangle$ indicates the average over the 626 snapshots in the current case. The free energy for each of these three entities is given as

$$G = E_{\text{bonded}} + E_{\text{Coul}} + E_{\text{LJ}} + G_{\text{polar}} + G_{\text{nonpolar}} - TS \quad (2)$$

where E_{bonded} describes the bonded interactions, which is like E_{Coul} and E_{LJ} obtained from the force field, G_{polar} and G_{nonpolar} are the polar and nonpolar contributions to the solvation free energy, and the last term is the absolute temperature, T , multiplied by the configurational entropy, S , which can be estimated by a normal-mode analysis of the vibrational frequencies. However, this entropy term is not calculated by *g_mmpbsa*. The polar energy term G_{polar} is obtained by solving the Poisson-Boltzmann equation, whereas the nonpolar term G_{nonpolar} is estimated from a linear relation to the solvent accessible surface area (SASA). The parameters for the calculation of ΔG_{bind} were set as $T = 310 \text{ K}$ ($37 \text{ }^\circ\text{C}$), $D_{\text{solvent}} = 80$ for the dielectric constant of the solvent (corresponding to water), $D_{\text{solute}} = 2$ for the dielectric constant of the solute (corresponding to a globular protein), $\gamma = 0.0226778 \text{ kJ}/(\text{mol}\cdot\text{\AA}^2)$ for the surface tension, $\text{sasrad} = 1.4 \text{ \AA}$ as probe radius for the SASA calculation. ΔG_{bind} was further decomposed into its per-residue contributions to determine the interaction strength with the catalytic dyad (H41/C145) or other residues in the binding site. This was accomplished with the Python script *MmPbSaDecomp.py*, while the script *MmPbSaStat.py* was used for the calculation of ΔG_{bind} . Both scripts are provided via the *g_mmpbsa* website.

The energetic analysis of the 3CL^{pro}-ligand interactions was augmented by an analysis of the hydrogen bonds formed between both entities. This was accomplished with *gmx hbond* applied to the binding site and ligand as interaction partners and using a cutoff of 3.5 \AA for the distance between hydrogen donor and acceptor and a maximal allowed deviation from linearity of 30° .

Visualization

The protein-ligand systems were visualized using the PyMOL software [39]. For the selection of the binding poses shown in this manuscript, the ligand structures and orientations were clustered using the algorithm of Daura et al. [40] as implemented in *gmx cluster*. A cutoff of 2 \AA was used for the clustering and the most populated cluster per ligand chosen for visualization. Interactions between protein and ligands were analyzed and plotted using LigPlot+ [12, 13]. Results were plotted using the Gnuplot software [41] and Python3 [42].

4.2. In vitro testing

Cloning, expression and purification of SARS-CoV-2 3CL^{pro}

The codon optimized cDNA encoding SARS-CoV-2 3CL^{pro} (Uniprot entry: P0DTD1) was synthesized and implemented

in the ampicillin resistant vector pGEX-6P-3 (BioCat GmbH, Heidelberg, Germany). The construct contains an N-terminal GST-tag and a PreScission protease cleavage site (LEFLFQGP).

SARS-CoV-2 3CL^{pro}-pGEX-6P-3 vectors were transformed into *E. coli* Lemo21 (DE3) (New England BioLabs, USA) competent cells and grown overnight at $37 \text{ }^\circ\text{C}$ in LB-medium. This pre-culture was added to fresh LB-medium (Ampicillin and Chloramphenicol) and grew at $37 \text{ }^\circ\text{C}$ until the cells reached an OD₆₀₀ of 0.6. Gene expression was induced with IPTG at final concentration of 0.5 mM (1 mM Rhamnose was added) and incubated for 3 h, at $37 \text{ }^\circ\text{C}$ and 120 rpm. Subsequently, the culture was harvested by centrifugation ($4,000 \text{ rpm}$) at $5 \text{ }^\circ\text{C}$ for 20 min (Sorvall RC-5B Plus Super-speed Centrifuge, Thermo Fisher Scientific, USA; GSA rotor). The supernatant was discarded and the cells containing the recombinant SARS-CoV-2 3CL^{pro}-GST was resuspended in 50 mM Tris-HCl pH 8.0, 200 mM NaCl (lysis buffer) and stored at $-20 \text{ }^\circ\text{C}$ for subsequent purification.

For purification, the cell-suspension was incubated on ice for 1h with addition of lysozyme, subsequently it was lysed by sonication in four pulses of 30 s each with amplitude of 30% interspersed by intervals of 10 s. The crude cell extract obtained was centrifuged ($7,000 \text{ rpm}$ for 90 min at $6 \text{ }^\circ\text{C}$). The supernatant containing SARS-CoV-2 3CL^{pro}-GST was loaded onto a GSH-Sepharose matrix which was previously equilibrated with the lysis buffer and was extensively washed with the same buffer. The protein was eluted with the same buffer plus addition of 10 mM GSH. The eluted fractions were concentrated and dialyzed against PreScission protease cleavage buffer (50 mM Tris pH 7.0, 200 mM NaCl, 1 mM DTT and 1 mM EDTA). PreScission protease was used to cleave the GST-tag from the SARS-CoV-2 3CL^{pro}-GST protein. For $100 \text{ }\mu\text{g}$ target protein, $10 \text{ }\mu\text{g}$ PreScission protease was added and the sample incubated for 36 h at $4 \text{ }^\circ\text{C}$. For separation of the target protein, the GST-tag and the PreScission protease was achieved using GSH-Sepharose. Further, to remove aggregated fraction, size exclusion chromatography was used (Superdex 200 10/300 GL GE Healthcare, USA), the column was equilibrated with 20 mM Tris-HCl pH 8.0, 150 mM NaCl. Sample purity after each purification step was assessed by 15% SDS-PAGE gels. The corresponding protein fraction was concentrated up to 2 mg/mL and stored at $-20 \text{ }^\circ\text{C}$.

Activity assay of SARS-CoV-2 3CL^{pro}

SARS-CoV-2 3CL^{pro} activity assay was performed as described earlier using a fluorogenic substrate DABCYL-KTSAVLQ↓SGFRKME-EDANS (Bachem, Switzerland) in a buffer containing 20 mM Tris pH 7.2, 200 mM NaCl, 1 mM EDTA and 1 mM TCEP [43, 10, 44]. The reaction mixture was pipetted in a Corning 96-Well plate (Sigma Aldrich) consisting of $0.5 \text{ }\mu\text{M}$ protein and the assay was initiated with the addition of the substrate at a final concentration of $50 \text{ }\mu\text{M}$.

The inhibitory potential against the SARS-CoV-2 3CL^{pro} activity of the best compounds identified in the virtual screening was investigated using the activity assay described above. $20 \text{ }\mu\text{M}$ of the compounds was used for the screening tests.

The mixtures were incubated for 30 min at RT. When the substrate with a final concentration of 50 μ M was added to the mixture, the fluorescence intensities were measured at 60 s intervals over 30 min using an Infinite 200 PRO plate reader (Tecan, Männedorf, Switzerland). The temperature was set to 37 °C. The excitation and emission wavelengths were 360 nm and 460 nm, respectively. Inhibition assays were performed as triplicates.

5. Funding sources

The authors gratefully acknowledge the computing time granted through JARA-HPC (project COVID19MD) on the supercomputer JURECA at Forschungszentrum Jülich [45], the hybrid computer cluster purchased from funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) project number INST 208/704-1 FUGG, and the Centre for Information and Media Technology at Heinrich Heine University Düsseldorf. R.J.E. recognizes with appreciation funding from FAPESP [grant numbers 2018/07572-3, 2019/05614-3].

6. Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

7. Acknowledgements

References

- [1] J. Luan, Y. Lu, X. Jin, L. Zhang, Spike protein recognition of mammalian ace2 predicts the host range and an optimized ace2 for sars-cov-2 infection., *Biochem. Biophys. Res. Commun.* (2020).
- [2] D. Wrapp, N. Wang, K. Corbett, J. Goldsmith, C. Hsieh, O. Abiona, B. Graham, J. McLellan, Cryo-em structure of the 2019-ncov spike in the prefusion conformation., *Science* 367 (2022) 1260–1263.
- [3] S. Xia, Y. Zhu, M. Liu, Q. Lan, W. Xu, Y. Wu, T. Ying, S. Liu, Z. Shi, S. Jiang, et al., Fusion mechanism of 2019-ncov and fusion inhibitors targeting hr1 domain in spike protein., *Cell. Mol. Immunol.* (2020).
- [4] M. Hoffmann, H. Kleine-Weber, S. Schroeder, N. Krüger, T. Herrler, S. Erichsen, T. Schiergens, G. Herrler, N. Wu, A. Nitsche, et al., Sars-cov-2 cell entry depends on ace2 and tmprss2 and is blocked by a clinically proven protease inhibitor., *Cell* (2020).
- [5] M. Romano, A. Ruggiero, F. Squeglia, G. Maga, R. Berisio, A structural view of sars-cov-2 rna replication machinery: Rna synthesis, proofreading and final capping., *Cell* 9 (2020) 1267.
- [6] O. Olubiyi, M. Olagunju, M. Keutmann, J. Loschwitz, B. Strodel, High throughput virtual screening to discover inhibitors of the main protease of the coronavirus sars-cov-2, *Molecules* (2020) 3193.
- [7] R. Amaro, J. Baudry, J. Chodera, O. Demir, J. McCammon, Y. Miao, J. Smith, Ensemble Docking in Drug Discovery, *Biophys. J.* 114 (2018) 2271–2278.
- [8] J.-M. Jin, P. Bai, W. He, F. Wu, X.-F. Liu, D.-M. Han, S. Liu, J.-K. Yang, Gender differences in patients with covid-19: Focus on severity and mortality, *Front. in Public Health* 8 (2020) 152.
- [9] 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, H. Yang, Structure of M^{pro} from COVID-19 virus and discovery of its inhibitors, *Nature* (2020).
- [10] L. Zhang, D. Lin, Y. Kusov, Y. Nian, Q. Ma, J. Wang, A. von Brun, P. Leyssen, K. Lanko, J. Neyts, A. de Wilde, E. J. Snijder, H. Liu, R. Hilgenfeld, α -Ketoamides as Broad-Spectrum Inhibitors of Coronavirus and Enterovirus Replication: Structure-Based Design, Synthesis, and Activity Assessment, *J. Med. Chem.* 63 (2020) 4562–4578.
- [11] V. Salmaso, S. Moro, Bridging Molecular Docking to Molecular Dynamics in Exploring Ligand-Protein Recognition Process: An Overview, *Front. Pharmacol.* 9 (2018) 923.
- [12] A. Wallace, R. Laskowski, J. Thornton, LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions, *Protein Eng. Des. Sel.* 8 (1995) 127–134.
- [13] R. Laskowski, M. Swindells, LigPlot+: Multiple Ligand-Protein Interaction Diagrams for Drug Discovery, *J. Chem. Inf. Model.* 51 (2011) 2778–2786.
- [14] I. Hung, K. Lung, E. Tso, R. Liu, T. Chung, M. Chu, Y. Ng, J. Lo, J. Chan, A. Tam, H. Shum, V. Chan, A. Wu, K. Sin, W. Leung, W. Law, D. Lung, S. Sin, P. Yeung, C. Yip, R. Zhang, A. Fung, E. Yan, K. Leung, J. Ip, A. Chu, W. Chan, A. Ng, R. Lee, K. Fung, A. Yeung, T. Wu, J. Chan, W. Yan, W. Chan, J. Chan, A. Lie, Q. Tsang, V. Cheng, T. Que, C. Lau, K. Chan, K. To, K. 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, *Lancet* 395 (2020) 1695–1704.
- [15] P. Norman, A novel syk kinase inhibitor suitable for inhalation: R-343(?) – wo-2009031011, *Expert. Opin. Ther. Pat.* 19 (2009) 1469–1472.
- [16] R. Kumari, R. Kumar, A. Lynn, g_mmpbsa–A GROMACS Tool for High-Throughput MM-PBSA Calculations, *J. Chem. Inf. Comp. Sci.* 54 (2014) 1951–1962.
- [17] S. Jo, S. Kim, D. H. Shin, M. S. Kim, Inhibition of SARS-CoV 3CL protease by flavonoids, *J. Enzyme. Inhib. Med. Chem.* 35 (2020) 145–151.
- [18] S. Rowland, E. O'Brien Bergin, Screening for low testosterone is needed for early identification and treatment of men at high risk of mortality from covid-19, *Crit. Care* 24 (2020) 367.
- [19] M. L. Agostini, E. L. Andres, A. C. Sims, R. L. Graham, T. P. Sheahan, X. Lu, E. C. Smith, J. B. Case, J. Y. Feng, R. Jordan, A. S. Ray, T. Cihlar, D. Siegel, R. L. Mackman, M. O. Clarke, R. S. Baric, M. R. Denison, Coronavirus susceptibility to the antiviral remdesivir (gs-5734) is mediated by the viral polymerase and the proofreading exonuclease, *mBio* 9 (2018) e00221–18.
- [20] E. P. Tchesnokov, C. J. Gordon, E. Woolner, D. Kocincova, J. K. Perry, J. Y. Feng, D. P. Porter, M. Gotte, Template-dependent inhibition of coronavirus rna-dependent rna polymerase by remdesivir reveals a second mechanism of action, *J. Biol. Chem.* (2020).
- [21] J. Wang, R. Wolf, J. Caldwell, P. Kollman, D. Case, Development and testing of a general amber force field., *J. Comput. Chem.* 25 (2004) 1157–1174.
- [22] M. Frisch, G. Trucks, H. Schlegel, G. Scuseria, M. Robb, J. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. Hratchian, A. Izmaylov, J. Bloino, G. Zheng, J. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. Montgomery, J. Peralta, F. Ogliaro, M. Bearpark, J. Heyd, E. Brothers, K. Kudin, V. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. Burant, S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. Millam, M. Klene, J. Knox, J. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. Stratmann, O. Yazyev, A. Austin, R. Cammi, C. Pomelli, J. Ochterski, R. Martin, K. Morokuma, V. Zakrzewski, G. Voth, P. Salvador, J. Dannenberg, S. Dapprich, A. Daniels, Ö. Farkas, J. Foresman, J. Ortiz, J. Cioslowski, D. Fox, Gaussian 09 Revision E.01, 2009. Gaussian Inc. Wallingford CT.
- [23] C. Bayly, P. Cieplak, W. Cornell, P. Kollman, A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model, *J. Phys. Chem.* 97 (1993) 10269–10280.

- [24] W. Cornell, P. Cieplak, C. Bayly, P. Kollman, Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation, *J. Phys. Chem.* 115 (1993) 9620–9631.
- [25] J. Wang, W. Wang, P. Kollman, D. Case, Automatic atom type and bond type perception in molecular mechanical calculations, *J. Mol. Graphics Modell.* 25 (2006) 247–260.
- [26] D. Case, I. Ben-Shalom, S. Brozell, D. Cerutti, T. Cheatham, V. Cruzeiro, III, T. Darden, R. Duke, D. Ghoreishi, G. Giambasu, T. Giese, M. Gilson, H. Gohlke, A. Goetz, D. Greene, R. Harris, N. Homeyer, Y. Huang, S. Izadi, A. Kovalenko, R. Krasny, T. Kurtzman, T. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, V. Man, D. Mermelstein, K. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, A. Onufriev, F. Pan, R. Qi, D. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C. Simmerling, J. Smith, J. Swails, R. Walker, J. Wang, H. Wei, L. Wilson, R. Wolf, X. Wu, L. Xiao, Y. Xiong, D. York, P. Kollman, Amber 2019, 2019. University of California, San Francisco.
- [27] A. Sousa da Silva, W. Vranken, ACPYPE – AnteChamber PYthon Parser interface, *BMC Res. Notes* 5 (2012) 367.
- [28] M. Abraham, T. Murtola, R. Schulz, S. Páll, J. Smith, B. Hess, E. Lindahl, GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers, *SoftwareX* 1–2 (2015) 19–25.
- [29] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, C. Simmerling, ff14sb: Improving the accuracy of protein side chain and backbone parameters from ff99sb, *J. Chem. Theory Comput.* 11 (2015) 3696–3713.
- [30] 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. Gelpí, C. González, M. Vendruscolo, C. Lughton, S. Harris, D. Case, M. Orozco, Parmbsc1: a refined force field for DNA simulations, *Nat. Methods* 13 (2016) 55.
- [31] W. Jorgensen, J. Chandrasekhar, J. Madura, R. Impey, M. Klein, Comparison of simple potential functions for simulating liquid water, *J. Chem. Theory Comput.* 79 (1983) 926–935.
- [32] M. A. Cauchy, Méthode générale pour la résolution des systèmes d'équations simultanées, *CR Hebd. Acad. Sci.* 25 (1847) 536–538.
- [33] S. Nosé, Molecular-Dynamics Method for Simulations in the Canonical Ensemble, *Mol. Phys.* 52 (1984) 255–268.
- [34] W. Hoover, Canonical Dynamics – Equilibrium Phase-Space Distributions, *Phys. Rev. A* 31 (1985) 1695–1697.
- [35] M. Parrinello, A. Rahman, Polymorphic Transitions in Single-Crystals - a New Molecular-Dynamics Method, *Mol. Phys.* 52 (1981) 7182–7190.
- [36] T. Darden, D. York, L. Pedersen, Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems, *J. Chem. Phys.* 98 (1993) 10089–10092.
- [37] U. Essmann, L. Perera, M. Berkowitz, A Smooth Particle Mesh Ewald Method, *J. Chem. Phys.* 103 (1995) 8577–8593.
- [38] B. Hess, H. Bekker, H. Berendsen, J. Fraaije, LINCS: A linear constraint solver for molecular simulations, *J. Comput. Chem.* 18 (1997) 1463–1472.
- [39] The PyMOL Molecular Graphics System, Version 1.8, 2015. Schrödinger, LLC.
- [40] X. Daura, K. Gademann, B. Jaun, D. Seebach, W. van Gunsteren, A. E. Mark, Peptide Folding: When Simulation Meets Experiment, *Angew. Chem. Int. Ed.* 38 (1999) 236–240.
- [41] T. Williams, C. Kelley, many others, Gnuplot 4.6: an interactive plotting program, <http://gnuplot.sourceforge.net/>, 2013.
- [42] G. Van Rossum, F. L. Drake, Python 3 Reference Manual, CreateSpace, Scotts Valley, CA, 2009.
- [43] L. Zhang, D. Lin, X. Sun, U. Curth, C. Drosten, L. Sauerhering, S. Becker, K. Rox, R. Hilgenfeld, Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α -ketoamide inhibitors, *Science* (2020).
- [44] C. Ma, M. D. Sacco, B. Hurst, J. A. Townsend, Y. Hu, T. Szeto, X. Zhang, B. Tarbet, M. T. Marty, Y. Chen, J. Wang, Boceprevir, gc-376, and calpain inhibitors ii, xii inhibit sars-cov-2 viral replication by targeting the viral main protease, *bioRxiv : the preprint server for biology* (2020) 678–692.
- [45] D. Krause, P. Thörnig, JURECA: Modular supercomputer at Jülich Supercomputing Centre, *JLSRF* 4 (2018) A132.