ACS Pharmacology & Translational Science

pubs.acs.org/ptsci

Article

A Blueprint for High Affinity SARS-CoV-2 Mpro Inhibitors from Activity-Based Compound Library Screening Guided by Analysis of Protein Dynamics

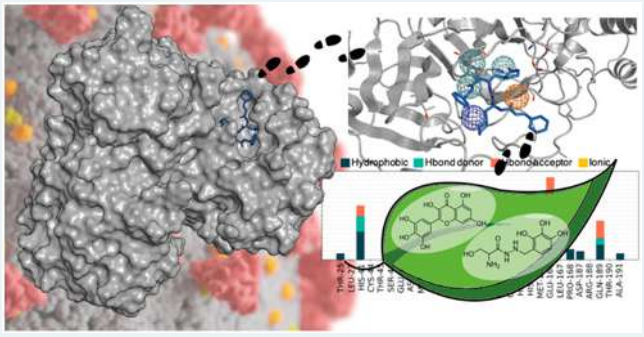
Jonas Gossen,[#] Simone Albani,[#] Anton Hanke,[#] Benjamin P. Joseph, Cathrine Bergh, Maria Kuzikov, Elisa Costanzi, Candida Manelfi, Paola Storici, Philip Gribbon, Andrea R. Beccari, Carmine Talarico, Francesca Spyrakis, Erik Lindahl,[#] Andrea Zaliani,[#] Paolo Carloni,[#] Rebecca C. Wade,[#] Francesco Musiani,[#] Daria B. Kokh,[#] and Giulia Rossetti*,[#]

Cite This: *ACS Pharmacol. Transl. Sci.* 2021, 4, 1079−1095

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The SARS-CoV-2 coronavirus outbreak continues to spread at a rapid rate worldwide. The main protease (Mpro) is an attractive target for anti-COVID-19 agents. Unexpected difficulties have been encountered in the design of specific inhibitors. Here, by analyzing an ensemble of ∼30 000 SARS-CoV-2 Mpro conformations from crystallographic studies and molecular simulations, we show that small structural variations in the binding site dramatically impact ligand binding properties. Hence, traditional druggability indices fail to adequately discriminate between highly and poorly druggable conformations of the binding site. By performing ∼200 virtual screenings of compound libraries on selected protein structures, we redefine the protein's druggability as the consensus chemical space arising from the multiple conformations of the binding site formed upon ligand binding. This procedure revealed a unique SARS-CoV-2 Mpro blueprint that led to a definition of a specific structure-based pharmacophore. The latter explains the poor transferability of potent SARS-CoV Mpro inhibitors to SARS-CoV-2 Mpro, despite the identical sequences of the active sites. Importantly, application of the pharmacophore predicted novel high affinity inhibitors of SARS-CoV-2 Mpro, that were validated by *in vitro* assays performed here and by a newly solved X-ray crystal structure. These results provide a strong basis for effective rational drug design campaigns against SARS-CoV-2 Mpro and a new computational approach to screen protein targets with malleable binding sites.



KEYWORDS: *conformational space, druggability, Mpro, molecular dynamics simulation, virtual screening, protein binding site flexibility, SARS-CoV-2*

In December 2019, a new coronavirus (CoV), belonging to the clade b of the Betacoronavirus viral genus, caused an outbreak of pulmonary disease in the Hubei province in China.[1,2] In the first months of 2020, the new pandemic spread globally and it is still continuing. The virus shares more than 80% of its genome with that of the SARS coronavirus discovered in 2002 (SARS-CoV).[1,2] Hence it has been named severe acute respiratory syndrome-coronavirus 2 (SARS-CoV-2) by the International Committee on Taxonomy of Viruses.

Interfering with viral replication is a promising strategy of treatment. In this context, the chymotrypsin-like proteinase (often referred to as the main protease, Mpro hereafter) is an excellent pharmaceutical target.[3,4] It does not depend on host immunogenic responses, and it is essential for generating the 16 nonstructural proteins, critical to the formation of the replicase complex.

Mpros are highly conserved enzymes across CoVs.[5,6] SARS-CoV Mpro was already suggested as one of the main drug targets in the pandemic associated with that virus, about 15 years ago.[7−9] Inhibitors of proteases (e.g. aspartyl protease) are also common drugs used in the clinic against other deadly viruses, e.g. HIV-1.[10]

The SARS-CoV Mpro active form is a homodimer (Figure 1A), with each monomer consisting of N-terminal, catalytic, and C-terminal regions (Figure 1B).[11] Mpros were shown to process

Received: December 17, 2020
Published: March 16, 2021

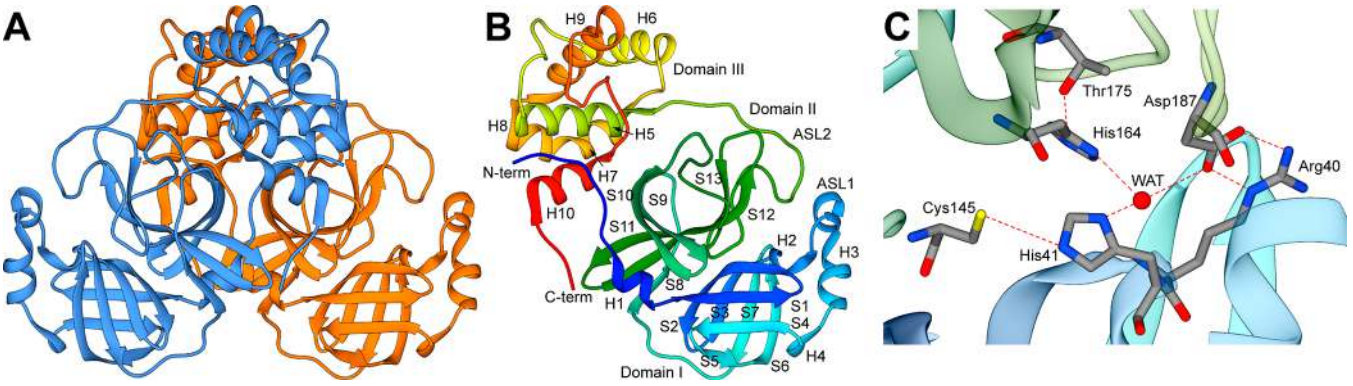


Figure 1. Structure of SARS-CoV-2 Mpro (PDB id: 6Y2E). (A) The enzyme is a homodimer.[14] (B) Each monomer consists of three domains (I−III): The chymotrypsin-like and picornavirus 3C protease−like domains I and II (in blue and green, respectively) form six-stranded antiparallel β-barrels, that harbor the substrate-binding site between them, including the ASL1 and ALS2 loops (residues 44−53 and 184−194, respectively). Domain III (in yellow-red) is a globular bundle formed by five helices, and it is involved in the dimerization of the protein. (C) Close-up of the active site and of the Hbond network. Atoms are in stick representation colored according to atom type, while Hbonds are indicated using dashed lines.

polyproteins on diverse cleavage sites, using a cysteine/histidine catalytic dyad:[12] the histidine (His41 in SARS-CoV-2) forms a hydrogen bond (Hbond) with a water molecule that, in turn, interacts with an aspartate (Asp187) and a histidine (His164) side chains. Asp187 is further stabilized through a salt-bridge with a nearby arginine (Arg40, Figure 1C). In this way, His41 can act as a base, extracting a proton from the catalytic cysteine (Cys145) side chain and activating it for the nucleophilic attack that cuts the polypeptide.

SARS-CoV-2 Mpro shares 96% sequence identity with Mpro from SARS-CoV (Supporting Information (SI), section S1, Figure S1A). Twelve residues differ between both Mpros, and only one, namely, Ser46 in SARS-CoV-2 (Ala46 in SARS-CoV), is located at the mouth of the active site cavity (Figure S1B). The binding sites share 100% of sequence identity (Figure S1A). Thus, exploiting the known libraries of SARS-CoV Mpro inhibitors has been a strategy followed by many research groups. Unfortunately, most SARS-CoV Mpro inhibitors with good (nM) activity against SARS-CoV Mpro *in vitro* and in cell-based assays, exhibited limited (sub-μM) potency against the protein from SARS-CoV-2 in enzymatic assays, and low-μM IC_{50} values (4−5 μM) in cell-based assays.[13,14]

A similar scenario has emerged from the virtual screening (VS) of Mpro inhibitors toward SARS-CoV-2. Indeed, none of these strategies, (i) repurposing of SARS-CoV drugs for SARS-CoV-2,[15,16] (ii) Deep Docking trained on SARS-CoV Mpro inhibitors,[17] (iii) libraries of the other SARS proteases,[18−20] and (iv) clinically approved drugs for other SARS Mpros or other similar proteases,[21−26] have led so far to clinical advances. Because only 12 residues, far from the binding site, differ between SARS-CoV Mpro and SARS-CoV-2 Mpro, the mutation of distant residues can substantially contribute to the binding site plasticity and to the ligand binding through allosteric regulation.[27] The inability to identify Mpro inhibitors for SARS-CoV-2 might also be due to the short amount of time passed since it is rise to urgency. Recently, however, it has been shown that the dipeptide prodrug GC376, and its parent compound GC373 inhibit the two proteases with IC_{50} values in the nanomolar range.[28] This suggests that, despite the intrinsic and significant differences between the two Mpros, common binding features against some classes of high-affinity ligands are retained.

Molecular dynamics (MD) simulations[15] provided hints to address this riddle: they showed that the SARS-CoV Mpro active sites display major differences in both shape and size. In particular, while both Mpros reduce their accessible volume upon inhibitor binding by approximately 20%, the maximal volume of the *holo* SARS-CoV Mpro active site is over 50% larger than that of SARS-CoV-2. In addition, the accessibility of the binding hotspots (i.e., the key residues for substrate binding) and the flexibility of one of the two loops delimiting the binding pockets (ASL1 in Figure 1B) differs between the two Mpros.[15] The simulations indicate that the binding sites of the two Mpros are dynamically diverse and that ligand binding can impact them differently.[29]

Therefore, transferable binding features across Mpros, as well as unique ones for SARS-CoV-2, are difficult to predict: the exceptional flexibility and plasticity of the binding site is here coupled with large adjustments of the cavity shape in response to the binding of an inhibitor. This clearly emerges from an analysis of the SARS-CoV-2 Mpro binding pocket's conformational changes (performed here) across the majority of the 196 X-ray crystal structures available in the Protein Data Bank up to September 30th 2020.[15,30−32] This flexibility makes a rational drug-design approach extremely challenging:[15,33] the screening potential of Mpro conformational space is too large, too flexible, and unpredictable, and the actual available binding space can differ significantly from ligand to ligand.[15,30,34]

It is therefore imperative to identify the relationship among SARS-CoV-2 conformational space, flexibility, druggability, and ligand binding. Here, we analyzed the mentioned 196 X-ray crystal structures, along with about ∼31 000 conformations extracted, not only from the longest (100 μs) MD simulation of SARS-CoV-2 Mpro so far,[35] but also from the binding site enhanced sampling simulations carried out here. Among these structures, we selected ∼30 000 conformations for which we systematically performed druggability analyses on the binding sites. The top 216 druggable structures were selected for virtual screening of a sample library of ∼13 500 ligands. The latter includes marketed drugs and compounds under development, the internal chemical libraries from Fraunhofer Institute and the Dompè pharma company, as well as the so-far known inhibitors of SARS-CoV Mpro from literature.

We redefine here the protein druggability in a new way exploiting the chemical space shaped by the different configurations of the binding site upon virtual screening. Specifically, we identified a consensus protein−ligand interaction fingerprint across the chemical space and the correspond-

https://doi.org/10.1021/acsptsci.0c00215
ACS Pharmacol. Transl. Sci. 2021, 4, 1079−1095

ing SARS-CoV-2 Mpro unique structure-based pharmacophore. The latter was able to identify known nM-binders ($IC_{50} \leq 400$ nM) of SARS-CoV-2 Mpro, distinguish these from micromolar inhibitors, and to identify Myricetin and Benserazide as novel nM inhibitors. The prediction was validated by an enzymatic activity binding assay. The calculated binding mode of Myricetin is in excellent agreement with a recently solved X-ray crystal structure (Kuzikov et al., submitted). The active pharmacophore also provides a rationale for the variability in ligand affinity of currently known SARS-CoV ligands and for the general lack of transferability of SARS-CoV ligands to SARS-CoV-2, shedding light on the complexity, plasticity, and druggability of SARS-CoV-2 Mpro.

RESULTS AND DISCUSSION

To understand how Mpro's binding site conformation affects the druggability and the chemical space of selected binders, we considered ~30 000 Mpro conformations spanning different binding site arrangements and flexibility obtained from different sources (Table 1). These include the following:

Table 1. Mpro Conformational Ensembles Considered in This Study

Mpro protein's conformational space			
name	total no. (N) of conformations analyzed	no. of binding site conformations analyzed (chain A and chain B)	no. of druggable binding site conformations screened (chain A and chain B)
X-ray ensemble	196	86	84
MSM ensemble (four-state model)/ MSM ensemble (three-state model)	40 (four-state model) 30 (three-state model)	80 (four-state model)	20 (four-state model)
MD10000	10 000	2000	16 + 40
LRIP/tC ensembles	21 994	21 994	16 + 40

ligands' conformational space		
name	total no. of molecules	notes
sample library	13 534	
active molecules	193	fraction of the sample library for which an experimental IC_{50} measure against SARS-CoV is available
crystallographic ligands	37	available from the Protein Data Bank
consensus molecules	33	fraction of the sample library that is found in common across the top 1% of the best-performing structures
consensus active	16	fraction of the active molecules found in the consensus molecules
SARS-CoV-2 Mpro binders	41	available from literature

(i) All the X-ray crystal structures deposited to date of *apo* and *holo* SARS-CoV-2 Mpro. After analyzing all of them, we selected 43 structures for follow-up analysis. These exhibit a RMSD lower than 1 Å with respect to the excluded ones. Therefore, they can be considered a good representative ensemble of the overall deposited SARS-CoV-2 X-ray crystal structures ("X-ray ensemble" hereafter, see Table S1).

(ii) Two sets of MD ensembles: (a) The "MD10000" ensemble, that is 10 000 frames, taken every 10 ns, of a 100 μs-long MD of *apo* Mpro from D. E. Shaw Research[35] and (b) the "MSM ensembles", two collections of 30 and 40 representative conformations extracted from a three-state and a four-state Markov state model (MSM), respectively. The MSM analysis was performed on the same 100 μs-long MD trajectory (see section S2, Figure S2). These are included to identify conformations representing structural changes potentially related to binding.

(iii) About 22 000 conformations obtained with two enhanced sampling methods implemented in the TRAPP (TRAnsient Pockets in Proteins) web server:[36] Enhanced sampling by Langevin Rotamerically Induced Perturbation (LRIP)[37] and constraint-based sampling by tConcoord[38] referred to as "LRIP/tC ensembles" hereafter. These ensembles were derived using the X-ray crystal structures of the protein in complex with the inhibitor N3 (PDB ID 6LU7) and with another α-ketoamide inhibitor (PDB ID 6Y2G).

Binding Site Features and Druggability. We calculated specific binding site parameters for the ensembles described above (see SI section S3 for the complete list). We discuss here the volume, and the hydrophobicity, because they turn out to be key descriptors of the SARS-CoV-2 Mpro active site shape and druggability (see below). The latter was evaluated with the "druggability" scores: SiteScore and Dscore, as derived from the SiteMap[39] tool implemented in the Schrödinger suite 2019−4 (Schrödinger, LLC, New York, NY, 2019) and the CNN and LR (Convolution Neural Network and Linear Regression) druggability models[40] as implemented in the TRAPP package.[41] Although TRAPP and SiteMap use different approaches for computing the pocket characteristics (3D grid-based versus residue-based), the trends in the computed parameters are similar.

From this analysis, we conclude the following:

(i) The binding site volumes computed with TRAPP for the *holo* X-ray crystal structures are distributed over a slightly larger and more variable range of values than that for the *apo* X-ray crystal structures (Figure 2A, Figure S3). The distribution of volumes is higher in the MD and LRIP/tC ensembles with respect to the X-ray. Indeed, the loop region shows a high amount of flexibility (see Figure S4 and paragraph below). The difference in volume distributions in the crystal structures could be caused by crystal packing as we show in SI section S3.2 (Table S3, Figure S5).

During the MD simulations, on the other hand, the large range and variability in binding site volume are associated with conformational changes of loops ASL1 (res. 44−53) and ASL2 (res. 184−194) (Figure 2B, Figure S4). It can be seen that ASL1 is more flexible than ASL2, from the MSM analysis (section S2) and by calculation of the residue occurrence in the binding site (section S3, Figure S3). Volume variability also results from the transient participation (with a frequency of ~25%) of the N- and C-terminal tails of the adjacent subunit in the binding site (Figure 2C). During the MD, these two termini move in the proximity of the pocket. For the seven *apo* X-ray crystal structures, where the termini are resolved (Tables S1,S2), the C-terminus is far away from the binding sites, whereas the N-terminus is always close by (see Figure 2D). In the *holo* crystal structures, the

https://doi.org/10.1021/acsptsci.0c00215
ACS Pharmacol. Transl. Sci. 2021, 4, 1079−1095

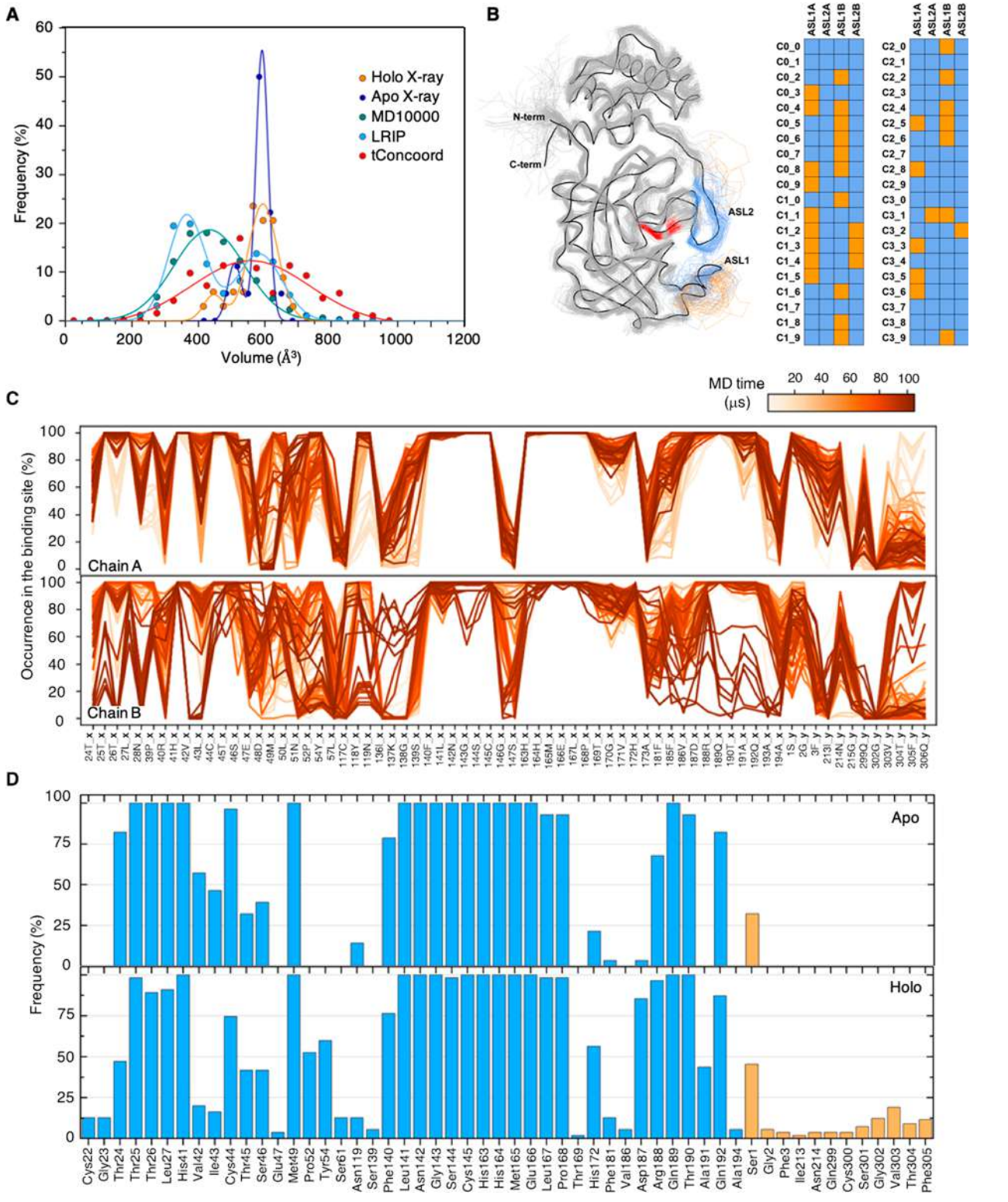


Figure 2. Binding site analysis of computational ensembles. (A) Binding site volume distributions from TRAPP for X-ray structures and MD/LRIP/tC ensembles. MD10000 contains conformations very similar to the frames extracted by the MSM analyses on the same trajectories, therefore these are not shown here. Binding site volume distributions from SiteMap are in Figure S3. (B) Analysis of ASL1 and ASL2 conformations (in chain A and B) in MSM (four-state model). Orange and blue squares refer to open and close conformations, respectively. The three-state model clusters are reported in Figure S3. (C) Binding site residues in the 100 μs-long MD trajectory: average occurrence in snapshots at 1 μs intervals for chains A and B, separately. Unlike chain A, extensive conformational changes are observed in the loop formed by residues 181−194 of chain B (see Figure S4). (D) Average occurrence of the binding site residues in the set of *apo* and *holo* X-ray structures (Table S1). Residues from the same chain are shown in blue, while residues from the adjacent chain are colored in orange. The percentage of each residue was calculated considering the number of structures for which that residue was resolved.

https://doi.org/10.1021/acsptsci.0c00215
ACS Pharmacol. Transl. Sci. 2021, 4, 1079−1095

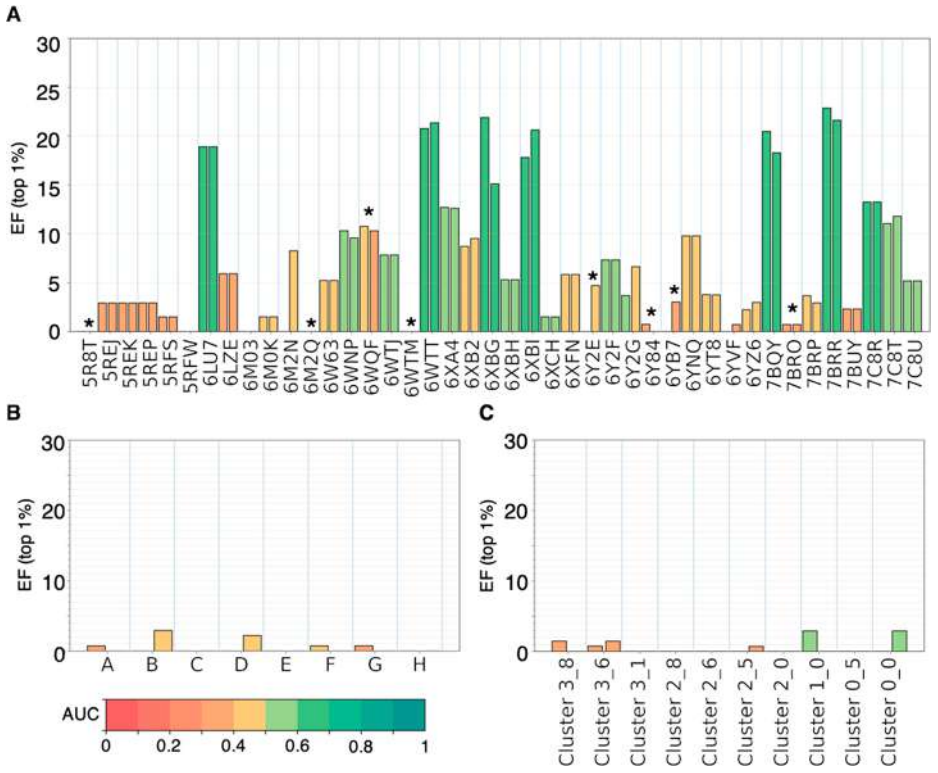


Figure 3. Virtual screening performance evaluation. Top 1% EF from the FRED virtual screenings performed on the binding sites in chains A and B in the X-ray crystal structures (A), in the MD10000 and LRIP/tC selection (section S3.4) (B) and in the MSM selection (C), that is, 10 structures with druggability index greater than 0.9 as extracted from the four-state model MSM ensemble. The bars are colored from red to dark green according to the value of the AUC in the ROC curves (color scale on the bottom left corner). The asterisk (∗) symbol in panel A highlights the *apo* structures. Glide results are shown in Figure S10.

N- and C-terminal tails of the adjacent subunit can both be found close to the binding site (see Figure 2D). A similar scenario is observed for the *holo* structures in the LRIP/tC ensemble, where both terminals are present in the binding pocket even more often (in about 40% of simulated structures, Figure S3).

(ii) The binding site hydrophobicity is here estimated in terms of hydrophobicity distribution across the different conformational ensembles, calculated with TRAPP and SiteMap. *Holo* X-ray crystal structures include conformations with higher hydrophobicity of the pocket with respect to the *apo* ones: these are complexes in which large ligands, with a molecular mass of 400 Da or greater, are covalently bound to Cys145 (PDB IDs 6LU7, 7BUY, and 7C8R). The MD and LRIP/tC ensembles instead can span from low to high hydrophobicity values (see section S3.3, Figure S6A).

Druggability. Here, we analyze the druggability as defined by the CNN and LR scores from TRAPP[40] along with DScore/SiteScore from SiteMap.[39,42] All pockets in the crystallographic structures (except one, PDB ID 6WTK) are scored as druggable: their druggability indices are above the scores' thresholds for druggability (0.8, 0.8, 0.5, 0.5 for SiteMap, DScore, LR, and CNN models, respectively, Figure S6. These thresholds were taken from Halgren *et al.*[42] and Yuan *et al.*[40], respectively). There is, however, no notable correlation between the druggability scores derived from the SiteMap and TRAPP methods. This is expected because the observed variations in the druggability index lie within the method prediction uncertainty. Despite the slightly lower druggability indices in SiteScore and TRAPP-LR for the *apo* X-ray crystal structures compared to the *holo* crystal structures, they are still predicted to be druggable within the uncertainty of the methods. In contrast, about 50% of the simulated structures (MD, LRIP, and tConcoord) were predicted not to be druggable (Figure S6).

The druggability scores of the simulated, and, more, of the X-ray crystal structures correlate with binding site hydrophobicity (Figure S6, section S3.3, and Tables S4,S5). The correlation with other binding site features is much smaller (see Tables S4,S5). We conclude that, as expected, the more hydrophobic the pocket is, the better it is scored.

Virtual Screening. We defined a sample library of a total of 13 534 compounds (Table 1, Figure S7). The library included commercialized drugs and compounds under development, the internal chemical library from Dompè pharma company, and compounds from the Fraunhofer Institute BROAD Repurposing Library, as well as known inhibitors of SARS-CoV Mpro. In particular the library included a set consisting of 180 compounds with pIC_{50} against SARS-CoV Mpro greater than 6 reported in the literature (active molecules, hereafter).[9,13,43−65] Our sample library is chemically very diverse as compared to the crystallographic ligands in complex with SARS-CoV-2 Mpro and the active molecules (see Figure S7). A more detailed chemoinformatic analysis is reported in section S4.1−2.

We selected SARS-CoV-2 Mpro conformations with scores above the druggability thresholds. These include all the X-ray crystal structures except 6WTK (42 structures), 10 MSM ensemble conformations (MSM selection), and 8 representatives of the top 10% scoring conformations from the MD10000

https://doi.org/10.1021/acsptsci.0c00215
ACS Pharmacol. Transl. Sci. 2021, 4, 1079−1095

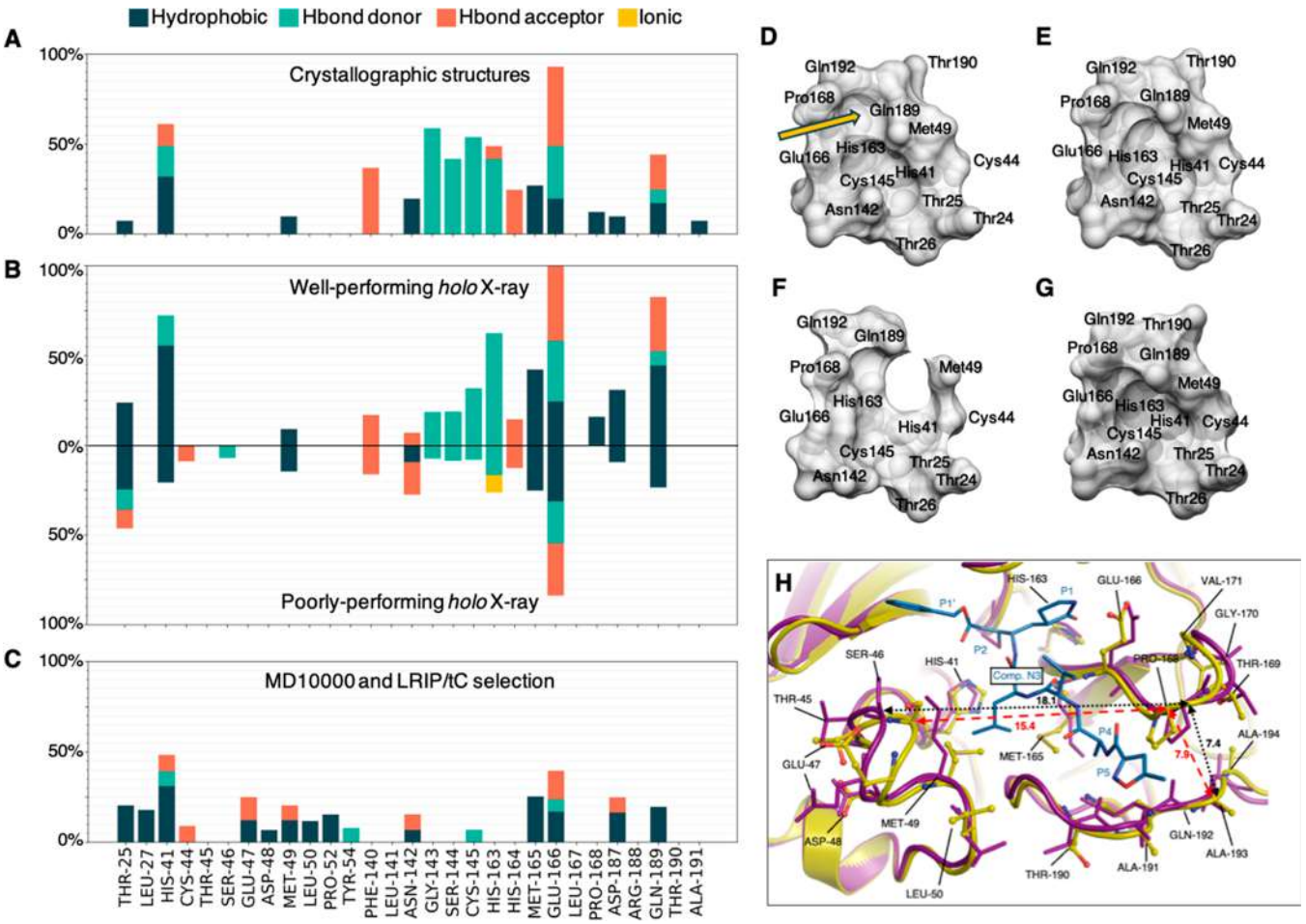


Figure 4. Virtual screening pose analysis of well performing and poorly performing receptor conformations. Average PLIF of (A) the crystal structures, (B) the top 1% of molecules from virtual screenings on well performing and poorly performing X-ray structures, and (C) the MD10000 and LRIP/tC selection. The PLIF of the top 1% of the MSM selection is reported in Figure S11. The occurrence of interactions between the ligands and the well- and poorly performing structures is plotted on the upper and lower half-plane, respectively. All bar plots were normalized with respect to the highest found occurrence (interactions with Glu166 in upper panel B). Binding site shape averaged over the well-performing (D), poorly performing (E), the MD10000/LRIP/tC selection (F), and the *apo* (G) structures (H). The yellow arrow in panel D is highlighting the "anchor" site. Superposition of a well performing and poorly performing crystal structures of SARS-CoV-2 Mpro (6LU7 and 5REK, well performing and poorly performing, respectively); ribbons are in purple and in yellow, respectively, while the N3 ligand is in blue carbon stick representation. The Ser46-Pro168 and Ala193-Pro168 Cα distances are highlighted. This latter panel was done following the scheme published in Kneller et al.[27]

and LRIP/tC ensembles (see Table 1, section S4.3, Table S6, Figures S8,S9).

The ligands were screened against the conformations using OpenEye FRED[66] and Schrödinger Suite Glide Version 85012.[67,68] We discuss here the results obtained with FRED. Those obtained with Glide present similar trends and are reported in the SI (section S4.4 and Figure S10). Also, we report here only the calculation results obtained with the MD10000/LRIP/tC and X-ray selections. The data obtained with the MSM selection are reported in the SI (section S4.5, Figure S11).

The quality of the virtual screenings was evaluated in terms of (i) enrichment factor (EF) defined as EF(1%) = (no. active molecules in the top 1%/no. molecules in the top 1%)/(active molecules in the whole set); (ii) receiver operating characteristic (ROC) curves, used to evaluate the true positive rate and the area under the curve (AUC).

The structures from the MD10000/LRIP/tC and MSM selections, along with the *apo* X-ray structures, exhibited a poor EF (below 5%), despite being identified as druggable by all the druggability prediction methods here implemented (Figure 3). The poorer performance of the *apo* X-ray crystal structures was expected since they exhibited overall lower druggability scores with respect to the *holo* structures (Figure S6F). This was not the case with the selected structures from the MD10000 and LRIP/tC ensembles (see section S4.3), where only the 10% of the structures with the highest druggability scores were used for screening. This suggests that the druggability prediction methods are not sensitive enough to distinguish between high- and low-EF conformations for the chemical space considered (see also discussion below). On the other hand, for the *holo* X-ray crystal structures, both "well-performing" (EF > 15%) and "poorly-performing" (EF < 5%) conformations were identified (Figure 3).

Next, we determined which of these ensembles exhibits a protein−ligand interaction fingerprint (PLIF) comparable to the one established by the ligands cocrystallized with SARS-CoV-2 Mpro. In the PLIF of the latter (Figure 4A), we observe an overall predominance of Hbond interactions over hydrophobic ones, with Cys145 (catalytic dyad), Gly143, Ser144, His163, and Glu166 as the most attractive residues to form

https://doi.org/10.1021/acsptsci.0c00215
ACS Pharmacol. Transl. Sci. 2021, 4, 1079−1095

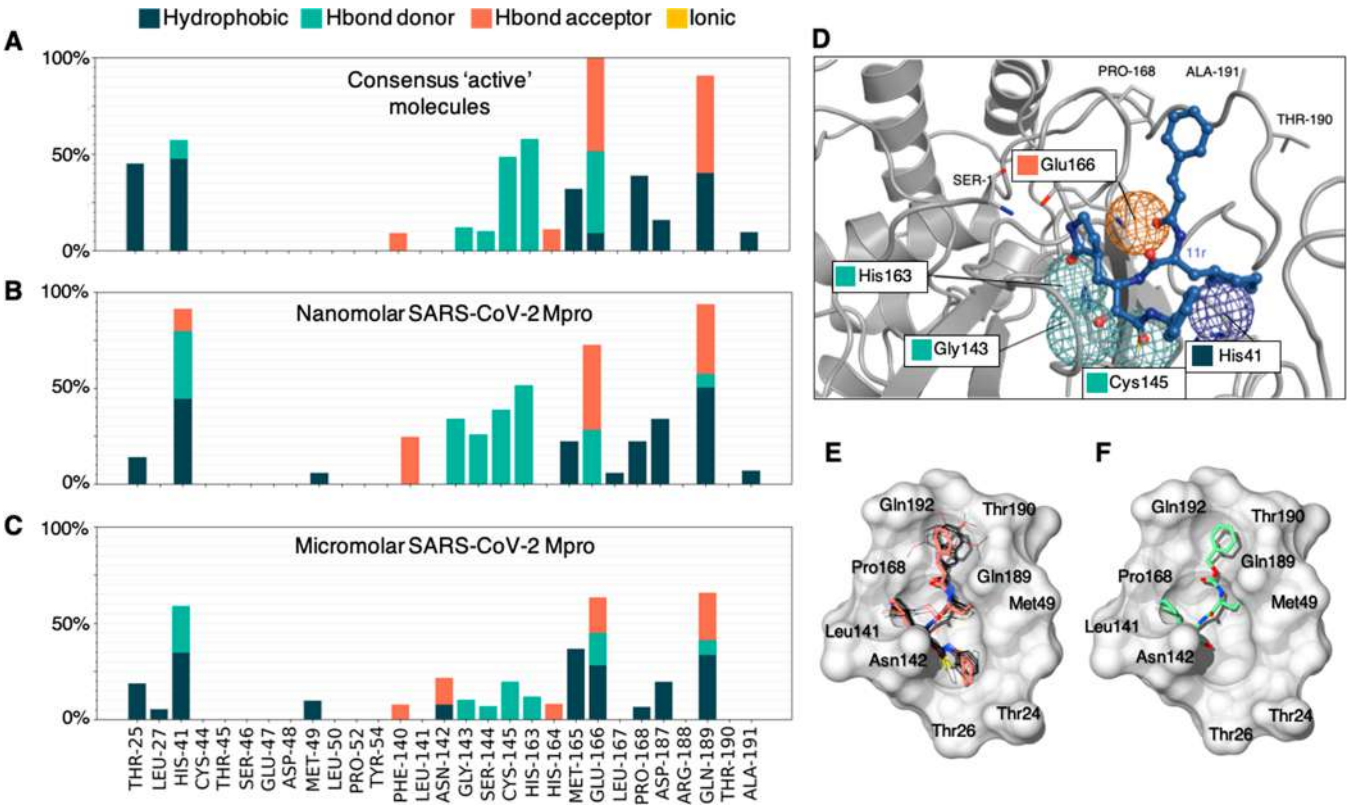


Figure 5. Virtual screening pose analysis of "consensus active" molecules. Average PLIF of the (A) 16 "consensus active" molecules, (B) 7 nM ($IC_{50} \leq$ 400 nM) inhibitors, and (C) 19 μM inhibitors docked onto the well-performing receptors. (D) Active-pharmacophore, where the five fundamental interactions (according to the selected cutoff, see methods) are displayed as spherical meshes. The docked pose of 11r, satisfying all the five interactions, is shown. (E) Docking poses of 16 "consensus active" molecules in SARS-CoV-2 Mpro (PDB ID 6LU7, chain A) binding site (black carbon representation). The 11r inhibitor pose is superimposed and highlighted in orange. (F) Docked pose of the inhibitor GC373.

Hbond interactions (comparable occurrence) with the ligands. The only exception is represented by His41 (catalytic dyad), which is similarly involved in Hbonds and hydrophobic interactions.

The PLIF of the well-performing conformations (Figure 4B, upper panel) matches the one of the crystallographic complexes (Figure 4A). Namely, the same hot-spot (i.e., preferential residues for ligand binding) residues emerge: the ligands form Hbonds with Cys145 (catalytic dyad), Gly143, Ser144, Gly166, and His163, as well as hydrophobic interactions with His41 (catalytic dyad). The main difference between the two PLIFs is in the lower occurrence of Hbonds involving Gly143, Ser144, and Cys145 (catalytic dyad) than in the crystallographic complexes, and in the higher occurrence of hydrophobic contacts with Thr25 and His41 (the second residue in the catalytic dyad). This change in the surrounding of the reactive cysteine, Cys145 (Thr25, Gly143, Ser144) might be due to the presence of several covalent ligands in the crystal structures. Covalent binding might locally alter the PLIF, and this effect is not considered in the virtual screening.

In all the other selections (i.e., poorly performing X-ray structures, Figure 4B lower panel, MD10000 and LRIP/tC selection, Figure 4C, and MSM selection, Figure S11), the key Hbonds above-discussed have an occurrence that is markedly lower than nonspecific hydrophobic interactions. Also, the latter interactions substantially decrease, including those with His41 (catalytic dyad). Moreover, in the MSM, MD10000 and LRIP/tC selections, ligands interact with almost all residues of the binding site (Figure 4C and Figure S11), but with an occurrence below 25% (for each interaction type) and with a strong predominance of hydrophobic interactions versus Hbonds. This points to a rather nonspecific binding of the screened molecules in the MD/MSM-selected structures.

Summarizing, we found that our evaluation of the virtual screening procedure correctly identifies the conformations able to provide the most similar PLIF to the known crystallized ligands of SARS-CoV-2 Mpro.

To rationalize these dramatic differences in the PLIF across the ensembles of structures, we compared their average binding site shapes. We found that the residues in the MSM, MD10000, and LRIP/tC selections are distributed over a larger volume than in the crystal structures (Figure 4D–F); therefore, the spatial location of the hotspots (i.e., Hbond donors/acceptors, hydrophobic patches, charges) is significantly different. Even reselecting the conformations from the MD-ensembles using the similarity with respect to well-performing structures as criterion (i.e., root mean square deviation, RMSD), did not provide a satisfying performance, nor the optimal spatial location of the hotspots (section S5, Figure S12), indicating that key features for obtaining high EFs in the virtual screening, that is, the precise placement of interacting residues, are missing.

Taken together, these results may explain why the druggability indices are unable to distinguish the binding site features linked to a good performance in virtual screening: very subtle variations of the conformations of the binding site residues induced upon binding, and therefore not present in the structural ensembles generated in the absence of a ligand, can lead to significant differences in the EF (see also Figure S13).

https://doi.org/10.1021/acsptsci.0c00215
ACS Pharmacol. Transl. Sci. 2021, 4, 1079−1095

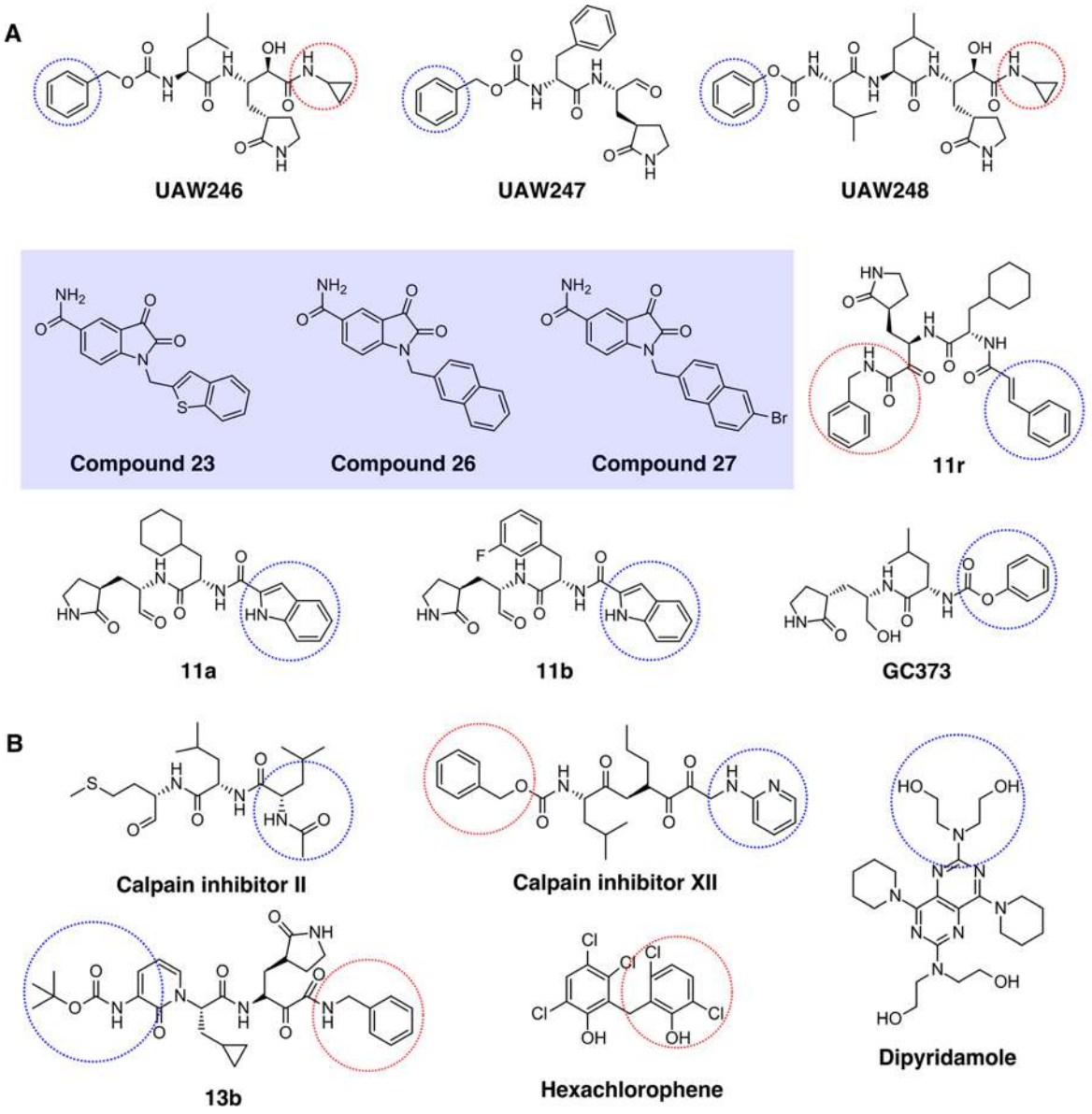


Figure 6. Chemical Structures of selected (A) nM and (B) sub-μM inhibitors of SARS-CoV-2 Mpro with a binding pose. The blue circles highlight the moieties buried in the "anchor site", while the orange circles show the portion of the molecules in proximity to Thr24, Thr25, Thr26. The docking poses of Compounds 23, 26, and 27, depicted in the blue rectangular area, do not present atoms in the surrounding of these two subpockets and were obtained from literature[78] during the submission process.

These subtle variations are very challenging to discriminate in terms of druggability indices.

The Active-Pharmacophore and the SARS-CoV-2 Blueprint on the Chemical Space. To identify the relevant ligand features that might relate to high affinity binding, we extract here the consensus chemical space, defined by the common ligands across the top 1% of the well-performing structures in the virtual screening. These are 32 molecules, 16 of which belong to the "active" molecules in our library ("consensus active" hereafter, Figure S14 and Table 1). We next calculated the corresponding pharmacophore (i.e., an ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target) and linearly combined it with the X-ray pharmacophore. This is done to consider possible additional features coming from covalent binding that cannot be covered by the virtual screening protocol. The consensus pharmacophore combined with the X-ray pharmacophore constitutes the "active-pharmacophore", hereafter (see Methods, Figure 5).

Next, we tested the predictive power of our active-pharmacophore in discriminating the higher affinity binders across all the so-far known SARS-CoV-2 Mpro inhibitors: these are 46 molecules coming from the papers published until November 20th 2020, which were not included in our sample library and that display a measured affinity spanning from 30 nM to 125 μM.[14,28,69−75] Some of these molecules were excluded by the docking software due to their excessive size or due to the presence of metals (e.g. Candesartan Cilexetil, Evans blue, Phenylmercuric acetate).

For this purpose, we calculated the Dice coefficient, which measures the number of features in common between the molecule and the active-pharmacophore, relative to the average number of features present.[76] When scoring the 46 known SARS-CoV-2 Mpro binders according to the Dice coefficient,

https://doi.org/10.1021/acsptsci.0c00215
ACS Pharmacol. Transl. Sci. 2021, 4, 1079−1095

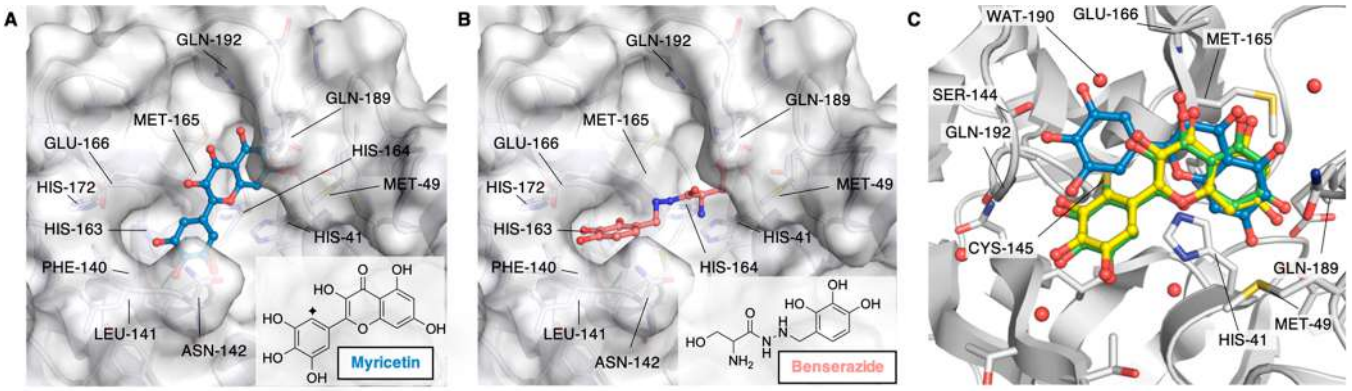


Figure 7. Binding poses of predicted high affinity ligands. Binding poses of Myricetin (EOS100914, A) and Benserazide (EOS100736, B), predicted to be high affinity binders using our active-pharmacophore model, and experimentally confirmed to be nM SARS-CoV-2 Mpro inhibitors. The protein structure is shown as a white surface (PDB ID 6WTT, chain A), while Myricetin and Benserazide are shown in blue and coral ball-and-sticks, respectively. The poses shown here are the best scored ones according to the Dice coefficient. The insert panels show the molecular formulas of Myricetin and Benserazide. The diamond symbol in the scheme of panel A highlights the position of the nucleophilic attack by Cys145 on Myricetin. (C) Overlay of crystal structure (PDB ID 7B3E, green), docked (blue, RMSD 3.14 Å), and refined (yellow, RMSD 0.46 Å) binding poses of Myricetin. Binding pocket residues are shown in white ribbons and sticks with heteroatoms colored according to the atom type. The orientation of panel C was rotated with respect of those of panels A and B to show the covalent bond found in the X-ray crystal structure between Cys145 and the Myricetin reactive carbon.

the highest scored molecules are 11a, 11b, 11r, UAWJ246, UAWJ247, UAWJ248, Compound 23, Compound 26, Compound 27, and CG373 (see Figure S15), which are also the highest affinity ($IC_{50} \leq 400$ nM) SARS-CoV-2 Mpro binders (nM-binders, hereafter). On the other hand, it is not possible to discriminate the sub-μM-binders of SARS-CoV-2 Mpro (400 nM < $IC_{50} \leq 1000$ nM) from the μM ones ($IC_{50} > 1000$ nM) (see Figure S16). These results have to be taken with great care given the fact we are comparing assay-dependent IC_{50} values coming from different laboratories. Also, several of these inhibitors are predicted to be covalent binders, which further complicates the use of their respective IC_{50} values (see the discussion offered in the Limitation paragraph). Therefore, in the next section, we analyzed the chemical space shaped by the well-performing conformations upon ligand binding, and offer a rationale for the predictive power of our active-pharmacophore in identifying nM-binders of SARS-CoV-2 Mpro.

Rationalization of the Active-Pharmacophore and SARS-CoV to -CoV-2 Mpro Ligands' Transferability. The PLIF of the consensus chemical space is dominated by the "consensus active" ones (the latter PLIF is almost identical to the former one, Figure S17A) and it shows a predominance of Hbond interactions with His163, Glu166, Gln189, and Cys145 (Figure 5A), as well as hydrophobic interactions with Thr25, His41, Met165, Pro168, and Gln189. Accordingly, the same trend can be seen for the PLIF of the SARS-CoV-2 Mpro nM-binders (Figure 5B), the only differences being (i) the additional high occurrence of Gly143 and Ser144 Hbonds (as already observed in the PLIF of the X-ray structures), and (ii) the lower occurrence of hydrophobic interactions with Thr25.

When the binding poses of the "consensus active" molecules and the nM-binders of SARS-CoV-2 Mpro (Section S6) are compared, we found indeed that the indole group (in 11a, 11b, and 7 of the 16 molecules of the "consensus active" set) or the benzyl group (in GC373, 11r, and 6 of the 16 molecules of the "consensus active" set), or the benzimidazole group (in 1 of the 16 of the "consensus active" set) is buried in the upper subcavity defined by residues Glu166, Pro168, Gln182, Gln189, and Thr190 (Figure 5E,F). Notably, this cavity region which consists of β-sheets (residues from Tyr161 to Asp176) and the coil (residues from Gly183 to Ala194) was previously denoted as the "anchor site".[77] This region was shrunk in the poorly performing structures, further validating the importance of this part of the binding side and also the quality of our model that correctly excluded the conformations potentially incompatible with nM-binders. Instead, the benzothiazole moiety of the "consensus active" is placed in the lower part of the binding cavity defined by Thr24, Thr25, and Thr26. This benzothiazole moiety is absent in the SARS-CoV-2 Mpro nM-binders, possibly explaining the lower occurrence of hydrophobic interactions with Thr25.

These findings suggest that the binding to the lower part of the binding site (Thr24, Thr25, Thr26) is not a relevant feature for the nM affinity of SARS-CoV-2 ligands. In contrast, the high occurrence of Gly143 and Ser144 Hbonds appears to be a signature of nM-binders of SARS-CoV-2 Mpro, also found in the PLIF of the known X-ray ligands of SARS CoV-2 complexes. Notably, the formation of these two Hbonds appear to be significantly hampered in the "consensus active" set due to the presence of the above-mentioned benzothiazole moiety, that seems to compromise the juxtaposition of the Hbond acceptors of the ligands. Accordingly, none of the SARS-CoV-2 nM-binders display benzothiazole or analogous bulky aromatic groups in such a position (Figure S14, Figure 6).

Analysis of all 166 *holo* SARS-CoV-2 Mpro crystal structures also showed that all their ligands except two (PDB IDs 7JKV and 6XR3) do not have a benzothiazole or analogous bulky aromatic groups in the Thr24, Thr25, Thr26 subpocket (Figure S18). Concerning the μM binders known so far, only very few of them are predicted to have a bulky group in such a position (Figure S16). In other words, our results suggest that the bottom part of the binding cavity in SARS-CoV-2 Mpro should only host small aromatic/hydrophobic moieties (or nothing at all) to facilitate the formation of Gly143 and Ser144 Hbonds, the latter being a signature of the currently known nM-binders to SARS-CoV-2 Mpro.

Currently, the X-ray structure of GC373 in complex with SARS-CoV-2 Mpro was solved (PDB ID: 7BRR, recently superseded by PDB ID 7D1M (October 28th 2020). The ligand in the crystal structure appears in two different conformations, one resembling the predicted pose from us, in which the benzyl

https://doi.org/10.1021/acsptsci.0c00215
ACS Pharmacol. Transl. Sci. 2021, 4, 1079−1095

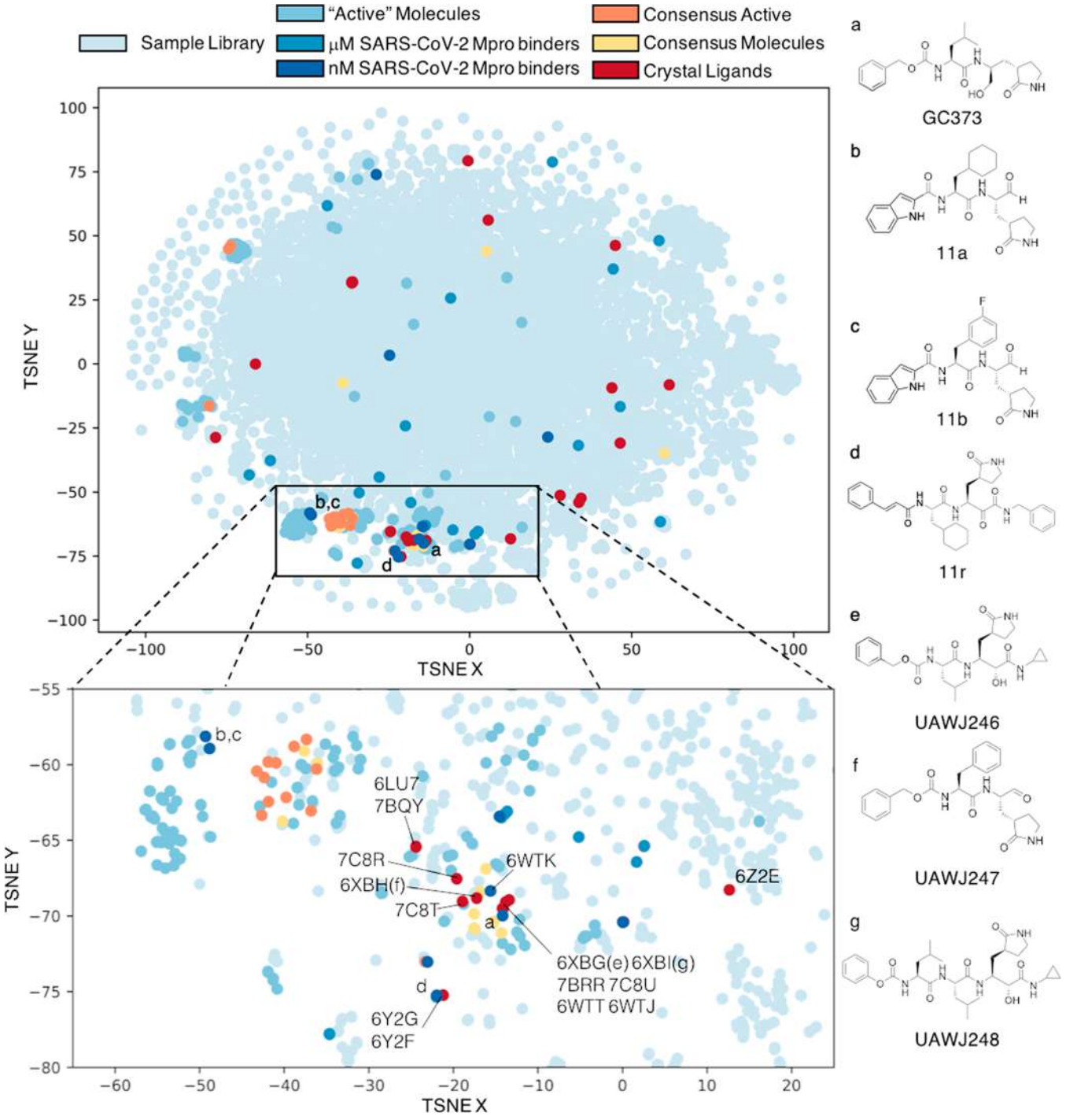


Figure 8. t-Distributed Stochastic Neighbor Embedding (t-SNE) plot of the sample chemical library screened (see SI equation S4.2 for details). The sample library, the molecules denoted as "active" (due to their experimental binding affinity toward SARS-CoV), as well as 7 known SARS-CoV-2 Mpro binders are all plotted in different shades of blue. The 2D representations of a selection of nM affinity SARS-CoV-2 Mpro binders colored in the darkest blue shade (labeled a–d) are shown on the side. Ligands identified in the top 100 of the well-performing structures ("consensus") are colored in yellow, while the subset of those that also have a high affinity toward SARS-CoV in experiments are plotted in orange. Lastly, the cocrystallized ligands are shown in red, with selected ligands shown in 2D representation on the side (labeled e–g). The inset with a magnified portion of the t-SNE plot is reported at the bottom of the figure. The "active" molecules appear to be more chemically diverse than the SARS CoV-2 Mpro cocrystallized ligands since they are spread over all the t-SNE plot, while the cocrystallized ligands mostly cluster in the bottom part of the plot. This region corresponds to peptides covalently bound to SARS-CoV-2 Mpro-C145 (PDB IDs: 6LU7, 6LZE, 6M0K, 6WTJ, 6WTK, 6WTT, 6XA4, 6XBH, 6XBG, 6XBI, 6Y2F, 6Y2G, 6YZ6, 6Z2E, 7BQY, 7BRR, 7C8R, 7C8T, and 7C8U, see Table S1).

group of GC373 is not buried in the upper subcavity defined by residues Glu166, Pro168, Gln182, Gln189, and Thr190; the other in which this benzyl group is exposed toward the solvent. Yet, when the crystallographic complex undergoes 500 ns of MD simulations, the pose where the aromatic ring is exposed toward the solvent rearranges as in the predicted docking pose (see Section S7 and Figure S19). Such results further validate our active-pharmacophore. The latter is not only able to

https://doi.org/10.1021/acsptsci.0c00215
ACS Pharmacol. Transl. Sci. 2021, 4, 1079–1095

discriminate the nM-binders of SARS-CoV-2 Mpro ($IC_{50} \leq 400$ nM) from the rest, but it also identifies key specific transferable and not-transferable binding features of nM SARS-CoV Mpro binders to SARS-CoV-2 Mpro ones.

Taken together, these results suggest that our active-pharmacophore is a fair representation of the SARS-CoV-2 Mpro blueprint in the chemical space. Namely, it correctly represents a set of binding features compatible with the induced SARS-CoV-2 conformational space of the binding site. The latter is in part determined by the ligand upon binding and in part it depends on the residues differing in SARS-CoV-2 Mpro with respect to SARS-CoV Mpro, as also shown in Bzówka et al.[15]

A cheminformatics analysis of the molecules that were active, but not part of our "consensus active" set is offered in the SI (see section S8, Figure S20).

Identification of nM-Binders of SARS-CoV-2 Mpro. We considered a set of publicly available compounds within the E4C network,[79] from the EU-OPENSCREEN Bioactive Compound-Library,[80] coming from the PROBE MINER repository.[79,81] The set was rescored based on the Dice similarity of their docked pose to our active-pharmacophore (see Methods). Benserazide (EOS100736) and Myricetin (EOS100814) compounds (see schemes in Figure 7A,B) were predicted as nM-binders of SARS-CoV-2 Mpro candidates. SARS-CoV-2 Mpro biochemical assays performed here established the accuracy of our predictions by measuring IC_{50} values as low as 140 nM and 220 nM, respectively (see section S9.1−2), Figure S21).

The best docking pose of Myricetin, as coming out from our virtual screening procedure, shows an orientation which is comparable to the one observed in the newly solved X-ray structure with PDB ID 7B3E (resolution 1.77 Å, see Figure 7C). In this pose, the bicyclic rings in the two structures nicely overlap, while the 3,4,5-trihydroxyphenyl moiety is rotated in our predicted pose with respect to the crystallographic one. By refining the docking pose (see section S9.3 and Figure 7C) both the bicyclic ring and the 3,4,5-trihydroxyphenyl moiety assumes an orientation nearly identical to the one found in the X-ray pose after covalent binding with Cys145, with an overall RMSD of 0.46 Å between the refined predicted pose and the crystallographic one. Interestingly, Baicalin features the same isoflavone scaffold as Myricetin, yet it binds the protein with a different orientation, as shown by X-ray structure determination (PDB ID 6M2N). Our procedure predicted this orientation, although the best binding pose differed more significantly from the crystallographic one than the one obtained with Myricetin (see section S9.4, Figure S22).

Myricetin and Benserazide contain polyhydroxy-phenolic moieties, which are considered promiscuous due to their redox features and also to the presence of a high number of close Hbond acceptor/donor sites that allow them to satisfy several 3D-pharmacophores. Nonetheless, these classes of compounds have, respectively, reached approved clinical usage (i.e., for Parkinson's disease[82] and alcohol use disorder[83]) and are in use in our diet like other polyhydroxyphenol-containing products.[84] Also, Quercetin, structurally similar to Myricetin, was identified as a mild inhibitor of SARS-CoV-2 Mpro ($K_i \sim 7\ \mu M$).[85]

SARS-CoV-2 Mpro is an important target for COVID-19 drug discovery because of its key role for viral replication and low similarity with human proteases.[3,4] Given its conserved nature with respect to the other Mpros across Coronaviruses and the presence of a huge number of crystallized structures (*apo* and *holo*), several drug repurposing and structure-based drug design campaigns have been conducted.[17−26] Unfortunately, this has so far led to only 10 SARS-CoV-2 Mpro inhibitors in the nM range ($IC_{50} \leq 400$ nM).[78,86] This contrasts with SARS-CoV Mpro, for which 127 inhibitors are known in this range.[44,45,50,52,54,58−60,63] The observed difficulties in identifying potent SARS-CoV-2 Mpro inhibitors was suggested to arise from the large plasticity of the binding site,[15] along with other factors (also observed for SARS-CoV Mpro), including induced-fit conformational changes and formation of covalent bonds upon ligand binding.[27,29,77] Therefore, the available binding space can differ significantly from ligand to ligand.

Accounting for receptor binding site flexibility in molecular docking is a significant challenge.[87] This can be partially overcome with a careful choice of the most appropriate receptor and reference ligand(s) or by performing ensemble docking approaches.[88,89] While it seems logical to employ multiple protein structures and ligands where available, very few published studies have systematically evaluated the impact of using additional information on proteins' and ligands' structure.[90] These studies arrive at the conclusion that an alternative structure-based design approach may be needed to define pharmacophores based on the binding site and use them to search large chemical databases.[91,92]

Our paper exploits the particularly large amount of structural information available for Mpro (∼200 X-ray structures in the *apo* and in the *holo* forms), along with a very long MD simulation from D. E. Shaw Research[35] and structures generated here by enhanced sampling of the binding site dynamics. First, we have determined the potential druggability of each of the ∼30 000 Mpro conformations generated from these sources, as calculated using TRAPP and SiteMap druggability tools.[36,39,41] Next, we have used a sample library to understand how selected potential high-druggable protein conformations perform when probed with a diverse chemical space, here defined by ∼13 000 compounds (see t-SNE plot in Figure 8, and methods for details).

Our library included also "active" molecules, that is, molecules that are known to bind with nM affinity ($pIC_{50} > 6$) to SARS-CoV Mpro. Therefore, virtual screenings against ∼200 protein conformations were performed. We found that only a few of these highly druggable ("well-performing") conformations recognize a sufficiently high percentage of such "active" molecules: the ones in common among the well-performing structures, "consensus active" molecules, represent a small subgroup of the overall "active" molecules' ensemble with specific features. In particular, the "consensus active" (16 molecules) are mostly clustered in two main areas of the t-SNE plot, both corresponding to peptidomimetic structures but differing from each other by the presence of a benzothiazole moiety and an additional peptide bond. The specific protein−ligand interaction fingerprint (PLIF) of the "consensus active" molecules strikingly resembles the one emerging from the SARS-CoV-2 Mpro cocrystallized ligands. The latter indeed cluster in the same region of the t-SNE plot. Notably, no consensus was found for the poorly performing structures.

We then combined the crystallographic and the virtual screening PLIFs (i.e., the chemical space emerging from experimental structures and the chemical space selected upon virtual screening).

Within the limitations of the procedure (see Limitations paragraph), we obtained an "active-pharmacophore" that we first used against a selection of SARS-CoV-2 Mpro binders (46

https://doi.org/10.1021/acsptsci.0c00215
ACS Pharmacol. Transl. Sci. 2021, 4, 1079−1095

molecules): the latter are very diverse and they are spread overall the t-SNE plot (Figure 8).

The active-pharmacophore could predict known nM-binders for SARS-CoV-2 Mpro (12 molecules out of the total of 46), which are also clustered in the peptides and peptidomimetics region of the t-SNE plot, and discriminate these from the μM ones. Moreover, it could also discriminate the transferable from the nontransferable binding features from SARS-CoV to SARS-CoV-2 Mpro.

The former include the interaction with the catalytic dyad residues along with (i) His163, the mutation of which to Ala inactivates SARS-CoV Mpro[93] and (ii) Glu166, which plays a role in the dimerization (required for enzymatic activity) in SARS-CoV.[94] In addition, its interactions with the N-finger of the other subunit assist the correct orientation of residues in the binding pocket for both proteins.[94,95] Also (iii) Gln189, which correlates evolutionally with residues from the Cys44-Pro52 loop in both proteins, which was shown to regulate ligand entrance to the binding site[15] in both proteins; and (iv) Ser144, the mutation of which to Ala hampers the catalytic activity in SARS-Cov Mpro.[96]

The nontransferable binding features include the ability to place large hydrophobic/aromatic groups in the part of the cavity defined by Thr25 and Thr26 that is partially lost in SARS-CoV-2 Mpro compared to SARS-CoV Mpro. This appears to affect Hbonds with Gly143 and Ser144. Accordingly, this cavity is empty or occupied by a smaller aromatic group such as the benzyl ring in all the known nM-binders and the cocrystallized ligands of SARS-CoV-2 Mpro. In contrast, several of the known nM-binders of SARS-CoV Mpro have benzothiazole or analogous bulky aromatic groups in this position.

We finally used our active-pharmacophore against a public library of compounds. We predicted two ligands to be nM for SARS-CoV-2: Benserazide and Myricetin. While in-cell anti-cytopathic effects and virus yield reduction assays showed a modest effect of those compounds, they are nevertheless confirmed to be nM binders of SARS-CoV-2 Mpro in binding assays (Figure S21).[97] Our predicted pose of Myricetin is in excellent agreement with the independently solved X-ray crystal structure of the complex (Figure 7C). This was not expected since Baicalin, bearing the same isoflavone scaffold of Myricitin, binds in a reversed orientation (PDB ID 6M2N). Thus, the pharmacophore not only successfully predicts poses in the highly flexible binding site of SARS-CoV-2 Mpro, but it also discriminates between different orientations of quite similar chemical scaffolds.

Our methodological approach also demonstrates that only a small fraction of the binding sites of the *apo* protein from crystal structures or simulations are similar to those of the well-performing *holo* structures. However, even the conformations with high structural similarity and high druggability scores, generated by molecular dynamics simulations, yielded low enrichment factors in virtual screening. Thus, druggability assessment methods fail to discriminate between small structural variations of the binding site that lead to successful performance in virtual screening. These small structural differences significantly impact ligand binding predictions as observed in our virtual screening campaigns. They could also be a source of disappointing results in other virtual screening campaigns on Mpro carried out so far by research groups world-wide.[19,25,30,31,98] These observations indicate that there is space to improve the discriminatory ability of druggability scores by training on a wider range of structures generated by simulation as well as crystallography. Moreover, they highlight the need to develop simulation methods to generate *holo*-like protein structures for virtual screening.

This work was carried out in part within the framework of the EXSCALATE4CORONAVIRUS (E4C)[79] project. We here used 400 000 core-hours on the JURECA supercomputer in the Jülich Supercomputing Centre for the virtual screenings. Our work demonstrates how an advanced computational procedure combined with experimental validation can correctly predict structure and affinity trends of effective hit molecules for a challenging protein target. This combined approach may provide a powerful drug discovery strategy, especially against pandemic viruses and other pathogens, for which the immediate identification of effective treatments is of paramount importance.

EXPERIMENTAL PROCEDURES

Protein Expression, enzymatic activity methodologies and primary screen and dose response are reported in the SI (section S9.1 and S9.2)

Site Analysis of SARS-CoV-2 Main Protease Structure. The SiteMap tool,[99] together with the TRAPP approach[41] were used for the characterization of the proteins' binding sites. TRAPP provides tools for (i) the exploration of binding pocket flexibility and (ii) the estimation of druggability variation in an ensemble of protein structures.[41]

Druggability Calculation. Two druggability indices from SiteMap tool were used: The SiteScore is based on a weighted sum of several properties accounting for the degree of pocket enclosure, size, and the balance between hydrophobic and hydrophilic character in the binding site. The Dscore uses almost the same properties as SiteScore but different coefficients are used and hydrophilicity is not considered.

In the TRAPP tool, the active site pocket of SARS-CoV-2 Mpro was defined by assigning a distance of 3.5 Å around all atoms of the ligand N3 from PDB ID 6LU7. This distance was used to detect residues that potentially may contribute to the binding site and to define dimensions of a 3D grid that was then used to compute the binding pocket shape. Then the binding pocket for each structure was mapped on the 3D grid. The druggability score of this pocket was computed using linear regression (LR) or a convolutional neural network (CNN) and scaled between 0 and 1.[40] Scores were calculated for all conformations generated for 6LU7 and 6Y2G from LRIP and tConcoord, as well as for frames collected every 10 ns from the 100 μs spanning classical MD trajectory generated by starting from the crystal structure with PDB ID6Y84by D. E. Shaw Research.[35] For each structure, a set of the binding site residues that line the binding pocket was detected using the procedure implemented in the TRAPP package. Specifically, each residue was characterized by the number of atoms that contact with the binding pocket.

Structure Selection. We selected all structures generated by tConcoord, LRIP (see SI section S10), or extracted from MD trajectories, within the top 10% of the LR and CNN druggability score. These structures were then clustered by the binding site similarity into eight clusters using k-means procedure (see Figure S8 and also section S4.3 for more details).

Virtual Screening and Library Preparation. Virtual screening studies were performed on a repurposing library including all the commercialized and under development drugs retrieved in the Clarivate Analytics Integrity database, merged with the internal chemical library from Dompè pharma company

https://doi.org/10.1021/acsptsci.0c00215
ACS Pharmacol. Transl. Sci. 2021, 4, 1079−1095

of already proven safe-in-man compounds and the Fraunhofer Institute BROAD Repurposing Library, removing duplicate structures. Known inhibitors of SARS-CoV Mpro were retrieved from several sources including the literature, the Clarivate Analytics Integrity database, the GOSTAR database, and the data repository shared by the Global Health Drug Discovery Institute. Library preparation is reported in the SI section S10. Virtual screening was conducted on the 216 selected receptors by both Fred[100] and Glide[68] docking programs. The protocols are reported in the SI section S10.

Fingerprints Generation. Results are reported in the SI (section S10).

Limitations. As with any modeling study, our models also have limitations. Protein mobility is most probably increased by the presence of moving and displaceable water molecules.[101] Solvation effects can account for up to 100-fold difference in binding affinity (corresponding to ∼3 kcal/mol in binding free energy[102]). Our docking protocols do not consider individual water molecules. Also, we in part rely on scoring functions to rank and select the best binding poses. Current docking/scoring methods[102,103] were suggested to provide reasonable predictions of ligand binding modes, but their performance can be disappointing. Additionally, those methods are often system-dependent, making it very hard to decide which scoring function is suitable for the chosen target protein. To partially overcome these issues, we carefully setup Glide and Fred docking procedures by reproducing a set of covalent and noncovalent SARS-CoV-2 Mpro-ligand crystal poses (Table S7) and by using other criteria, as ROC and EF to establish the quality of our docking procedure.

Moreover, we compare assay-dependent IC_{50} data coming from different laboratories. In addition, several of the known Mpros inhibitors are covalently bound to the protein. Irreversible (covalent) enzyme inhibitors cannot be unambiguously ranked for biochemical potency by using IC_{50} values, because the same IC_{50} value could originate either from relatively low initial binding affinity accompanied by high chemical reactivity, or the other way around.[104] In other words, the important quantity to be considered would be the rate of covalent modification, (k_{inact}/K_i), that describes the efficiency of covalent bond formation resulting from the potency (K_i) of the first reversible binding event and the maximum potential rate (k_{inact}) of inactivation.[105] This information is unfortunately not available for most of the ligands here considered.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsptsci.0c00215.

Supplemental experimental procedures, Figures S1−S22, Table S1−S6, and supplemental computational methods section S10 (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Giulia Rossetti − *Institute for Neuroscience and Medicine (INM-9) and Institute for Advanced Simulations (IAS-5) "Computational biomedicine", Forschungszentrum Jülich, Jülich 52425, Germany; Jülich Supercomputing Center (JSC), Forschungszentrum Jülich, Jülich 52425, Germany; Department of Hematology, Oncology, Hemostaseology, and Stem Cell Transplantation, RWTH Aachen University, Aachen 44517, Germany;* orcid.org/0000-0002-2032-4630; Email: g.rossetti@fz-juelich.de

Authors

Jonas Gossen − *Institute for Neuroscience and Medicine (INM-9) and Institute for Advanced Simulations (IAS-5) "Computational biomedicine", Forschungszentrum Jülich, Jülich 52425, Germany; Faculty of Mathematics, Computer Science and Natural Sciences, RWTH Aachen, Aachen 52062, Germany;* orcid.org/0000-0002-0730-6796

Simone Albani − *Institute for Neuroscience and Medicine (INM-9) and Institute for Advanced Simulations (IAS-5) "Computational biomedicine", Forschungszentrum Jülich, Jülich 52425, Germany; Faculty of Mathematics, Computer Science and Natural Sciences, RWTH Aachen, Aachen 52062, Germany*

Anton Hanke − *Molecular and Cellular Modeling Group, Heidelberg Institute for Theoretical Studies (HITS), Heidelberg 69118, Germany; Institute of Pharmacy and Molecular Biotechnology (IPMB), Heidelberg University, Heidelberg 69120, Germany*

Benjamin P. Joseph − *Institute for Neuroscience and Medicine (INM-9) and Institute for Advanced Simulations (IAS-5) "Computational biomedicine", Forschungszentrum Jülich, Jülich 52425, Germany; Faculty of Mathematics, Computer Science and Natural Sciences, RWTH Aachen, Aachen 52062, Germany*

Cathrine Bergh − *Science for Life Laboratory & Swedish e-Science Research Center, Department of Applied Physics, KTH Royal Institute of Technology, Stockholm 11428, Sweden*

Maria Kuzikov − *Department of Screening Port, Fraunhofer Institute for Translational Medicine and Pharmacology ITMP, Hamburg 22525, Germany;* orcid.org/0000-0001-8771-1865

Elisa Costanzi − *Elettra-Sincrotrone Trieste S.C.p.A., Basovizza, Trieste 34149, Italy*

Candida Manelfi − *Dompé Farmaceutici SpA, L'Aquila 67100, Italy*

Paola Storici − *Elettra-Sincrotrone Trieste S.C.p.A., Basovizza, Trieste 34149, Italy*

Philip Gribbon − *Department of Screening Port, Fraunhofer Institute for Translational Medicine and Pharmacology ITMP, Hamburg 22525, Germany*

Andrea R. Beccari − *Dompé Farmaceutici SpA, L'Aquila 67100, Italy*

Carmine Talarico − *Dompé Farmaceutici SpA, L'Aquila 67100, Italy*

Francesca Spyrakis − *Department of Drug Science and Technology, University of Turin, Turin 10125, Italy;* orcid.org/0000-0002-4016-227X

Erik Lindahl − *Science for Life Laboratory & Swedish e-Science Research Center, Department of Applied Physics, KTH Royal Institute of Technology, Stockholm 11428, Sweden; Science for Life Laboratory, Department of Biochemistry and Biophysics, Stockholm University, Solna SE-106 91, Sweden*

Andrea Zaliani − *Department of Screening Port, Fraunhofer Institute for Translational Medicine and Pharmacology ITMP, Hamburg 22525, Germany;* orcid.org/0000-0002-1740-8390

Paolo Carloni − *Institute for Neuroscience and Medicine (INM-9), Institute for Molecular Neuroscience and Neuroimaging (INM-11), and Institute for Advanced Simulations (IAS-5) "Computational biomedicine", Forschungszentrum Jülich,*

Jülich 52425, Germany; Faculty of Mathematics, Computer Science and Natural Sciences, RWTH Aachen, Aachen 52062, Germany; orcid.org/0000-0002-9010-0149

Rebecca C. Wade − *Molecular and Cellular Modeling Group, Heidelberg Institute for Theoretical Studies (HITS), Heidelberg 69118, Germany; Zentrum für Molekulare Biologie der University Heidelberg, DKFZ-ZMBH Alliance, Heidelberg 69120, Germany; Interdisciplinary Center for Scientific Computing (IWR), Heidelberg University, Heidelberg 69120, Germany;* orcid.org/0000-0001-5951-8670

Francesco Musiani − *Laboratory of Bioinorganic Chemistry, Department of Pharmacy and Biotechnology, University of Bologna, Bologna 40126, Italy;* orcid.org/0000-0003-0200-1712

Daria B. Kokh − *Molecular and Cellular Modeling Group, Heidelberg Institute for Theoretical Studies (HITS), Heidelberg 69118, Germany;* orcid.org/0000-0002-4687-6572

Complete contact information is available at: https://pubs.acs.org/10.1021/acsptsci.0c00215

Author Contributions

#J.G., S.A., and A.H. equally contributed. E.L., A.Z., P.C., R.C.W., F.M., D.B.K., and G.R. shared senior authorship. J.G. and S.A. performed all the docking experiments, the PLIF and pharmacophore calculations, as well as editing most of the figures and tables of the manuscript. A.H. and D.B.K. performed the TRAPP analyses. B.P.J. collected and analyzed the X-ray structures. F.M. and G.R. performed the SiteMap analyses. C.M., C.T., and A.B. took care of the library collection. C.B. and E.L. performed the MSM analyses. M.K., P.G., and A.Z. performed the experiments. E.C. and P.S. solved the crystal structure. F.S., A.Z., P.C., R.C.W., F.M., D.B.K., and G.R. wrote the manuscript and contributed to the design of the research and the analysis of the data. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

Tables S1−S3 in .xlsx format are available at DOI: 10.5281/zenodo.4299967. The data sets (.csv, .xlsx) generated during this study are available at DOI: 10.5281/zenodo.4299967.

■ ACKNOWLEDGMENTS

This research was conducted under the project "EXaSCale smArt pLatform Against paThogEns for Corona Virus−Exscalate4CoV" founded by the EU's H2020-SC1-PHE-CORONAVIRUS-2020 call, Grant No. 101003551. Additional computational resources were provided by the Swedish National Infrastructure for Computing (SNIC) and the Knut and Alice Wallenberg Foundation. D.B.K and R.C.W. acknowledge the support of the Klaus Tschira Foundation. G.R., P.C., D.B.K., and R.C.W. acknowledge the Human Brain Project funded by the European Union's Horizon 2020 Framework Programme for Research and Innovation under the Specific Grant Agreement No. 945539 (Human Brain Project SGA3). G.R., A.Z., P.S., and P.C. acknowledge the E4C consortium. We would also like to thank Dr. Katja Herzog (EU-OPENSCREEN ERIC) for providing access to the EU-OPENSCREEN ERIC Bioactive Compound Library data.[80]

■ REFERENCES

(1) Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., Hu, Y., Tao, Z.-W., Tian, J.-H., Pei, Y.-Y., Yuan, M.-L., Zhang, Y.-L., Dai, F.-H., Liu, Y., Wang, Q.-M., Zheng, J.-J., Xu, L., Holmes, E. C., and Zhang, Y.-Z. (2020) A new coronavirus associated with human respiratory disease in China. *Nature 579* (7798), 265−269.

(2) Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., Huang, C.-L., Chen, H.-D., Chen, J., Luo, Y., Guo, H., Jiang, R.-D., Liu, M.-Q., Chen, Y., Shen, X.-R., Wang, X., Zheng, X.-S., Zhao, K., Chen, Q.-J., Deng, F., Liu, L.-L., Yan, B., Zhan, F.-X., Wang, Y.-Y., Xiao, G.-F., and Shi, Z.-L. (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature 579* (7798), 270−273.

(3) Chen, Y., Liu, Q., and Guo, D. (2020) Emerging coronaviruses: Genome structure, replication, and pathogenesis. *J. Med. Virol. 92* (4), 418−423.

(4) Xue, X., Yu, H., Yang, H., Xue, F., Wu, Z., Shen, W., Li, J., Zhou, Z., Ding, Y., Zhao, Q., Zhang, X. C., Liao, M., Bartlam, M., and Rao, Z. (2008) Structures of Two Coronavirus Main Proteases: Implications for Substrate Binding and Antiviral Drug Design. *J. Virol. 82* (5), 2515−2527.

(5) Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J. R., and Hilgenfeld, R. (2003) Coronavirus Main Proteinase (3CLpro) Structure: Basis for Design of Anti-SARS Drugs. *Science 300* (5626), 1763.

(6) Naqvi, A. A. T., Fatima, K., Mohammad, T., Fatima, U., Singh, I. K., Singh, A., Atif, S. M., Hariprasad, G., Hasan, G. M., and Hassan, M. I. (2020) Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: Structural genomics approach. *Biochim. Biophys. Acta, Mol. Basis Dis. 1866* (10), 165878−165878.

(7) Wang, M., Yan, M., Xu, H., Liang, W., Kan, B., Zheng, B., Chen, H., Zheng, H., Xu, Y., Zhang, E., Wang, H., Ye, J., Li, G., Li, M., Cui, Z., Liu, Y.-F., Guo, R.-T., Liu, X.-N., Zhan, L.-H., Zhou, D.-H., Zhao, A., Hai, R., Yu, D., Guan, Y., and Xu, J. (2005) SARS-CoV infection in a restaurant from palm civet. *Emerging Infect. Dis. 11* (12), 1860−1865.

(8) Haagmans, B. L., and Osterhaus, A. D. M. E. (2006) Nonhuman primate models for SARS. *PLoS Med. 3* (5), e194−e194.

(9) Yang, S., Chen, S.-J., Hsu, M.-F., Wu, J.-D., Tseng, C.-T. K., Liu, Y.-F., Chen, H.-C., Kuo, C.-W., Wu, C.-S., Chang, L.-W., Chen, W.-C., Liao, S.-Y., Chang, T.-Y., Hung, H.-H., Shr, H.-L., Liu, C.-Y., Huang, Y.-A., Chang, L.-Y., Hsu, J.-C., Peters, C. J., Wang, A. H. J., and Hsu, M.-C. (2006) Synthesis, Crystal Structure, Structure−Activity Relationships, and Antiviral Activity of a Potent SARS Coronavirus 3CL Protease Inhibitor. *J. Med. Chem. 49* (16), 4971−4980.

(10) Patick, A. K., and Potts, K. E. (1998) Protease inhibitors as antiviral agents. *Clin. Microbiol. Rev. 11* (4), 614−627.

(11) Zhong, N., Zhang, S., Zou, P., Chen, J., Kang, X., Li, Z., Liang, C., Jin, C., and Xia, B. (2008) Without Its N-Finger, the Main Protease of Severe Acute Respiratory Syndrome Coronavirus Can. Form a Novel Dimer through Its C-Terminal Domain. *J. Virol. 82* (9), 4227−4234.

(12) Paasche, A., Zipper, A., Schäfer, S., Ziebuhr, J., Schirmeister, T., and Engels, B. (2014) Evidence for substrate binding-induced zwitterion formation in the catalytic Cys-His dyad of the SARS-CoV main protease. *Biochemistry 53* (37), 5930−46.

(13) Zhang, L., Lin, D., Kusov, Y., Nian, Y., Ma, Q., Wang, J., von Brunn, A., Leyssen, P., Lanko, K., Neyts, J., de Wilde, A., Snijder, E. J., Liu, H., and Hilgenfeld, R. (2020) α-Ketoamides as Broad-Spectrum Inhibitors of Coronavirus and Enterovirus Replication: Structure-Based Design, Synthesis, and Activity Assessment. *J. Med. Chem. 63* (9), 4562−4578.

(14) Zhang, L., Lin, D., Sun, X., Curth, U., Drosten, C., Sauerhering, L., Becker, S., Rox, K., and Hilgenfeld, R. (2020) Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. *Science 368* (6489), 409−412.

(15) Bzówka, M., Mitusińska, K., Raczyńska, A., Samol, A., Tuszyński, J. A., and Góra, A. (2020) Structural and Evolutionary Analysis Indicate That the SARS-CoV-2 Mpro Is a Challenging Target for Small-Molecule Inhibitor Design. *Int. J. Mol. Sci. 21* (9), 3099.

(16) Joshi, R. S., Jagdale, S. S., Bansode, S. B., Shankar, S. S., Tellis, M. B., Pandya, V. K., Chugh, A., Giri, A. P., and Kulkarni, M. J. (2020) Discovery of potential multi-target-directed ligands by targeting host-specific SARS-CoV-2 structurally conserved main protease. *J. Biomol. Struct. Dyn.*, 1−16.
(17) Ton, A.-T., Gentile, F., Hsing, M., Ban, F., and Cherkasov, A. (2020) Rapid Identification of Potential Inhibitors of SARS-CoV-2 Main Protease by Deep Docking of 1.3 Billion Compounds. *Mol. Inf. 39* (8), 2000028.
(18) André, F., Manuel, S., Santhosh, N., Markus A, L., and Martin, S. (2020) Inhibitors for Novel Coronavirus Protease Identified by Virtual Screening of 687 Million Compounds. *Chemrxiv*; https://chemrxiv.org/articles/preprint/Inhibitors_for_Novel_Coronavirus_Protease_Identified_by_Virtual_Screening_of_687_Million_Compounds/11923239/2, ver. 2, revised 2020-04-15.
(19) Gentile, D., Patamia, V., Scala, A., Sciortino, M. T., Piperno, A., and Rescifina, A. (2020) Putative Inhibitors of SARS-CoV-2 Main Protease from A Library of Marine Natural Products: A Virtual Screening and Molecular Modeling Study. *Mar Drugs 18* (4), 225.
(20) Adem, S., Eyupoglu, V., Sarfraz, I., Rasul, A., and Ali, M. (2020) Identification of Potent COVID-19 Main Protease (Mpro) Inhibitors from Natural Polyphenols: An in Silico Strategy Unveils a Hope against CORONA. *Preprints*, No. 0333.
(21) Xu, Z., Peng, C., Shi, Y., Zhu, Z., Mu, K., Wang, X., and Zhu, W. (2020) Nelfinavir was predicted to be a potential inhibitor of 2019-nCov main protease by an integrative approach combining homology modelling, molecular docking and binding free energy calculation. *BioRxiv*, No. 921627.
(22) Liu, X., and Wang, X.-J. (2020) Potential inhibitors against 2019-nCoV coronavirus M protease from clinically approved medicines. *J. Genet. Genomics 47* (2), 119−121.
(23) Li, Y., Zhang, J., Wang, N., Li, H., Shi, Y., Guo, G., Liu, K., Zeng, H., and Zou, Q. (2020) Therapeutic Drugs Targeting 2019-nCoV Main Protease by High-Throughput Screening. *BioRxiv*, No. 922922.
(24) Nguyen, D. D., Gao, K., Chen, J., Wang, R., and Wei, G.-W. (2020) Potentially highly potent drugs for 2019-nCoV. *Bioxriv*, No. 936013.
(25) Sekhar, T. (2020) Molecular Docking and Virtual Screening based prediction of drugs for COVID-19. *Combinatorial Chemistry & High Throughput Screening 23*, 1−12.
(26) Chen, Y. W., Yiu, C.-P. B., and Wong, K.-Y. (2020) 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*, 129−129.
(27) Kneller, D. W., Phillips, G., O'Neill, H. M., Jedrzejczak, R., Stols, L., Langan, P., Joachimiak, A., Coates, L., and Kovalevsky, A. (2020) Structural plasticity of SARS-CoV-2 3CL Mpro active site cavity revealed by room temperature X-ray crystallography. *Nat. Commun. 11* (1), 3202.
(28) Vuong, W., Khan, M. B., Fischer, C., Arutyunova, E., Lamer, T., Shields, J., Saffran, H. A., McKay, R. T., van Belkum, M. J., Joyce, M. A., Young, H. S., Tyrrell, D. L., Vederas, J. C., and Lemieux, M. J. (2020) Feline coronavirus drug inhibits the main protease of SARS-CoV-2 and blocks virus replication. *Nat. Commun. 11* (1), 4282.
(29) Huynh, T., Wang, H., and Luan, B. (2020) Structure-based lead optimization of herbal medicine rutin for inhibiting SARS-CoV-2's main protease. *Phys. Chem. Chem. Phys. 22* (43), 25335−25343.
(30) Grottesi, A., Bešker, N., Emerson, A., Manelfi, C., Beccari, A. R., Frigerio, F., Lindahl, E., Cerchia, C., and Talarico, C. (2020) Computational Studies of SARS-CoV-2 3CLpro: Insights from MD Simulations. *Int. J. Mol. Sci. 21* (15), 5346.
(31) Gervasoni, S., Vistoli, G., Talarico, C., Manelfi, C., Beccari, A. R., Studer, G., Tauriello, G., Waterhouse, A. M., Schwede, T., and Pedretti, A. (2020) A Comprehensive Mapping of the Druggable Cavities within the SARS-CoV-2 Therapeutically Relevant Proteins by Combining Pocket and Docking Searches as Implemented in Pockets 2.0. *Int. J. Mol. Sci. 21* (14), 5152.
(32) Cannalire, R., Stefanelli, I., Cerchia, C., Beccari, A. R., Pelliccia, S., and Summa, V. (2020) SARS-CoV-2 Entry Inhibitors: Small Molecules and Peptides Targeting Virus or Host Cells. *Int. J. Mol. Sci. 21* (16), 5707.
(33) Nguyen, D. D., Gao, K., Chen, J., Wang, R., and Wei, G.-W. (2020) Unveiling the molecular mechanism of SARS-CoV-2 main protease inhibition from 92 crystal structures. *arXiv:2005.13653 [q-bio.BM]*, ver. 1, https://arxiv.org/abs/2005.13653v1.
(34) Lee, T.-W., Cherney, M. M., Liu, J., James, K. E., Powers, J. C., Eltis, L. D., and James, M. N. G. (2007) Crystal structures reveal an induced-fit binding of a substrate-like Aza-peptide epoxide to SARS coronavirus main peptidase. *J. Mol. Biol. 366* (3), 916−932.
(35) Shaw, D. E., *Molecular Dynamics Simulations Related to SARS-CoV-2*; D. E. Shaw Research, 2020.
(36) Stank, A., Kokh, D. B., Horn, M., Sizikova, E., Neil, R., Panecka, J., Richter, S., and Wade, R. C. (2017) TRAPP webserver: predicting protein binding site flexibility and detecting transient binding pockets. *Nucleic Acids Res. 45* (W1), W325−W330.
(37) Kokh, D. B., Czodrowski, P., Rippmann, F., and Wade, R. C. (2016) Perturbation Approaches for Exploring Protein Binding Site Flexibility to Predict Transient Binding Pockets. *J. Chem. Theory Comput. 12* (8), 4100−4113.
(38) Seeliger, D., and de Groot, B. (2009) tCONCOORD-GUI: Visually supported conformational sampling of bioactive molecules. *J. Comput. Chem. 30*, 1160−6.
(39) Halgren, T. (2007) New Method for Fast and Accurate Binding-site Identification and Analysis. *Chem. Biol. Drug Des. 69* (2), 146−148.
(40) Yuan, J.-H., Han, S. B., Richter, S., Wade, R. C., and Kokh, D. B. (2020) Druggability Assessment in TRAPP Using Machine Learning Approaches. *J. Chem. Inf. Model. 60* (3), 1685−1699.
(41) Kokh, D. B., Richter, S., Henrich, S., Czodrowski, P., Rippmann, F., and Wade, R. C. (2013) TRAPP: a tool for analysis of transient binding pockets in proteins. *J. Chem. Inf. Model. 53* (5), 1235−52.
(42) Halgren, T. A. (2009) Identifying and Characterizing Binding Sites and Assessing Druggability. *J. Chem. Inf. Model. 49* (2), 377−389.
(43) Wu, C.-Y., King, K.-Y., Kuo, C.-J., Fang, J.-M., Wu, Y.-T., Ho, M.-Y., Liao, C.-L., Shie, J.-J., Liang, P.-H., and Wong, C.-H. (2006) Stable Benzotriazole Esters as Mechanism-Based Inactivators of the Severe Acute Respiratory Syndrome 3CL Protease. *Chem. Biol. 13* (3), 261−268.
(44) Thanigaimalai, P., Konno, S., Yamamoto, T., Koiwai, Y., Taguchi, A., Takayama, K., Yakushiji, F., Akaji, K., Kiso, Y., Kawasaki, Y., Chen, S.-E., Naser-Tavakolian, A., Schön, A., Freire, E., and Hayashi, Y. (2013) Design, synthesis, and biological evaluation of novel dipeptide-type SARS-CoV 3CL protease inhibitors: Structure−activity relationship study. *Eur. J. Med. Chem. 65*, 436−447.
(45) Turlington, M., Chun, A., Tomar, S., Eggler, A., Grum-Tokars, V., Jacobs, J., Daniels, J. S., Dawson, E., Saldanha, A., Chase, P., Baez-Santos, Y. M., Lindsley, C. W., Hodder, P., Mesecar, A. D., and Stauffer, S. R. (2013) Discovery of N-(benzo[1,2,3]triazol-1-yl)-N-(benzyl)-acetamido)phenyl) carboxamides as severe acute respiratory syndrome coronavirus (SARS-CoV) 3CLpro inhibitors: Identification of ML300 and noncovalent nanomolar inhibitors with an induced-fit binding. *Bioorg. Med. Chem. Lett. 23* (22), 6172−6177.
(46) Zhang, J., Pettersson, H. I., Huitema, C., Niu, C., Yin, J., James, M. N. G., Eltis, L. D., and Vederas, J. C. (2007) Design, Synthesis, and Evaluation of Inhibitors for Severe Acute Respiratory Syndrome 3C-Like Protease Based on Phthalhydrazide Ketones or Heteroaromatic Esters. *J. Med. Chem. 50* (8), 1850−1864.
(47) Galasiti Kankanamalage, A. C., Kim, Y., Damalanka, V. C., Rathnayake, A. D., Fehr, A. R., Mehzabeen, N., Battaile, K. P., Lovell, S., Lushington, G. H., Perlman, S., Chang, K.-O., and Groutas, W. C. (2018) Structure-guided design of potent and permeable inhibitors of MERS coronavirus 3CL protease that utilize a piperidine moiety as a novel design element. *Eur. J. Med. Chem. 150*, 334−346.
(48) Prior, A. M., Kim, Y., Weerasekara, S., Moroze, M., Alliston, K. R., Uy, R. A. Z., Groutas, W. C., Chang, K.-O., and Hua, D. H. (2013) Design, synthesis, and bioevaluation of viral 3C and 3C-like protease inhibitors. *Bioorg. Med. Chem. Lett. 23* (23), 6317−6320.
(49) Nascimento, J. A. C., Jr., Santos, A. M., Quintans-Júnior, L. J., Walker, C. I. B., Borges, L. P., and Serafini, M. R. (2020) SARS, MERS

and SARS-CoV-2 (COVID-19) treatment: a patent review. *Expert Opinion on Therapeutic Patents*, 1−13.
(50) Shao, Y.-M., Yang, W.-B., Kuo, T.-H., Tsai, K.-C., Lin, C.-H., Yang, A.-S., Liang, P.-H., and Wong, C.-H. (2008) Design, synthesis, and evaluation of trifluoromethyl ketones as inhibitors of SARS-CoV 3CL protease. *Bioorg. Med. Chem. 16* (8), 4652−4660.
(51) Shie, J.-J., Fang, J.-M., Kuo, T.-H., Kuo, C.-J., Liang, P.-H., Huang, H.-J., Wu, Y.-T., Jan, J.-T., Cheng, Y.-S. E., and Wong, C.-H. (2005) Inhibition of the severe acute respiratory syndrome 3CL protease by peptidomimetic α,β-unsaturated esters. *Bioorg. Med. Chem. 13* (17), 5240−5252.
(52) Jacobs, J., Grum-Tokars, V., Zhou, Y., Turlington, M., Saldanha, S. A., Chase, P., Eggler, A., Dawson, E. S., Baez-Santos, Y. M., Tomar, S., Mielech, A. M., Baker, S. C., Lindsley, C. W., Hodder, P., Mesecar, A., and Stauffer, S. R. (2013) Discovery, Synthesis, And Structure-Based Optimization of a Series of N-(tert-Butyl)-2-(N-arylamido)-2-(pyridin-3-yl) Acetamides (ML188) as Potent Noncovalent Small Molecule Inhibitors of the Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) 3CL Pr. *J. Med. Chem. 56* (2), 534−546.
(53) Jain, R. P., Pettersson, H. I., Zhang, J., Aull, K. D., Fortin, P. D., Huitema, C., Eltis, L. D., Parrish, J. C., James, M. N. G., Wishart, D. S., and Vederas, J. C. (2004) Synthesis and Evaluation of Keto-Glutamine Analogues as Potent Inhibitors of Severe Acute Respiratory Syndrome 3CLpro. *J. Med. Chem. 47* (25), 6113−6116.
(54) Kumar, V., Shin, J. S., Shie, J.-J., Ku, K. B., Kim, C., Go, Y. Y., Huang, K.-F., Kim, M., and Liang, P.-H. (2017) Identification and evaluation of potent Middle East respiratory syndrome coronavirus (MERS-CoV) 3CLPro inhibitors. *Antiviral Res. 141*, 101−106.
(55) Wang, X., and Liu, Z. (2014) Prevention and treatment of viral respiratory infections by traditional Chinese herbs. *Chin. Med. J. 127* (7), 1344.
(56) Chuck, C., Ke, Z., Chen, C., Wan, D., Chow, H., and Wong, K. (2014) Profiling of substrate-specificity and rational design of broadspectrum peptidomimetic inhibitors for main proteases of coronaviruses. *Hong Kong Med. J. 20* (Suppl. 4), 22.
(57) Pillaiyar, T., Manickam, M., Namasivayam, V., Hayashi, Y., and Jung, S.-H. (2016) An overview of severe acute respiratory syndrome−coronavirus (SARS-CoV) 3CL protease inhibitors: peptidomimetics and small molecule chemotherapy. *J. Med. Chem. 59* (14), 6595−6628.
(58) Regnier, T., Sarma, D., Hidaka, K., Bacha, U., Freire, E., Hayashi, Y., and Kiso, Y. (2009) New developments for the design, synthesis and biological evaluation of potent SARS-CoV 3CLpro inhibitors. *Bioorg. Med. Chem. Lett. 19* (10), 2722−2727.
(59) Kumar, V., Tan, K.-P., Wang, Y.-M., Lin, S.-W., and Liang, P.-H. (2016) Identification, synthesis and evaluation of SARS-CoV and MERS-CoV 3C-like protease inhibitors. *Bioorg. Med. Chem. 24* (13), 3035−3042.
(60) Ramajayam, R., Tan, K.-P., Liu, H.-G., and Liang, P.-H. (2010) Synthesis, docking studies, and evaluation of pyrimidines as inhibitors of SARS-CoV 3CL protease. *Bioorg. Med. Chem. Lett. 20* (12), 3569−3572.
(61) Wen, C.-C., Kuo, Y.-H., Jan, J.-T., Liang, P.-H., Wang, S.-Y., Liu, H.-G., Lee, C.-K., Chang, S.-T., Kuo, C.-J., Lee, S.-S., Hou, C.-C., Hsiao, P.-W., Chien, S.-C., Shyur, L.-F., and Yang, N.-S. (2007) Specific plant terpenoids and lignoids possess potent antiviral activities against severe acute respiratory syndrome coronavirus. *J. Med. Chem. 50* (17), 4087−4095.
(62) Nguyen, T. T. H., Ryu, H.-J., Lee, S.-H., Hwang, S., Cha, J., Breton, V., and Kim, D. (2011) Discovery of novel inhibitors for human intestinal maltase: virtual screening in a WISDOM environment and in vitro evaluation. *Biotechnol. Lett. 33* (11), 2185−2185.
(63) Mandadapu, S. R., Weerawarna, P. M., Prior, A. M., Uy, R. A. Z., Aravapalli, S., Alliston, K. R., Lushington, G. H., Kim, Y., Hua, D. H., Chang, K.-O., and Groutas, W. C. (2013) Macrocyclic inhibitors of 3C and 3C-like proteases of picornavirus, norovirus, and coronavirus. *Bioorg. Med. Chem. Lett. 23* (13), 3709−3712.
(64) Pakravan, P., Kashanian, S., Khodaei, M. M., and Harding, F. J. (2013) Biochemical and pharmacological characterization of isatin and its derivatives: from structure to activity. *Pharmacol. Rep. 65* (2), 313−335.
(65) Shimamoto, Y., Hattori, Y., Kobayashi, K., Teruya, K., Sanjoh, A., Nakagawa, A., Yamashita, E., and Akaji, K. (2015) Fused-ring structure of decahydroisoquinolin as a novel scaffold for SARS 3CL protease inhibitors. *Bioorg. Med. Chem. 23* (4), 876−890.
(66) McGann, M. (2011) FRED Pose Prediction and Virtual Screening Accuracy. *J. Chem. Inf. Model. 51* (3), 578−596.
(67) Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shelley, M., Perry, J. K., Shaw, D. E., Francis, P., and Shenkin, P. S. (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem. 47* (7), 1739−1749.
(68) Halgren, T. A., Murphy, R. B., Friesner, R. A., Beard, H. S., Frye, L. L., Pollard, W. T., and Banks, J. L. (2004) Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem. 47* (7), 1750−1759.
(69) Jin, Z., Du, X., Xu, Y., Deng, Y., Liu, M., Zhao, Y., Zhang, B., Li, X., Zhang, L., Peng, C., Duan, Y., Yu, J., Wang, L., Yang, K., Liu, F., Jiang, R., Yang, X., You, T., Liu, X., Yang, X., Bai, F., Liu, H., Liu, X., Guddat, L. W., Xu, W., Xiao, G., Qin, C., Shi, Z., Jiang, H., Rao, Z., and Yang, H. (2020) Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature 582* (7811), 289−293.
(70) Ma, C., Sacco, M. D., Hurst, B., Townsend, J. A., Hu, Y., Szeto, T., Zhang, X., Tarbet, B., Marty, M. T., Chen, Y., and Wang, J. (2020) Boceprevir, GC-376, and calpain inhibitors II, XII inhibit SARS-CoV-2 viral replication by targeting the viral main protease. *bioRxiv*, No. 0356-z.
(71) Dai, W., Zhang, B., Jiang, X.-M., Su, H., Li, J., Zhao, Y., Xie, X., Jin, Z., Peng, J., Liu, F., Li, C., Li, Y., Bai, F., Wang, H., Cheng, X., Cen, X., Hu, S., Yang, X., Wang, J., Liu, X., Xiao, G., Jiang, H., Rao, Z., Zhang, L.-K., Xu, Y., Yang, H., and Liu, H. (2020) Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease. *Science 368* (6497), 1331.
(72) Fan, S., Xiao, D., Wang, Y., Liu, L., Zhou, X., and Zhong, W. (2020) Research progress on repositioning drugs and specific therapeutic drugs for SARS-CoV-2. *Future Med. Chem. 12* (17), 1565−1578.
(73) Li, Z., Li, X., Huang, Y.-Y., Wu, Y., Liu, R., Zhou, L., Lin, Y., Wu, D., Zhang, L., Liu, H., Xu, X., Yu, K., Zhang, Y., Cui, J., Zhan, C.-G., Wang, X., and Luo, H.-B. (2020) Identify potent SARS-CoV-2 main protease inhibitors via accelerated free energy perturbation-based virtual screening of existing drugs. *bioRxiv*, No. 004580.
(74) Sacco, M. D., Ma, C., Lagarias, P., Gao, A., Townsend, J. A., Meng, X., Dube, P., Zhang, X., Hu, Y., Kitamura, N., Hurst, B., Tarbet, B., Marty, M. T., Kolocouris, A., Xiang, Y., Chen, Y., and Wang, J. (2020) Structure and inhibition of the SARS-CoV-2 main protease reveals strategy for developing dual inhibitors against Mpro; and cathepsin L. *Science Advances 6*, eabe0751.
(75) Coelho, C., Gallo, G., Campos, C. B., Hardy, L., and Würtele, M. (2020) Biochemical screening for SARS-CoV-2 main protease inhibitors. *PLoS One 15* (10), e0240079.
(76) Dice, L. R. (1945) Measures of the Amount of Ecologic Association Between Species. *Ecology 26* (3), 297−302.
(77) Huynh, T., Wang, H., and Luan, B. (2020) In Silico Exploration of the Molecular Mechanism of Clinically Oriented Drugs for Possibly Inhibiting SARS-CoV-2's Main Protease. *J. Phys. Chem. Lett. 11* (11), 4413−4420.
(78) Liu, P., Liu, H., Sun, Q., Liang, H., Li, C., Deng, X., Liu, Y., and Lai, L. (2020) Potent inhibitors of SARS-CoV-2 3C-like protease derived from N-substituted isatin compounds. *Eur. J. Med. Chem. 206*, 112702.
(79) Exscalate for COVID. https://www.exscalate4cov.eu (accessed 2020-11-20).
(80) EU-OPENSCREEN. https://www.probes-drugs.org/compounds/standardized#compoundset=353@AND (accessed 2020-11-20).
(81) PROBE MINER repository. https://www.probes-drugs.org/compoundsets (accessed 2020-11-20).

(82) Goldstein, M., Gopinathan, G., Neophytides, A., Hiesiger, E., Walker, R., and Nelson, J. (1984) Combined use of benserazide and carbidopa in Parkinson's disease. *Neurology 34* (2), 227−227.
(83) Liang, J., and Olsen, R. W. (2014) Alcohol use disorders and current pharmacological therapies: the role of GABA A receptors. *Acta Pharmacol. Sin. 35* (8), 981−993.
(84) Watson, R. R., Preedy, V. R., and Zibadi, S. (2013) *Polyphenols in human health and disease*, Academic press.
(85) Abian, O., Ortega-Alarcon, D., Jimenez-Alesanco, A., Ceballos-Laita, L., Vega, S., Reyburn, H. T., Rizzuti, B., and Velazquez-Campoy, A. (2020) Structural stability of SARS-CoV-2 3CLpro and identification of quercetin as an inhibitor by experimental screening. *Int. J. Biol. Macromol. 164*, 1693−1703.
(86) Ghahremanpour, M. M., Tirado-Rives, J., Deshmukh, M., Ippolito, J. A., Zhang, C.-H., Cabeza de Vaca, I., Liosi, M.-E., Anderson, K. S., and Jorgensen, W. L. (2020) Identification of 14 Known Drugs as Inhibitors of the Main Protease of SARS-CoV-2. *ACS Med. Chem. Lett. 11*, 2526.
(87) Kokh, D. B., Wade, R. C., and Wenzel, W. (2011) Receptor flexibility in small-molecule docking calculations. *Wiley Interdiscip. Rev.: Comput. Mol. Sci. 1* (2), 298−314.
(88) Amaro, R. E., Baudry, J., Chodera, J., Demir, Ö., McCammon, J. A., Miao, Y., and Smith, J. C. (2018) Ensemble Docking in Drug Discovery. *Biophys. J. 114* (10), 2271−2278.
(89) Evangelista Falcon, W., Ellingson, S. R., Smith, J. C., and Baudry, J. (2019) Ensemble Docking in Drug Discovery: How Many Protein Configurations from Molecular Dynamics Simulations are Needed To Reproduce Known Ligand Binding? *J. Phys. Chem. B 123* (25), 5189−5195.
(90) Walters, W. P., and Wang, R. (2020) New Trends in Virtual Screening. *J. Chem. Inf. Model. 60* (9), 4109−4111.
(91) Cleves, A. E., and Jain, A. N. (2020) Structure- and Ligand-Based Virtual Screening on DUD-E+: Performance Dependence on Approximations to the Binding Pocket. *J. Chem. Inf. Model. 60* (9), 4296−4310.
(92) Jiang, S., Feher, M., Williams, C., Cole, B., and Shaw, D. E. (2020) AutoPH4: An Automated Method for Generating Pharmacophore Models from Protein Binding Pockets. *J. Chem. Inf. Model. 60* (9), 4326−4338.
(93) Tan, J., Verschueren, K. H., Anand, K., Shen, J., Yang, M., Xu, Y., Rao, Z., Bigalke, J., Heisen, B., Mesters, J. R., Chen, K., Shen, X., Jiang, H., and Hilgenfeld, R. (2005) pH-dependent conformational flexibility of the SARS-CoV main proteinase (M(pro)) dimer: molecular dynamics simulations and multiple X-ray structure analyses. *J. Mol. Biol. 354* (1), 25−40.
(94) Cheng, S.-C., Chang, G.-G., and Chou, C.-Y. (2010) Mutation of Glu-166 blocks the substrate-induced dimerization of SARS coronavirus main protease. *Biophys. J. 98* (7), 1327−1336.
(95) Goyal, B., and Goyal, D. (2020) Targeting the Dimerization of the Main Protease of Coronaviruses: A Potential Broad-Spectrum Therapeutic Strategy. *ACS Comb. Sci. 22* (6), 297−305.
(96) Bacha, U., Barrila, J., Velazquez-Campoy, A., Leavitt, S. A., and Freire, E. (2004) Identification of novel inhibitors of the SARS coronavirus main protease 3CLpro. *Biochemistry 43* (17), 4906−12.
(97) Kuzikov, M., Costanzi, E., Reinshagen, J., Esposito, F., Vangeel, L., Wolf, M., Ellinger, B., Claussen, C., Geisslinger, G., Corona, A., Iaconis, D., Talarico, C., Manelfi, C., Cannalire, R., Rossetti, G., Gossen, J., Albani, S., Musiani, F., Herzog, K., Ye, Y., Giabbai, B., Demitri, N., Jochmans, D., De Jonghe, S., Rymenants, J., Summa, V., Tramontano, E., Beccari, A. R., Leyssen, P., Storici, P., Neyts, J., Gribbon, P., and Zaliani, A. (2020) Identification of inhibitors of SARS-CoV-2 3CL-Pro enzymatic activity using a small molecule in-vitro repurposing screen. *BioRxiv*, No. 422677.
(98) Vatansever, E. C., Yang, K., Kratch, K. C., Drelich, A., Cho, C.-C., Mellot, D. M., Xu, S., Tseng, C.-T. K., and Liu, W. R. (2020) Targeting the SARS-CoV-2 Main Protease to Repurpose Drugs for COVID-19. *BioRxiv*, No. 112235.
(99) Schmidtke, P., Souaille, C., Estienne, F., Baurin, N., and Kroemer, R. T. (2010) Large-Scale Comparison of Four Binding Site Detection Algorithms. *J. Chem. Inf. Model. 50* (12), 2191−2200.
(100) McGann, M. (2012) FRED and HYBRID docking performance on standardized datasets. *J. Comput.-Aided Mol. Des. 26* (8), 897−906.
(101) Spyrakis, F., and Cavasotto, C. N. (2015) Open challenges in structure-based virtual screening: Receptor modeling, target flexibility consideration and active site water molecules description. *Arch. Biochem. Biophys. 583*, 105−119.
(102) Wong, S. E., and Lightstone, F. C. (2011) Accounting for water molecules in drug design. *Expert Opin. Drug Discovery 6* (1), 65−74.
(103) Muegge, I., and Rarey, M. (2001) Small molecule docking and scoring, in *Reviews in Computational Chemistry* (Lipkowitz, K. B., and Boyd, D. B., Eds.) Vol. *18*, Chapter 1, Wiley-VCH, New York.
(104) Kuzmič, P. (2020) A two-point IC_{50} method for evaluating the biochemical potency of irreversible enzyme inhibitors. *BioRxiv*, No. 171207.
(105) Strelow, J. M. (2017) A Perspective on the Kinetics of Covalent and Irreversible Inhibition. *SLAS DISCOVERY: Advancing the Science of Drug Discovery 22* (1), 3−20.

https://doi.org/10.1021/acsptsci.0c00215
ACS Pharmacol. Transl. Sci. 2021, 4, 1079−1095