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 (IBI-7), Forschungszentrum Jülich, 52428 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, M^pro of the COVID-19 virus. With the high-resolution X-ray crystallography structure of this viral enzyme recently being solved, CADD provides a veritable tool for rapidly screening diverse sets of compounds with the aim of identifying ligands capable of forming energetically favorable complexes with M^pro. From our screening of 1,082,653 compounds derived from the ZINC, the DrugBank, and our in-house African natural product libraries, and a rescreening protocol incorporating enzyme dynamics via ensemble docking, we have been able to identify a range of prospective M^pro inhibitors, which include FDA-approved drugs, drug candidates in clinical trials, as well as natural products. The top-ranking compounds are characterized by the presence of an extended ring system combined with functional groups that allow the ligands to adapt flexibly to the M^pro active site as, for example, present in the biflavonoid amentoflavone, one of the most promising compounds identified here. This particular chemical architecture leads to considerable stronger binding than found for reference compounds with *in vitro* demonstrated M^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.

1   Introduction

The recently identified COVID-19 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, COVID-19, another member of the Coronaviridae family previously unknown 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 April 6th 2020, there are almost 1,300,000 confirmed cases globally, with a 5% confirmed-case fatality rate [10]. The virus causing COVID-19 has been named SARS-CoV-2, because its RNA genome is about 82% identical to SARS-CoV [11]. 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 [12]. With its highly effective mode of transmission, COVID-19, in spite of its relatively low fatality rate, 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 (3CLpro) also called main protease (M^pro), papain-like protease, helicase, and RNA-dependent RNA polymerase [13, 14]. Since the main protease M^pro is crucial for viral replication and well conserved across the *Coronaviridae* family, it represents a viable target for drug design [11]. M^pro cleaves the large polyprotein 1ab (replicase 1ab, ∼790 kDa) at 11 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 M^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..

There are around 20 3D protein structures for SARS-CoV-2, which were resolved by X-ray diffraction or cryo-electron microscopy and are deposited in the Protein Data Bank (PDB). Most of these structures are for M^{pro} in apo- and holo-states; one of them is a M^{pro} crystal structure in complex with a peptidomimetic inhibitor called N3 (PDB code 6LU7, Figure 1) [15].

While the catalytically active form of M^{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 [16]. The proteolytic process in the active site of M^{pro} is enabled by the catalytic C145-H41 dyad with the cysteine thiol group acting as the nucleophile (Figure 1C) [17]. The initial drug discovery efforts after the SARS outbreak in 2003 aimed at electrophilic attack to the Cys residue of the catalytic dyad via covalent Michael inhibitors [18]. 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 [19]. 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 drug design and development to treat this disease have already been published [15, 20–26]. In one of these studies, the mechanism-based inhibitor N3 was designed (Figure 1) using computer-aided drug design (CADD) techniques [15]. Six 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 M^{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 [27].
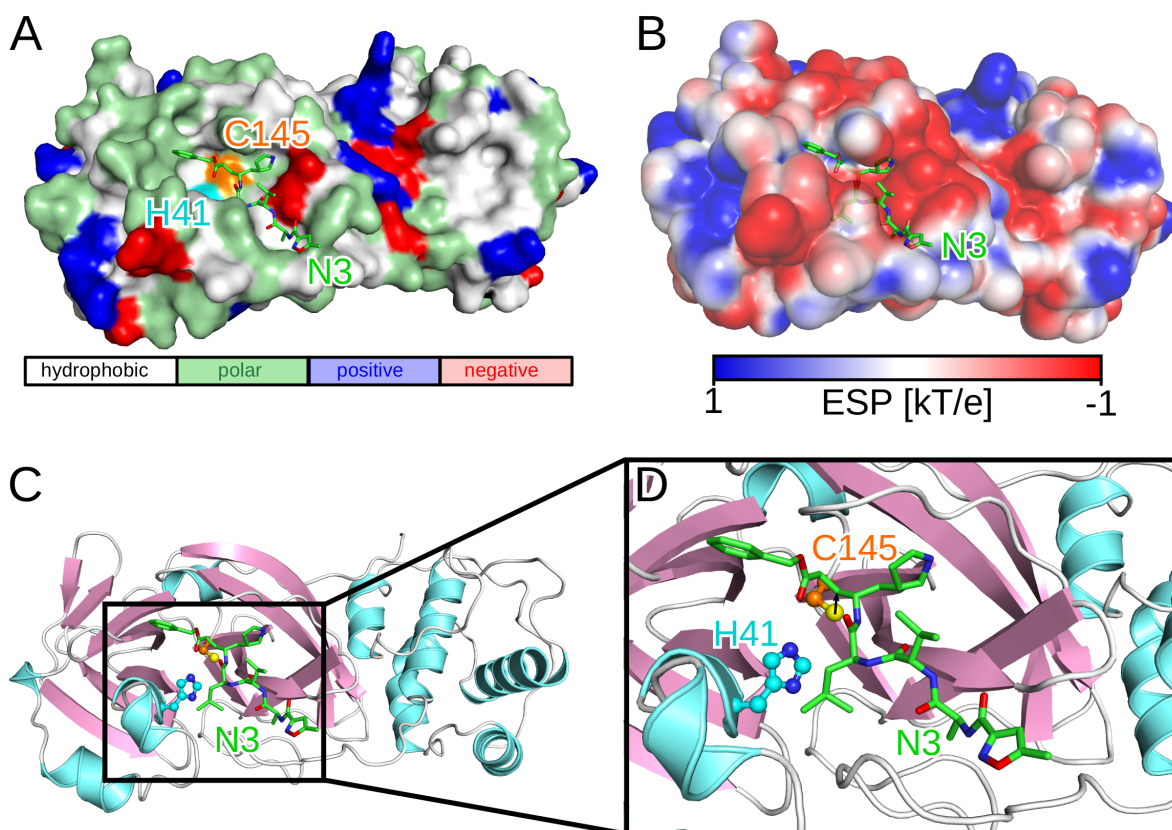
Figure 1: Structure and binding site of M^pro including the ligand N3 (PDB code 6LLU7). (A) The surface of M^pro is shown and colored according to residue type using the color code given at the bottom. The catalytic dyad residues H41 and C145 are colored in cyan and orange, respectively. The stick representation 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 M^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). (C) Cartoon representation of M^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 active site of M^pro. In panels (B)–(D) the same representation for N3 is used as in panel (A).

Here, we use high-throughput virtual screening to discover potential inhibitors of M^pro of SARS-CoV-2. We screened more than 1 million compounds considering three different databases: the ZINC database, which is a curated collection of more than 230 million commercially available chemical compounds prepared for virtual screening [28–30], DrugBank containing 13,540 drug entries including 2,630 approved small molecule drugs, 1,372 approved biologics (proteins, peptides, vaccines, and allergenics), 131 nutraceuticals and over 6,358 experimental (discovery-phase) drugs [31,32], as well as our in-house database with more than 3,200 natural compounds isolated from African plants. More than 11,000 of the best predictons are then further assessed using ensemble

docking [33]. The 100 top-performing compounds from this redocking round are analyzed in more detail, and the properties and binding to M^{pro} for nine of these compounds are discussed here and compared to the results obtained in another drug design initiative targeting M^{pro} [15].

2   Results and Discussion

2.1   High throughput virtual screening

In this work we have employed CADD approaches in search of chemical compounds that are capable of forming thermodynamically feasible binary complexes with the SARS-CoV-2 main protease enzyme M^{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 [15, 20]. First, we have not limited our virtual screening to previously known inhibitors of viral proteases, and second, our screening procedure incorporated protein dynamics. The immediate objective of our research work is the identification of small molecule inhibitors of M^{pro} and 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 1,082,653 compounds derived from the ZINC library [28–30], the Drugbank database [31, 32], as well as our in-house library of natural products of African origin (unpublished), against the crystallographic structure of the functional form of the SARS-CoV-2 M^{pro} enzyme (PDB code 6LU7 [15]). We also included in our screening nine recently reported inhibitors of the SARS-CoV2 M^{pro}: eight of these — the peptidomimetic N3, ebselen, disulfiram, carmofur, cinanserin, shikonin, tideglusib, and PX-12 — were reported alongside the crystal structure of M^{pro} complexed with N3 [15], while ivermectin is currently attracting some research attention as a prospective M^{pro} inhibitor [34]. All nine reference compounds were able to bind to the active site of the crystal structure of the enzyme. A binding free energy value (ΔG) of -6.7 kcal/mol was obtained for N3, while PX-12 was indicated in our calculations as possessing the least binding affinity with a ΔG value of -4.2 kcal/mol. Interestingly, the experimental IC$_{50}$ value obtained for PX-12 also suggested it to be the least binding out of the eight experimentally validated M^{pro} inhibitors [15]. Ivermectin similarly showed a moderate binding affinity for the enzyme.

   From the screened library, we obtained a total of 11,155 compounds whose computed affinities in terms of ΔG were less than or equal to -8.0 kcal/mol. The choice of the cutoff was to improve the chances of obtaining compounds with binding strengths significantly superior to those reported for the reference inhibitors. Moreover, as affinities computed from computer simulations may be affected by the neglect of a number of factors such as receptor flexibility and explicit solvation, we selected a cutoff considerably below the -6.7 kcal/mol obtained for N3, the primary reference inhibitor. We analyzed to which extent compounds from the three different databases contribute

to the selected 11,155 compounds (Figure 2A). While compounds from all three libraries are represented in the top-performing compounds, compounds from the ZINC library dominate; this is of course not surprising since compounds from the ZINC library represented the majority of the screened compounds. We further analyzed the physicochemical properties of the selected compounds. Figure 2B shows that compounds captured by the -8.0 kcal/mol cutoff were observed to belong to several distinct chemical groups with molecular weights varying from just around 200 to almost 1000 g/mol. Small and large molecular weight compounds are present in the top predictions from all three libraries. Since the intention at this time is to identify any compounds with the potential to bind to the substrate site of M$^{\text{pro}}$ regardless of its pharmacokinetic attributes, we have decided to deprioritize filtering based on physicochemical properties such as the molecular weight employed in rule of thumb principles to predict oral bioavailability. There appears no apparent relationship between the molecular weight of the compounds and the predicted affinities, possibly highlighting the role of specific binding site interactions rather than size in determining the strength of binding. We analyzed the dominant ligand chemical fragments to find out if some chemical groups were more favored in the 11,115 strong-binding compounds. Figure S1 shows the most preponderant chemical fragments, most of which are associated with known drug molecules.
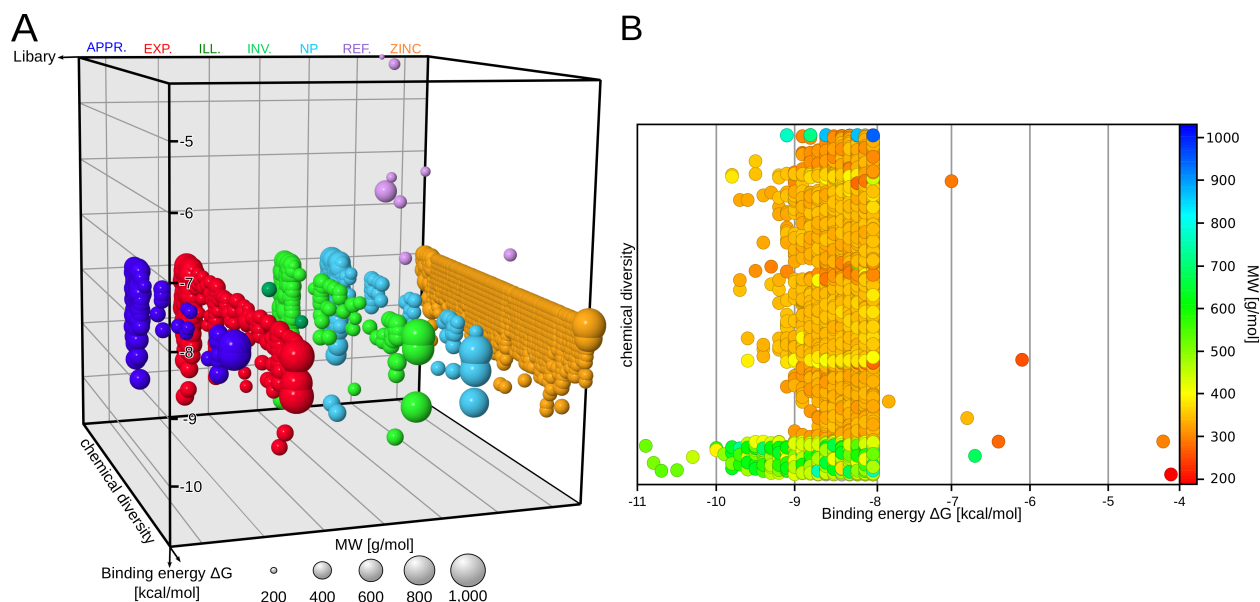
Figure 2: Distribution of the best 11,155 compounds selected from docking screen against M^pro. (A) A 3D plot of the compounds showing the distribution across ligand libraries, in terms of chemical diversity, molecular weight (MW), and ΔG. The DrugBank library is composed of FDA-approved drugs (APPR.), experimental drugs (EXP.), illicit drugs (ILL.), and investigational drugs (INV.). Our in-house ligand library of natural products from African plants is shown as NP, while the reference inhibitors and compounds from the ZINC database are presented as REF. and ZINC, respectively. (B) Distribution of the compounds in terms of chemical diversity and binding free energies (ΔG cutoff = -8.0 kcal/mol). The distribution of the molecular weight is given by the color bar on the left. At $\Delta G > -8.0$ kcal/mol the reference inhibitors can also be seen as representing a chemically diverse set of ligands. Plots were generated with Data Warrior chemoinformatic software [35].

2.2   Redocking to account for protein flexibility of M^pro

To account for the influence of protein flexibility in ligand binding we performed a 100 ns MD simulation of the M^pro-N3 complex in solution and clustered the active site conformations collected during that simulation. The five most dominant conformations were then employed to perform a rescreening of the 11,155 best compounds obtained from the initial screening against the protein crystal structure. The ΔG values obtained for each compound were then averaged over the five protease conformations and the top 100 compounds resulting from this step were selected for analysis. All 100 compounds along with information on their structure and from which of the three databases they originate are provided in Table S1. Analysis of the structural fragments in this smaller subset of protease binders (Figure S2) suggests a relative selection for multicyclic ring structures over monocyclic rings that dominate in Figure S1. While in the top 11,155 binders resulting from high throughput screening in the absence of considerations of M^pro dynamics monocyclic ring

fragments have a 58.3% representation in the most frequently encountered fragments, in the top 100 compounds after incorporating M$^{\text{pro}}$ dynamics their share falls to 16.7%. This suggests a structural selection by the SARS-CoV-2 binding site for larger fused ring systems.

To further analyze the binding of the ligands to the substrate site of M$^{\text{pro}}$, we evaluated the closeness of the ligands to the catalytic dyad C145 and H41. We found that a substantial number of the top 100 compounds does not directly bind to the dyad despite having very low ΔG values. This especially applies to majorly hydrophobic residues like testosterone, which is listed as the 6$^{\text{th}}$ best binding compound in Table S1 and is thus referred to as compound **5** in the following. Figure 3 shows that testosterone, as other ligands which do not bind to C145 and H41, associate with M$^{\text{pro}}$ at the entry to the active site. Comparison with the electrostatic potential (ESP) of M$^{\text{pro}}$ in Figure 1B reveals that this binding is driven by mainly hydrophobic interactions as at the entry to the active site the ESP is close to zero, whereas it is partly negative in the active site. The nine compounds that bind best to the catalytic dyad and also have overall very low ΔG values are shown in their bound M$^{\text{pro}}$ state in Figure 4. According to their position in Table S1, these compounds are referred to as **1**, **2**, **4**, **6**, **14**, **19**, **23**, **27**, and **28** in the following. Comparison to the reference structures [15] that we also docked to M$^{\text{pro}}$ (Figure S3) shows that our CADD approach involving a larger collection of compounds and included protteiptpros flexibility at the second docking stage favors larger compounds with extended ring structures. This explains the smaller ΔG values for the compounds identified in this work. Apart from this, the current best and the reference compounds occupy a similar region in the active site.
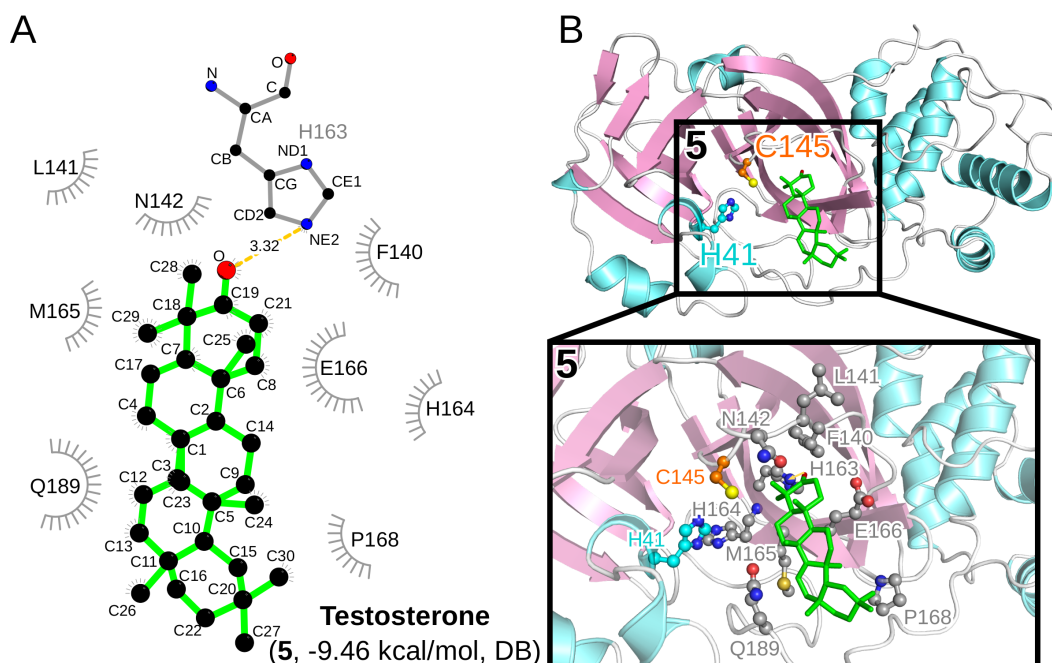
Figure 3: The M^pro–testosterone interactions and the pose of testerone in its binding site. (A) The interactions were analyzed and plotted with LigPlot+ [36, 37]. Hydrogen bonds are indicated by orange dashed lines between the atoms involved and the donor–acceptor distance is given in Å, while hydrophobic contacts are represented by gray arcs with spokes radiating towards the ligand atoms they contact. The contacted atoms are shown with spokes radiating back. (B) The binding pose of testosterone is shown using the same protein and ligand representation as well as color scheme as used in Figure 1.
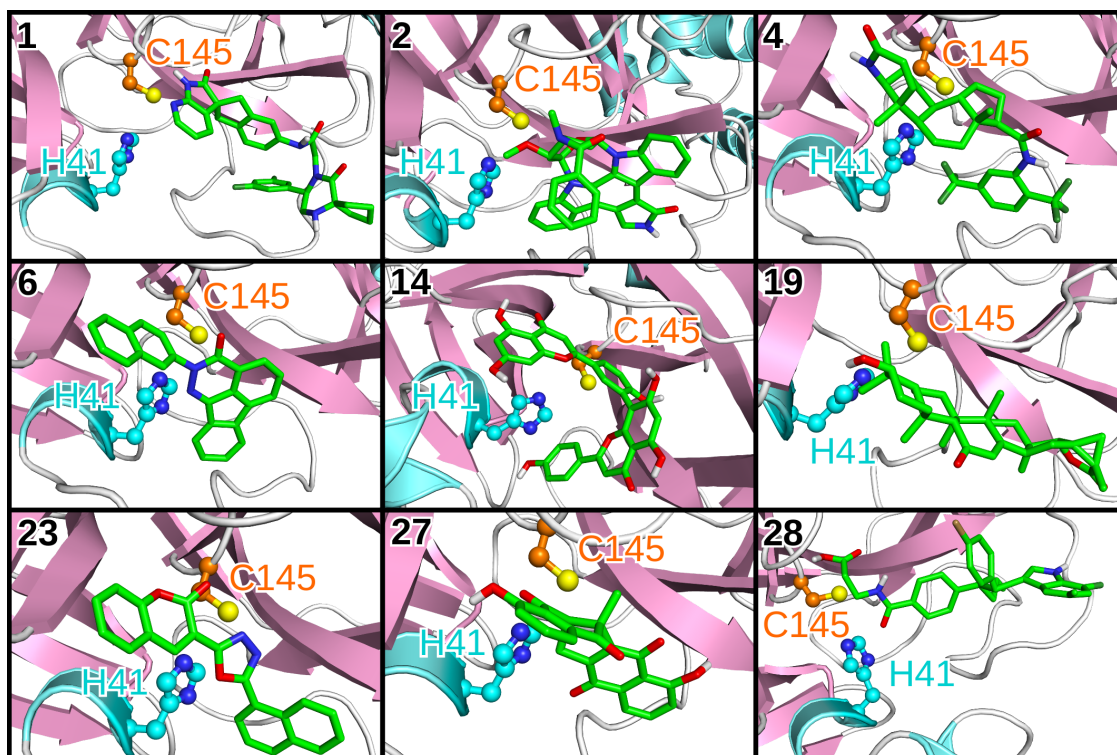
Figure 4: The poses of the nine compounds 1, 2, 4, 6, 14, 19, 23, 27, and 28. These compounds bind well to the active site in terms of both ΔG and closeness to the catalytic dyad of M^pro as indicated by these plots. The same protein and ligand representation as well as color scheme as in Figure 1 is used.

2.3   Binding of the top nine predicted compounds

In this section the binding of the best predictions **1**, **2**, **4**, **6**, **14**, **19**, **23**, **27**, and **28** is discussed in detail and further information for each of these compounds is provided. In order to understand the M^pro — compound interactions we plotted the protein — ligand interactions using LigPlot+ [36, 37] (Figure 5). While all front runner compounds exhibited strong binding interactions with the substrate binding site of M^pro, we have decided to accord special attention to those compounds that additionally interact with at least one of the two catalytic dyad residues. The binding site because of its peculiar topology enables the binding of various chemical groups at different subsites within the rather large accommodation substrate site. Compounds with the steroidal cyclopentanophenanthrene skeleton such as testosterone bind strongly to a hydrophobic cavity at the entrance of the active site, while the catalytic dyad is buried further within the interior. All the nine compounds that we have selected for further discussion possess the chemical architecture which, though widely diverse, allows for direct interaction with the dyad.
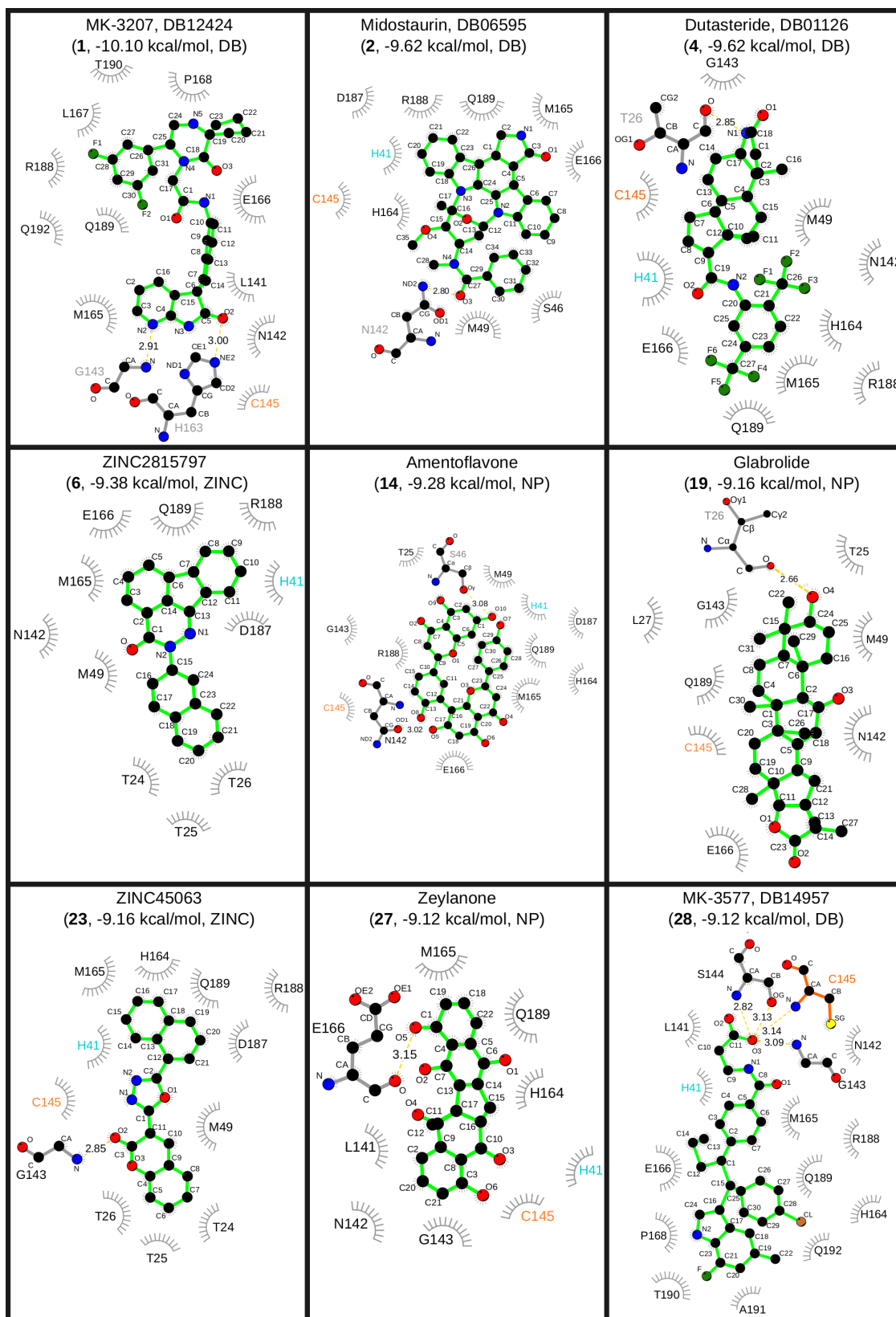
Figure 5: The M^pro–compound interactions for the nine compounds 1, 2, 4, 6, 14, 19, 23, 27, and 28. The interactions were analyzed and plotted with LigPlot+ [36, 37]. Hydrogen bonds are indicated by orange dashed lines between the atoms involved and the donor–acceptor distance is given in Å, while hydrophobic contacts are represented by gray arcs with spokes radiating towards the ligand atoms they contact. The contacted atoms are shown with spokes radiating back.

Compound 1.   The investigational drug MK-3207 (molecular weight (MW) 557.6 g/mol) is a piperazinone derivative that has been used in trials studying the treatment of migraine and migraine disorders [38]. MK-3207 represents the culmination of structural improvements on earlier candidate compounds studied as inhibitors of calcitonin gene-related peptide receptor (CGRPR) antagonists whose roles have been confirmed in migraine [39]. MK-3207 is employed for the inhibition of the binding of CGRPR to its substrate, the calcitonin gene-related peptide, a thirty-seven amino acid neuropeptide. Thus, MK-3207 represents a chemical framework that has been optimized for interacting with a peptide-binding substrate site similar to the M^{pro} active site. In binding to M^{pro}, the piperazinone fragment of MK-3207 is buried within a portion of the binding site where it is able to form hydrophobic contacts with L167, P168, Q189, T190 and Q192 while the amide linker allows further contacts with active site residues like E166 and L141 (Figure 5A). The pyrrolopyridinone fragment of MK-3207 forms hydrophobic contacts with the catalytic residue C145 as well as other residues (N142 and M165) in the vicinity. At the same time, this fragment, through its lactam oxygen and its pyridine nitrogen atoms, establishes two hydrogen bonding contacts with H163 sidechain and G143 backbone atoms, respectively. In general, MK-3207 possesses an interesting combination of structural features that support its strong binding to M^{pro}, and in particular the presence of three amide groups, two of which are cyclized.

Compound 2.   This compound is an approved drug, PKC412 (midostaurin, MW 570.6 g/mol) for the treatment of acute myeloid leukemia in adults with an FMS-like tyrosine kinase-3 (FLT3) mutation, as well as for the treatment of advanced systemic mastocytosis [40]. The drug development journey of midostaurin was rather interesting, especially on account of its ability to interact with multiple targets including protein kinase C implicated in different solid tumors and B-cell malignancies, the angiogenic vascular endothelial growth factor involved in diabetic retinopathy, and FLT3 and KIT representing drug targets in FLT+ acute myeloid leukaemia and systemic mastocytosis for which it was eventually approved by the US Food and Drug Administration (FDA) and the European Medicines Agency [41]. Similar to MK-3207, midostaurin is able to establish extensive contacts that are mostly hydrophobic in nature with the active site residues of M^{pro}. Some of these contacts include S46, M49, H164, M165, E166, D187, R188, and Q189. Unlike MK 3207, midostaurin possesses a significantly rigid core composed of an eight-fused ring system and a relatively flexible acetamide-linked phenyl group. The linker amide is able to form a hydrogen bond with N162, which complements the hydrophobic interactions in other parts of that group. The pose of midostaurin in the active site allows for hydrophobic contacts with C145 and H41. It is also important to point out that midostaurin also features two amide bonds, one of which is a lactam, considering that proteins are the substrate of M^{pro}.

Compound 3.   This compound belongs to the steroid class. It is an approved drug called dutasteride (MW 528.5 g/mol) and is used because of its effectiveness as a 5α-reductase inhibitor, which converts testosterone to dihydrotestosterone. It is therefore employed to treat the symptoms of an enlarged prostate, for scalp hair loss in men and as a part of hormone therapy in transgender women. It is on the market since 2001. A recent study indicated that this steroid-based drug may be of use as a downregulator of transcriptional processes associated with proteins crucial to SARS-CoV-2 entry into host cells [42]. This study, however, did not examine the possible role of dutasteride as an inhibitor of the viral main protease enzyme. The structural architecture of dutasteride, as that of midostaurine, is characzerized by a rigid structural skeleton, which is composed of the steroidal backbone, and by possessing an acetamide linker separating a strongly fluorinated phenyl ring from the larger steroid structure. Moreover, also dutasteride possesses one proper amide group and a lactam ring in its A ring. The smaller overall structure of dutasteride, however, permits a bound conformational arrangement that places the fluorinated ring, rather than the larger core structure, in a location where it forms hydrophobic contacts with H41. The hydrophobic steroidal ring makes contact with Cys145 while a single hydrogen bond connects the lactam nitrogen atom of the drug with T26.

Compound 6.   This ZINC database compound (ZINC2815797) with IUPAC name (5R)-3-(naphthalen-2-yl)-2,3-diazatetracyclo[7.6.1.0^{5,16}.0^{10,15}]hexadeca-1,6,9(16),10(15),11,13-hexaen-4-one (MW 348.4 g/mol) features an internal amide at the center of its structure and fits snugly into the M^{pro} binding site with its hydrophobic naphthalene system oriented perpendicularly to the plane of the four-fused rings. In this pose, the naphthalene makes hydrophobic contacts with three threonine residues (T24, T25, and T26), while the bulkier fragment of the compound is inserted into a hydrophobic core where it directly interacts with H41, M165 and other hydrophobic residues. This candidate compound relies exclusively on non-polar contacts without any indication of hydrogen bonds.

Compound 14.   Amentoflavone (MW 538.5 g/mol), a natural product and a biflavonoid present in multiple plant sources such as *Ginkgo biloba* and St. John's Wort, was one of the candidate compounds identified in our systematic virtual screening approach. A strong binding ΔG value of -9.28 kcal/mol was obtained for this polyphenolic compound, a value which is at least 2 kcal/mol better than that obtained for the co-crystallized inhibitor N3. 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 S46. Multiple hydrophobic contacts involving 11 residues, including both catalytic dyad residues, account for the strong binding to the M^{pro} active 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 [43]. 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 correlation with existing data for the main protease target of a virus that bears high similarity with SARS-CoV-2 provides important validation for the computational results presented in this work. It is further important to note that amentoflavone has similarly been reported with antiviral activities mediated by interaction with other viral targets in viral infections including Dengue virus [44], Coxsackievirus B3 [45], the human immunodeficiency virus [46], and respiratory syncytial virus [47].

Compound 19.   Glabrolide (MW 468.7 g/mol) is a natural product found in herbs and spices, such as *Glycyrrhiza glabra* (licorice) and *Glycyrrhiza uralensis* (Chinese licorice). Glabrolide, while possessing structural similarity to the steroid structure, 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 M^pro it establishes an hydrogen bond with T26 and several hydrophobic contacts, including a critical contact with C145.

Compound 23.   This molecule is another compound from the ZINC library (ZINC45063, IUPAC name 3-[5-(naphthalen-1-yl)-1,3,4-oxadiazol-2-yl]chromen-2-one, MW 340.3 g/mol) that is able to form essential contacts with the active site of M^pro. While it lacks the amide function that appears common to some of the top-ranking compounds, it is able to form hydrogen bond via its lactone group with the backbone nitrogen atom of G143 while maintaining hydrophobic contacts with both residues of the catalytic dyad as well as most other contacts identified for other candidate compounds.

Compound 27.   Zeylanone (MW 374.3 g/mol) is a quinone from plants like *Plumbago zeylanica*. With its characteristic centrally bent five-ring system it was able to fit comfortably within the hydrophobic cavity and is stabilized both by hydrogen bonding with E166 and extensive networks of hydrophobic contacts that include both catalytic dyad amino acids. While zeylanone has not yet been experimentally examined in SARS-CoV-2 M^pro enzyme-based replication assays, the extract of the plant has been shown to possess inhibitory activities against Coxsackie virus B3, influenza A virus and herpes simplex virus type 1 Kupka [48], while zeylanone epoxide inhibited influenza virus replication *in vitro* [49].

Compound 28.   MK-3577 (MW 521.0 g/mol) is an investigational drug owned by Merck that is being examined for the treatment of type 2 diabetes mellitus. Out of all examined candidate

compounds in this study, it is the only one that has been able to establish networks of hydrogen bonding that involve S144 and C145, with the later contact similar to the pattern obtained in the crystal structure of the M^{pro}-N3 complex [15] (Figure S5). It also makes contact with H41 and several other binding site amino acid residues.

3   Conclusions

Using a computational drug design approach, we have identified a number of compounds whose ability to establish contacts with critical active site residues of M^{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. Since most of these contacts involve hydrophobic attraction the likelihood of covalent interaction with C145, and as a result the toxicity associated with such covalent Cys modifications, is remote. 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 M^{pro} active site. This closely mimics the experience with the reference inhibitors which are found to belong to diverse chemical subsets ranging from aliphatic systems to compounds with aromatic centers [15]. However, as a result of the presence of multiple and multicyclic aromatic rings combined with linker groups that allow the compounds to adapt flexibly to the SARS-CoV-2 M^{pro} active site, the compounds identified from our investigation were able to outperform the reference inhibitors with respect to M^{pro} interaction. This includes ivermectin, an FDA-approved antiparasitic with demonstrated broad-spectrum antiviral activity *in vitro*, which was very recently shown to effect a 5000-fold reduction in viral RNA at 48 h after infection of cells with SARS-CoV-2 [34]. Figure S7 shows that ivermectin can bind to the active site of M^{pro} with similar binding free energies as the compounds identified in ref. [15] and which were demonstrated *in vitro* to possess M^{pro} inhibition capabilities. Thus, downregulation of SARS-CoV-2 by ivermectin may indeed result from M^{pro} inhibition; however, *in vitro* and cell assays have yet to establish (or also disprove) this hypothesis. Apart from the top nine compounds identified and presented in this work, we also identified testosterone as one of the compounds able to interact strongly within an hydrophobic cleft adjacent to but not contacting the catalytic dyad. Further study will be required to investigate the role of testosterone in COVID-19, especially since there are reports linking vulnerability to the viral infection with gender and age, two variables that bear significant consequence on testosterone titer. In addition, the unbiased virtual screening protocol employed in this study was able to identify amentoflavone, a biflavonoid that was reported a decade ago with micromolar range inhibitory effect on the closely related main protease enzyme of SARS-CoV [43]. This rare finding provides a validation for the virtual screening method

incorporating protein dynamics employed in this work. 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 1,082,653 ligand structural models from different ligand libraries, including the ZINC databasei [28–30], the DrugBank library [31,32] (DB), and an in-house generated library of natural products of Nigerian origin (NP). The DB compounds included compounds described as of FDA-approved drugs (APPR.), experimental drugs (EXP.), investigational drugs (INV.), illicit (ILL.) drugs, as well as withdrawn drugs. 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. For the same reason, we had not selected a screening protocol focused on known protease inhibitors. While random visual inspection of the ligand models obtained from the different sources indicated *prima facie* structural correctness, models from the DrugBank library required a preparatory phase of steepest descent energy minimization employing the *obminimize* functionality of the OpenBabel programme [50]. We retrieved the X-ray crystallographic structure of the SARS-CoV-2 (M$^{\text{pro}}$) enzyme (PDB code 6LU7) [15] from the RCSB website [51]. The structure represents the functional form of M$^{\text{pro}}$ in complex with a peptidomimetic inhibitor (called 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 M$^{\text{pro}}$ active site was generated with AutoDock Tool [52, 53] and centered at *x, y, z* position of -10.745, 12.328, 68.844 Å. After adding hydrogen atoms to M$^{\text{pro}}$, Gasteiger partial charges were computed and added separately to M$^{\text{pro}}$ and N3 using AutoDock Tool [52,53], and the resulting models were saved in the PDBQT format. PDBQT files were also generated for each of the ligands in the screening library, following which the library together with the co-crystallized inhibitor compound was subjected to a virtual screen against the M$^{\text{pro}}$ molecule using AutoDock Vina [54], which treats the ligand as fully flexible while keeping the receptor rigid. After the initial screening against the M$^{\text{pro}}$ crystal structure, 11,155 *virtual hits* were selected whose computed binding free energy ΔG values were less than or equal to -8.0 kcal/mol. These were subsequently employed in the second round of screening incorporating protein dynamics.

Virtual rescreening protocol – Incorporating protein dynamics.   To account for protein flexibility in the second docking round we performed ensemble docking [33]. To this end, we performed a 100 ns molecular dynamics (MD) simulation of the M^{pro}-N3 complex in solution using the MD software Gromacs [55]. As protein force field GROMOS54A7 [56] was used and combined with the SPC water model [57, 58] to explicitly simulate water. As starting structure the M^{pro}-N3 crystal structure was taken, which was centered in a dodecahedral box of size $117.5 \times 117.6 \times 83.1$ Å^3, solvated, and NaCl added at a concentration of 150 mM while at the same time neutralizing the system. The system was first energy-minimized, equilibrated in the NVT ensemble (i.e., with a constant number of molecules, volume, and temperature) for 0.1 ns and then simulated for 100 ns in the NpT ensemble at 300 K (Nosé-Hoover thermostat [59, 60]) and 1.0 bar (Parrinello-Rahman barostat [61]). Electrostatic interactions were treated with the particle-mesh Ewald method [62, 63] in conjunction with periodic boundary conditions and a real-space cutoff of 11 Å. The Lennard-Jones interactions were also cut at 11 Å. A leapfrog stochastic dynamics integrator was used for the integration of equations of motion using a time step of 2 fs. The LINCS algorithm [64] was used to constrain all bond lengths during the MD simulation. The resulting trajectory was subjected to geometric clustering [65] based on the conformations sampled by M^{pro} substrate binding site residues (within 10 Å of N3). With a clustering cutoff of 3.5 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 virtual rescreening. As described above, a docking grid was generated for each of the five MD-generated M^{pro} conformations with a 3.0 Åbuffer around the bound position of N3. The 11,155 virtual hits from the screening against the M^{pro} crystal structure were then subjected to another round of screening against the five MD-generated M^{pro} conformers using AutoDock Vina.

Acknowledgments

The authors gratefully acknowledge the computing time granted through JARA-HPC (project COVID19MD) on the supercomputer JURECA at Forschungszentrum Jülich [66], 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] Susan R. Weiss and Julian 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, Susanna 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] L. Zhang, D. Lin, X. Sun, U. Curth, C. Drosten, L. Sauerhering, S. Becker, K. Rox, and R. Hilgenfeld.

[12] 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*, n/a(n/a).

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

[14] 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.

[15] 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^pro from covid-19 virus and discovery of its inhibitors. *bioRxiv*, 2020.

[16] 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. PMID: 26878082.

[17] 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.

[18] 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 Mpro with Enhanced Activity: Application as a Novel Tag-cleavage Endopeptidase for Protein Overproduction. *Journal of Molecular Biology*, 366(3):965–975, 2007.

[19] 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.

[20] 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.

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

[22] 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.

[23] 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.

[24] 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.

[25] 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.

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

[27] 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.

[28] 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. PMID: 15667143.

[29] 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. PMID: 22587354.

[30] T. Sterling and J.J. Irwin. Zinc 15 – ligand discovery for everyone. *Journal of Chemical Information and Modeling*, 55(11):2324–.2337, 2015. PMID: 26479676.

[31] 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.

[32] 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.

[33] Rommie E. Amaro, Jerome Baudry, John Chodera, Özlem Demir, J. Andrew McCammon, Yinglong Miao, and Jeremy C. Smith. Ensemble docking in drug discovery. *Biophysical Journal*, 114(10):2271–2278, 2018.

[34] Leon Caly, Julian D. Druce, Mike G. Catton, David A. Jans, and Kylie M. Wagstaff. The FDA-approved Drug Ivermectin inhibits the replication of SARS-CoV-2 in vitro. *Antiviral Research*, page 104787, 2020.

[35] 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.

[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] D.J. Hewitt, S.K. Aurora, D.W. Dodick, P.J. Goadsby, Y. (Joy) Ge, R. Bachman, D. Taraborelli, X. Fan, C. Assaid, C. Lines, and T.W. Ho. Randomized controlled trial of the cgrp receptor antagonist mk-3207 in the acute treatment of migraine. *Cephalalgia*, 31(6):712–722, 2011.

[39] I.M. Bell, S.N. Gallicchio, M.R. Wood, A.G. Quigley, C.A. Stump, C.B. Zartman, J.F. Fay, C.-C. Li, J.J. Lynch, E.L. Moore, S.D. Mosser, T. Prueksaritanont, C.P. Regan, S. Roller, C.A. Salvatore, S.A. Kane, J.P. Vacca, and H.G. Selnick. Discovery of mk-3207: A highly potent, orally bioavailable cgrp receptor antagonist. *ACS medicinal chemistry letters*, 1(1):24–29, 2010.

[40] M.R. Luskin and D.J. DeAngelo. Midostaurin/pkc412 for the treatment of newly diagnosed flt3 mutation-positive acute myeloid leukemia. *Expert Review of Hematology*, 10(12):1033–1045, 2017.

[41] R.M. Stone, P.W. Manley, R.A. Larson, and R. Capdeville. Midostaurin: its odyssey from discovery to approval for treating acute myeloid leukemia and advanced systemic mastocytosis. *Blood advances*, 2(4):444–453, 2018.

[42] X. Wang, R. Dhindsa, G. Povysil, A. Zoghbi, J. Motelow, J. Hostyk, and D. Goldstein. Transcriptional Inhibition of Host Viral Entry Proteins as a Therapeutic Strategy for SARS-CoV-2. 2020.

[43] 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.

[44] 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.

[45] 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.

[46] 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.

[47] 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.

[48] 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.

[49] 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.

[50] N.M. O'Boyle, M. Banck, C.A. James, C. Morley, T. Vandermeersch, and G.R. Hutchison. Open babel: An open chemical toolbox. *Journal of cheminformatics*, 3(1):33, 2011.

[51] 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.

[52] 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.

[53] 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.

[54] 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.

[55] 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.

[56] N. Schmid, A.P. Eichenberger, A. Choutko, S. Riniker, M. Winger, A.E. Mark, and W.F. van Gunsteren. Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *European Biophysics Journal*, 40(7):843, Apr 2011.

[57] H.J.C Berendsen, J.P.M Postma, W.F. Van Gunsteren, and a J. Hermans. Intermolecular forces, 1981.

[58] G.W. Robinson. *Water in biology, chemistry, and physics: experimental overviews and computational methodologies*, volume 9. World Scientific, 1996.

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

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

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

[62] D. York T. Darden and L. Pedersen. Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.*, 98(12):10089–10092, 1993.

[63] U. Essmann, L. Perera, and M. L. Berkowitz. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.*, 103(19):8577–8593, 1995.

[64] B. Hess, H. Bekker, H. J. C. Berendsen, and J. Fraaije. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.*, 18(12):1463–1472, 1997.

[65] Xavier Daura, Karl Gademann, Bernhard Jaun, Dieter Seebach, Wilfred F. van Gunsteren, and Alan E. Mark. Peptide Folding: When Simulation Meets Experiment. *Angew. Chem. Int. Ed.*, 38(1-2):236–240, 1999.

[66] Dorian Krause and Philipp Thörnig. Jureca: Modular supercomputer at Jülich supercomputing centre. *Journal of large-scale research facilities*, 4:A132, 2018.

Supporting Information

Figure S1: Chemical fragments majorly featured in the top performing 11,155 compounds obtained from high throughput virtual screening against the crystal structure of the SARS-CoV-2 main protease M^pro. The numbers represent the occurrence in absolute numbers.
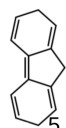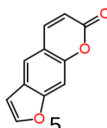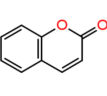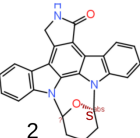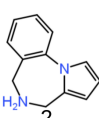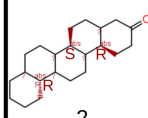
Figure S2: Chemical fragments majorly featured in the top performing 100 compounds obtained from rescreening against different M^pro structures generated by MD simulation. The numbers represent the occurrence in absolute numbers.
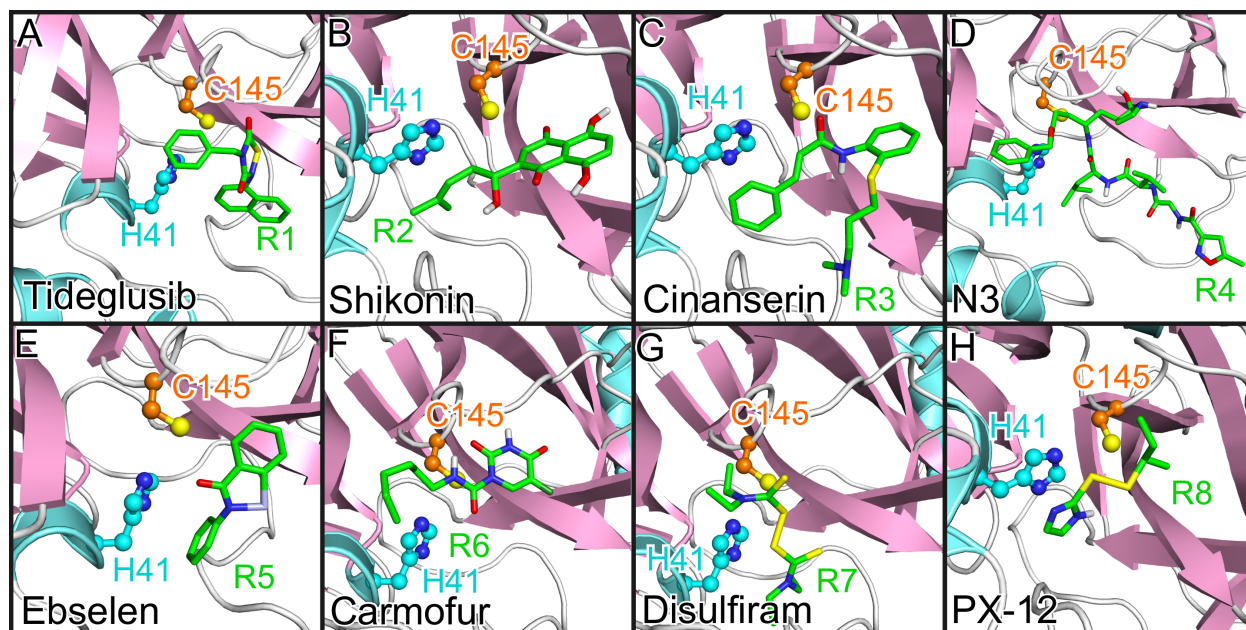
Figure S3: The poses of the reference structures taken from ref. [15] binding to M^{pro}. The same protein and ligand representation as well as color scheme as in Figure 1 is used.
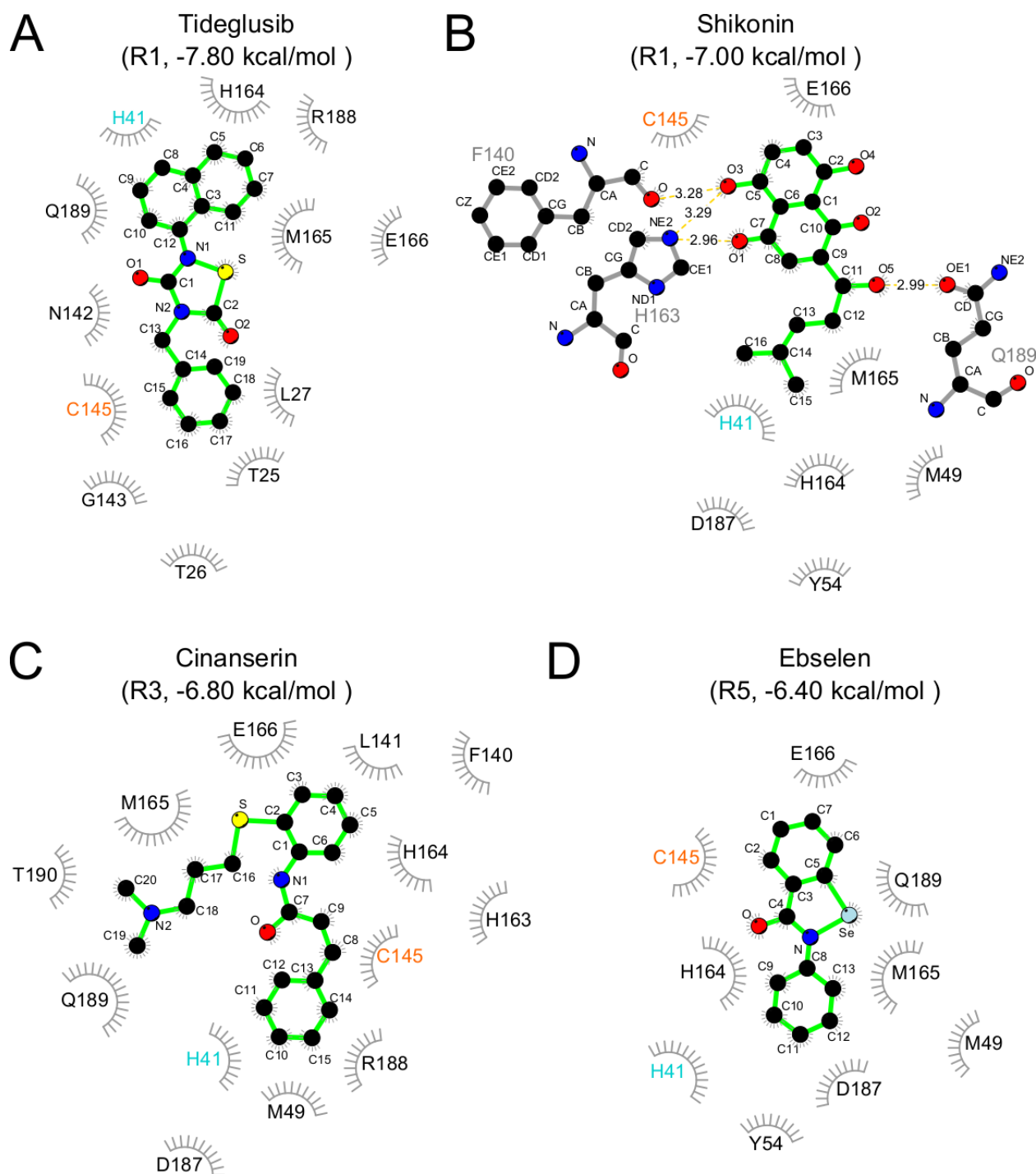
Figure S4: The M^pro–compound interactions for the reference compounds 1–4 taken from ref. [15]. The interactions were analyzed and plotted with LigPlot+ [36, 37]. See Figure 3 for an explanation of the representation.
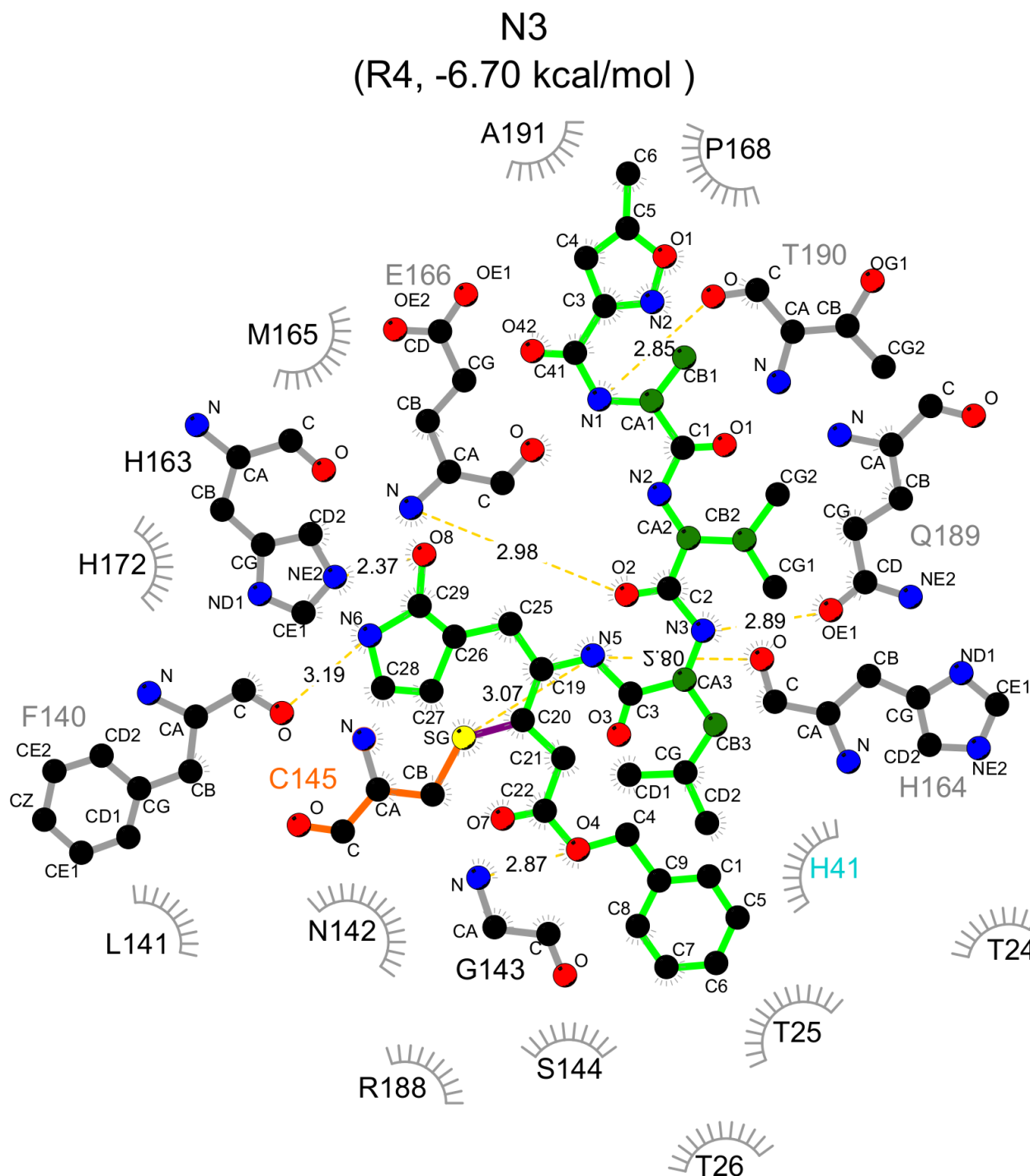
Figure S5: The M^pro–compound interactions for the reference compound 5, which corresponds to N3, taken from ref. [15]. The interactions were analyzed and plotted with LigPlot+ [36, 37]. See Figure 3 for an explanation of the representation. In addition, the violet line between the S atom of C145 and the ligand N3 indicates the protein–ligand bond via Micheal addition.

High throughput virtual screening for M^{pro} inhibitors   A PREPRINT

Figure S6: The M^{pro}–compound interactions for the reference compounds 6–8 taken from ref. [15]. The interactions were analyzed and plotted with LigPlot+ [36, 37]. See Figure 3 for an explanation of the representation.
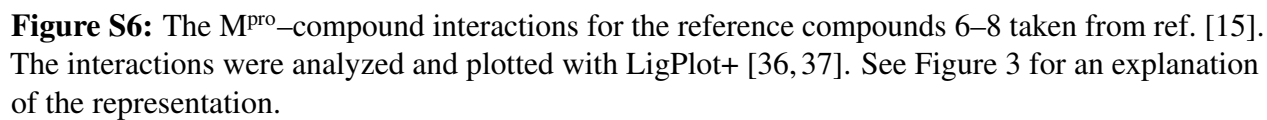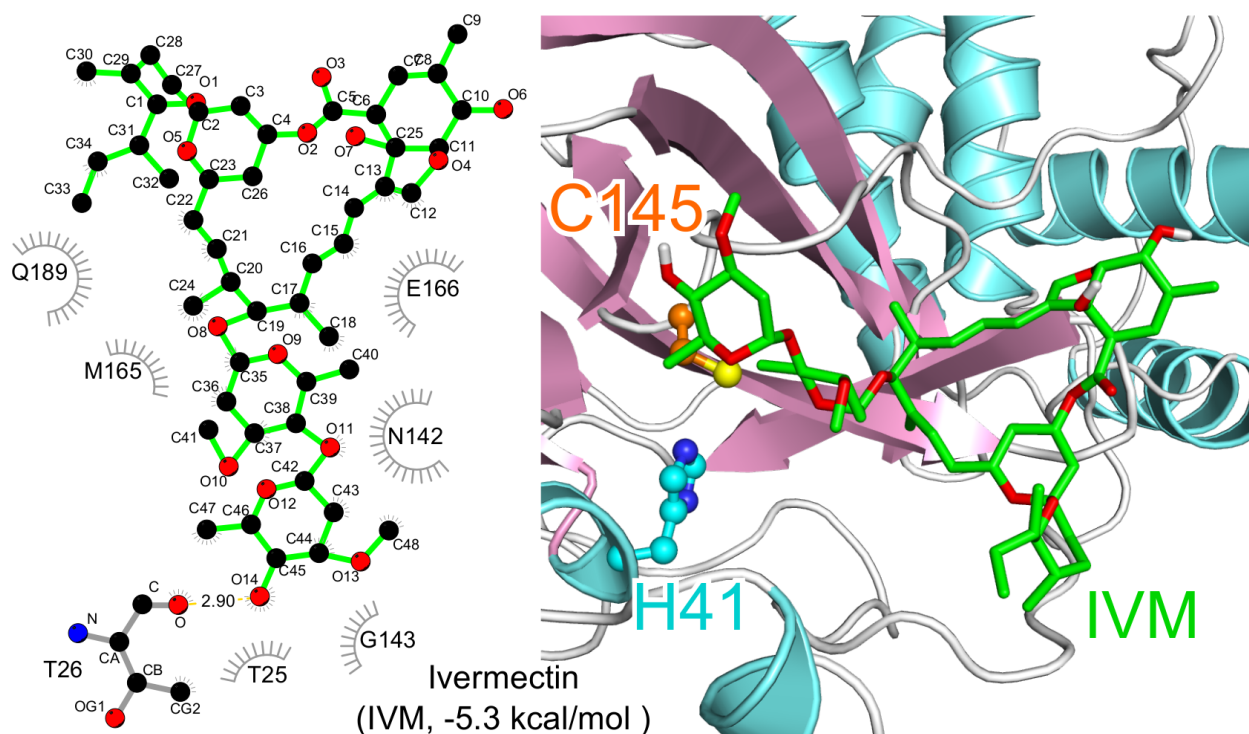
Figure S7: The M^pro–ivermectin interactions and binding pose of ivermectin in the active site of M^pro. The interactions are analyzed and plotted with LigPlot+ [36, 37]. See Figure 3 for an explanation of this representation. For the right panel, the same protein and ligand representation as well as color scheme as in Figure 1 is used.

High throughput virtual screening for M^pro inhibitors   A PREPRINT

Table S1: Names and properties of the top 100 compounds with lowest ΔG values and eight reference compounds obtained from redocking considering protein flexibility of M^pro.

No.	Library	Accession ID, Compound	ΔG (kcal/mol)	SMILES structure
1	DB	DB12424, MK-3207	−10.10	FC1=CC(=CC(F)=C1)[C@@H]1CNC2(CCCC2)C(=O)N1CC(=O)NC1=CC=C2C[C@]3(CC2=C1)C(=O)NC1=C3C=CC=N1
2	DB	DB06595, Midostaurin	−9.62	CO[C@@H]1[C@@](C[C@@H]2O[C@@]1(C)N1C3=C(C=CC=C3)C3=C1C1=C(C4=C(C=CC=C4)N21)C1=C3CNC1=O)N(C)C(=O)C1=CC=CC=C1
3	ZINC	ZINC2258952	−9.52	C1=C[C@H]2Cc3cc4nc(-c5ccc6ccccc6n5)ccc4cc3[C@H]2C=C1
4	DB	DB01126, Dutasteride	−9.48	[H][C@@]1(CC[C@@]2([H])[C@]3([H])CC[C@@]4([H])NC(=O)C=C[C@]4(C)[C@@]3([H])CC[C@]12C)C(=O)NC1=CC(=CC=C1C(F)(F)F)C(F)(F)F
5	DB	DB00624, Testosterone	−9.46	[C@]12([C@@H]([C@@]3([C@@H](CC2)C(C(=O)CC3)(C)C)C)CC[C@@]2(C1=CC[C@@]1([C@H]2CC(CC1)(C)C)C)C
6	ZINC	ZINC2815797	−9.42	O=C1[C@@H]2C=CCC3=C2C(=NN1c1ccc2ccccc2c1)c1c3ccccc1
7	ZINC	ZINC71784951	−9.38	c1cc2c3c([C@@H]4C[C@@H]5c6c7c(cccc7ccc6)[C@@H]5C[C@@H]24)cccc3c1
8	ZINC	ZINC1856250	−9.34	c1cc2ccc3OC4(c5ccccc5Nc5ccccc45)C=Cc3c2cc1
9	ZINC	ZINC2273497	−9.34	c1cc2ccc3OC4(c5ccccc5Oc5ccccc45)C=Cc3c2cc1
10	ZINC	ZINC68603990	−9.34	O=C1c2ccccc2C(c2ccccc12)[C@@H]1C=Cc2ccccc2C=C1
11	DB	DB15367, LY-2623091	−9.30	n1(C[C@H](CN1CCOCC1)N1C(NC(N)=O)=NC2=C1C=CC(\C=C1/C3=CC=CC=C3OC3=CC(F)=CC=C13)=C2
12	ZINC	ZINC101020226	−9.28	c1cc2c3c([C@@H]4C[C@@H]5c6c7c(cccc7ccc6)[C@H]5C[C@@H]24)cccc3c1
13	ZINC	ZINC2132060	−9.28	O=c1cc(c2cc3c(OC4(CC3)CCCCC4)cc2o1)c1ccccc1
14	NP	Amentoflavone	−9.28	c1(cc(c2c(c1)oc(cc2=O)c1cc(c(cc1)O)c1c(cc(c2c1oc(cc2=O)c1ccc(cc1)O)O)O)O)O
15	ZINC	ZINC107176	−9.26	O=C1[C@H]2[C@H]([C@H]3c4ccccc4[C@@H]2c2c3ccccc2)c2nc3ccccc3n12
16	ZINC	ZINC12378054	−9.24	O=C1[C@@H]2[C@H]([C@H]3c4ccccc4[C@@H]2c2c3ccccc2)c2nc3ccccc3n12
17	NP	Epitaraxerol	−9.20	[C@@H]12[C@@]([C@@H]3[C@](C4=CC[C@@]5([C@H]([C@@]4(CC3)C)CC(CC5)(C)C)C)(CC2)C)(CC[C@@H](C1(C)C)O)C
18	DB	DB02741, CD564	−9.18	C(=O)(c1ccc2C(C)(CCC(c2c1)(C)C)C)c1ccc2cc(ccc2c1)C(=O)O
19	NP	Glabrolide	−9.16	[C@]12([C@@H]([C@@]3([C@@H](CC2)C(C([C@H](CC3)O)C)C)C(=O)C=C2[C@]1(CC[C@@]1([C@H]2C[C@@]23[C@H]1OC(=O)[C@]3(C2)C)C)C)C
20	DB	DB07189	−9.16	[H][C@]1(C[C@@]11C[C@]([H])(NC1=O)C1=CC=C(OC2=CC(=NC3=C2C=CC=C3)C2=CC=CC=C2)C=C1)C(O)=O
21	ZINC	ZINC2679623	−9.16	O=C1OC(=N/C/1=C1c2ccccc2ccc1)c1c2ccccc2ccc1
22	ZINC	ZINC1687008	−9.16	C1C=C[C@H]2C(=C1)Cc1c3[C@@H]4C=CC=C4Cc3c3[C@@H]4C=CC=C4Cc3c21
23	ZINC	ZINC45063	−9.16	c1(nnc(o1)c1c2ccccc2ccc1)c1c(=O)oc2ccccc2c1
24	ZINC	ZINC21982660	−9.14	c1cc2c(c3c(ccc(c3)c3ccc4c([nH]c5ccccc45)cc3)[nH]2)cc1
25	ZINC	ZINC489027	−9.14	Cc1cc(=O)oc2c(c)c3occ(c3cc12)c1ccc2ccccc2c1
26	NP	Chamaecydin	−9.14	[C@H]12[C@@](C3=C(C(=O)C4=C5C3=C(C1)C(=O)[C@]15[C@@H]3[C@@](C3)(C[C@@H]1C[C@@H]4C)C)C)O)(CCCC2(C)C)C
27	NP	Zeylanone	−9.12	c1(c2c(C(=O)C3=C(C2=O)[C@]2(C)C(=O)c4cccc(c4C(=O)[C@@H]2C3)O)ccc1)O
28	DB	DB14957, MK-3577	−9.12	[C@@H](c1ccc(cc1)C(=O)NCCC(=O)O)(CCC)[C@@H](c1c2cc(cc(F)c2[nH]c1)C)c1ccc(C)cc1
29	DB	DB11913, LY-2090314	−9.12	c1(c2cc(F)cc3CN(C(=O)N4CCCCC4)CCn(c23)c1)C1=C(C(=O)NC1=O)c1cnc2ccccn12
30	ZINC	ZINC485867	−9.12	Cc1cc(=O)oc2cc3occ(c3cc12)c1ccc2ccccc2c1
31	ZINC	ZINC11157723	−9.12	O=C(c1cc2c(ccc3ccccc23)o1)c1ccc2NC(=O)CCc2c1
32	NP	Friedelinol	−9.10	[C@]12([C@]([C@H]3[C@](CC2)(CCC(C3)(C)C)C)(CC[C@@]2([C@H]1CC[C@]1([C@H]2CC[C@@H]([C@@H]1C)O)C)C)C
33	NP	Friedelin-2	−9.10	O=C1CC[C@@H]2[C@]([C@H]1C)(CC[C@H]1[C@@]2(C)CC[C@]2([C@@H]3CC(C)(C)CC[C@]3(C)CC[C@]12C)C
34	ZINC	ZINC55095248	−9.10	c1cn2c([C@@H](NCc3ccccc23)[C@H]2C=CC3=C4C(=CCC=C4)CC3=C2)c1
35	ZINC	ZINC488772	−9.08	Cc1c(c2c(oc1=O)cc1occ(c1c2)c1ccc2ccccc2c1)C
36	ZINC	ZINC71782709	−9.08	c1(cc(ccc1)c1c2ccccc2ccc1)c1cccc2ccccc12
37	ZINC	ZINC2084768	−9.06	CC1(CC(=O)c2c(nc3ncnn3c2c2ccc3ccccc3c2)C1)C
38	DB	DB08683	−9.04	[nH]1c(=O)c2c3c4ccccc4n4[C@H]5CC[C@@H](n6c(c(c7ccccc67)c2c1=O)c34)O5
39	DB	DB01897	−9.04	n1(n(nc([nH]1)/C=C/c1ccccc1)c1sc2ccccc2n1)c1ccc2c(=O)[nH][nH]c(=O)c2c1
40	ZINC	ZINC2258950	−9.04	c1(nc2cc3c([C@H]4CCCC[C@@H]4C3)cc2cc1)c1ccc2ccccc2n1
41	ZINC	ZINC436932948	−9.04	c1(noc(n1)c1ccc2ccccc2c1O)c1cc2c(c(C)c1)nc[nH]2
42	DB	DB08058	−9.02	c1(c(ccc(c1)Cc1n[nH]c(=O)c2ccccc12)F)C(=O)N1CCCNCC1
43	ZINC	ZINC3120897	−9.02	C(=C1c2ccccc2c2ccccc12)/C=c\1/n(c)c2ccccc2c1
44	ZINC	ZINC184381406	−9.02	O=C(N1C[C@H]2CCC(=O)c3cccc1c23)c1c(cc2ccccc2c1)O
45	ZINC	ZINC1018028	−9.02	O=C1C[C@@H](c2ccc3ccccc3c2N1)c1c2ccccc2ccc1
46	NP	Corilagin	−9.02	c12c(/C/3=C(/C(=C(\C=C=C4O[C@@]1(O3)O[C@H]1[C@H]([C@H](CO4)O[C@H]([C@@H]1O)OC(=O)c1cc(c(c(c1)O)O)O)O)O)/O)/O)\O)c(c(c(c2)O)O)O
47	NP	Ajugalactone	−9.02	[C@@H]1(CC[C@@]2([C@@]1(C(=O)C[C@H]1C2=CC(=O)[C@H]2[C@@]1(C[C@@H]([C@@H](C2)O)C)C)C)O)[C@@](C)(O)[C@@H]1OC(=C(C(C1)CC)C
48	NP	Sarsasapogenin	−9.00	[C@H]12[C@@]([C@H]3[C@H]([C@H]4[C@]([C@H]5[C@@H](C4)O[C@@]4([C@H]5C)CC[C@@H](C)C)(CC3)C)CC2)(CC[C@@H](C1)O)C
49	ZINC	ZINC55095248	−9.00	c1cn2c([C@@H]([NH2]Cc3ccccc23)[C@H]2C=CC3=C4C(=CCC=C4)CC3=C2)c1
50	ZINC	ZINC334157361	−9.00	ON1c2ccc3ccccc3c2CC=C1c1ccc2C=CC=c2n1O
51	DB	DB11769, Funapide	−8.98	c1cc2[C@]3(C(=O)N(c2cc1)Cc1oc(cc1)C(F)(F)F)c1cc2OCc2cc1OC3
52	DB	DB12457, Rimegepant	−8.96	[C@@H]1([C@H](N[C@H]1[C@@H](OC(=O)N2CCC(CC2)N2C(=O)NC3=NC=CC=C23)C2=C1C=CC=N2)C1=C(F)C(F)=CC=C1
53	NP	Friedelan-3-one	−8.96	[C@]12([C@]([C@H]3[C@](CC2)(CCC(C3)(C)C)C)(CC[C@@]2([C@H]1CC[C@]1([C@H]2CCC(=O)[C@@H]1C)C)C)C
54	ZINC	ZINC4557337	−8.96	c1cc2c(c3c(c(ccc3)c3cccc4c5ccccc5oc34)o2)cc1
55	ZINC	ZINC32105390	−8.96	O=C1c2ccccc2c2ccccc2C21c1ccccc1c1ccccc21
56	ZINC	ZINC126553	−8.96	c1(c2ccccc2c(nn1)Oc1ccc2ccccc2c1)c1ccccc1
57	ZINC	ZINC436932948	−8.96	c1(noc(n1)c1ccc2ccccc2c1O)c1cc2c(c(C)c1)[nH]cn2
58	ZINC	ZINC685683035	−8.96	c1(nc(no1)c1ccc2[nH]cnc2c1)c1cc2CCCC(=O)c2cc1
59	DB	DB09280, Lumacaftor	−8.96	N(C(=O)C1(CC1)c1cc2OC(F)(F)Oc2cc1)c1ccc(c(n1)c1cc(ccc1)C(=O)O)C
60	NP	Friedelin	−8.94	O=C1[C@@H]([C@@]2([C@H]([C@]3([C@@H]([C@@]4([C@]([C@H]5[C@@](CC4)(CCC(C5)(C)C)C)(CC3)C)C)CC2)C)CC1)C)C

High throughput virtual screening for M^{pro} inhibitors   A PREPRINT

Nr.	Library	Accession ID, Compound	ΔG (kcal/mol)	SMILES structure
61	DB	DB04408, Ncs-Chromophore	−8.94	[C@H]1([C@H](/C/2=C/C#C[C@@H]3[C@](C#CC2=C1)(O3)[C@H]1OC(=O)OC1)O[C@@H]1O [C@@H]([C@H]([C@@H]([C@H]1NC)O)O)C)OC(=O)c1c2c(c(C)cc(c2)OC)ccc1O
62	ZINC	ZINC55095247	−8.94	c1cn2c([C@H](NCc3ccccc23)[C@H]2C=CC3=C4C(=CCC=C4)CC3=C2)c1
63	ZINC	ZINC137526589	−8.94	O=C1[C@H]2[C@@H]([C@H]3c4ccccc4[C@@H]2c2c3cccc2)c2nc3ccccc3n12
64	ZINC	ZINC436932667	−8.94	c1(noc(n1)c1c(cc2ccccc2c1)O)c1cc2c(c(C)c1)[nH]cn2
65	ZINC	ZINC436931125	−8.94	c1(nc(no1)c1n[nH]c2ccccc12)c1c(O)c2ccccc2cc1
66	ZINC	ZINC4540403	−8.94	O=c1ccc2c(cc3occ(c3c2)c2ccc3ccccc3c2)o1
67	ZINC	ZINC13637421	−8.94	c1nn2c(nc(cc2c2ccccc2)c2ccc3ccccc3c2)n1
68	NP	Daturalactone	−8.94	C[C@@]12[C@@H]3[C@H]([C@H]4[C@]([C@H](CC4)[C@H](C)[C@H]4C[C@]5(C) O[C@@]5(C(=O)O4)C)(C)CC3)[C@H]3[C@@H]([C@@H]1OC=CC2=O)O3.C
69	DB	DB00872, Conivaptan	−8.94	c1(ccc(cc1)NC(=O)c1ccccc1c1ccccc1)C(=O)N1c2ccccc2c2c(CC1)[nH]c(C)n2
70	NP	Friedelane-3-one-28-al	−8.92	C1C(=O)[C@@H]([C@@]2([C@@H](C1)[C@]1([C@H](CC2)[C@@]2 ([C@@](CC1)([C@]1([C@@](CC2)(CCC(C1)(C)C)C=O)C)C)C)C)C)C
71	DB	DB08237	−8.92	[H][C@]1(O)C[C@@]([H])(O[C@]1([H])COP(O)(O)=O)N1C=NC2=C1N=C(NCC1=CC=CC3=C1C=CC=C3)NC2=O
72	DB	DB06896	−8.92	c1(F)c(ccc1)NC(=O)c1cccn(c1=O)c1ccc(F)cc1)Oc1ccnc2[nH]ccc12
73	ZINC	ZINC55095247	−8.92	c1cn2c([C@H]([NH2]Cc3ccccc23)[C@H]2C=CC3=C4C(=CCC=C4)CC3=C2)c1
74	ZINC	ZINC305332	−8.92	O=c1oc2cc(ccc2c2CCCc12)OCc1ccc2ccccc2c1
75	ZINC	ZINC2258952	−8.92	c1(nc2cc3c([C@@H]4C=CC=C[C@@H]4C3)cc2cc1)c1ccc2ccccc2n1
76	ZINC	ZINC436932078	−8.92	c1(noc(n1)c1cc2c(CCCC)cc1O)c1cc2c(c(C)c1)[nH]cn2
77	ZINC	ZINC436932078	−8.92	O=C(N1C[C@@H]2CCC(=O)c3ccccc123)c1cc(c2ccccc2c1)O
78	DB	DB15291, BMS-986142	−8.90	c1(c(F)cc(c2[nH]c3C[C@H](CCc3c12)C(C)(C)O)C(=O)N)c1cccc(c1C)n1c(=O)c2c(n(C)c1=O)c(ccc2)F
79	DB	DB12411, R-428	−8.90	NC1=NC(NC2=CC=C3CC[C@@H](CCC3=C2)N2CCCC2)=NN1C1=CC2=C(N=N1)C1=CC=CC=C1CCC2
80	ZINC	ZINC488832	−8.90	c1(cc(=O)oc2cc3occ(c3cc12)c1ccc2ccccc2c1)CC
81	ZINC	ZINC2258953	−8.90	c1(nc2cc3c([C@@H]4CCCC[C@@H]4C3)cc2cc1)c1ccc2ccccc2n1
82	ZINC	ZINC71784949	−8.90	c1cc2c3c([C@@H]4C[C@@H]5c6c7c(cccc7ccc6)[C@H]5C[C@@H]24)cccc3c1
83	ZINC	ZINC489321	−8.90	Cc1cc(=O)oc2cc(c3c(coc3c12)c1ccc2ccccc2c1)C
84	ZINC	ZINC137526643	−8.90	O=C1[C@@H]2[C@@H]([C@H]3c4ccccc4[C@@H]2c2c3cccc2)c2nc3ccccc3n12
85	ZINC	ZINC436930815	−8.90	c1(noc(n1)c1ccc2CCCCc2c1O)c1cc2c(c(C)c1)[nH]cn2
86	ZINC	ZINC439992388	−8.90	O=C(N1CCC[C@@]2(CCCc3ccccc23)C1)c1c2ccccc2n[nH]1
87	ZINC	EIADMN84449	−8.90	O=C(n1ccc2cc(ccc12)c1ccccc1)c1c2cn[nH]c2ccc1
88	ZINC	ZINC1169311334	−8.90	N(C(=O)c1c2CCCCCc2n[nH]1)C(C)(C)c1ccc2ccccc2c1
89	ZINC	ZINC436888106	−8.90	c1(nc(no1)c1ccc2OCCc2c1)c1c(O)c2ccccc2cc1
90	ZINC	ZINC1188912248	−8.90	O=C(Nc1c2ccccc2ccc1)Nc1cc2ccc(cc2oc1=O)O
91	ZINC	ZINC564612208	−8.90	c1(noc(n1)c1c2nccnc2ccc1)c1cc2ccccc2cc1
92	NP	Catechin-3-5-O-digallatE	−8.90	c1c(cc2c(c1OC(=O)c1cc(c(c(c1)O)O)O)C[C@@H]([C@H](O2)c1cc(c(cc1)O)OC(=O)c1cc(c(c(c1)O)O)O)O)O
93	ZINC	ZINC253590138	−8.88	C[C@@H]1C[C@@H]2CCc3ccc4ccccc4c3[C@H]2c2c3ccccc3ccc12
94	ZINC	ZINC970584	−8.88	O=C(N[C@H]1C=CC2=C3C(=CCC=C3)CC2=C1)c1c2ccccc2ccc1
95	ZINC	ZINC436933875	−8.88	c1(nc(no1)c1ccc2nccnc2c1)c1c(O)c2ccccc2cc1
96	ZINC	ZINC771478839	−8.88	c1(nc(no1)Cc1cc2n[nH]cc2cc1)c1c(O)c2ccccc2cc1
97	ZINC	ZINC184381390	−8.88	O=C(N1C[C@@H]2CCC(=O)c3ccccc123)c1c(cc2ccccc2c1)O
98	NP	Apigenin-4-O-α-L-rhamnopyranoside	−8.88	c1cc(ccc1c1oc2cc(cc(c2c(=O)c1)O)O)O[C@@H]1O[C@@H](C)[C@@H]([C@H]([C@H]1O)O)O
99	NP	Amataine	−8.88	O1[C@@]23N4C[C@]5([C@@H]1N1[C@H]6[C@@]7([C@H]5OCC7)CC(=C5Nc7c([C@]65CC1)cccc7) C(=O)OC)C[C@@H]3C[C@@]13[C@@H]5N(CC[C@]25c2c4c(ccc2)OC)CC[C@@H]1OCC3
100	NP/ZINC	ZINC3979028, Bilobetin	−8.86	c1(c2c(c(cc1O)O)c(=O)cc(o2)c1ccc(cc1)O)c1c(ccc(c1)c1oc2cc(cc(c2c(=O)c1)O)O)OC
101	REF.	Tideglusib	−7.94	n1(c(=O)n(c(=O)s1)Cc1ccccc1)c1c2c(CC=CC2)ccc1
102	REF.	Shikonin	−7.20	c12c(ccc(c1C(=O)C=C(C2=O)[C@H](O)CC=C(C)C)O)O
103	REF.	Cinanserin	−6.56	c1(c(cccc1)SCCCN(C)C)NC(=O)/C=C/c1ccccc1
104	REF.	Ebselen	−6.48	O=C1N([Se]C2=CC=CC12)C1=CC=CC=C1
105	REF.	N3	−6.38	[C@H](CC(C)C)(C(=O)N[C@@H](C[C@@H](CO)CCN)CCC(=O)OCc1ccccc1) NC(=O)[C@H](C(C)C)NC(=O)[C@H](C)NC(=O)c1noc(C)c1
106	REF.	Carmofur	−6.24	n1(C(=O)NCCCCCC)cc(F)c(=O)[nH]c1=O
107	REF.	Disulfiram	−4.22	S(C(=S)N(CC)CC)SC(=S)N(CC)CC
108	REF.	PX-12	−4.16	c1nc([nH]c1)SS[C@H](C)CC