



Article

Computational Selectivity Assessment of Protease Inhibitors against SARS-CoV-2

André Fischer ^{1,†} , Manuel Sellner ^{1,†} , Karolina Mitusińska ^{2,†} , Maria Bzówka ^{2,†} , Markus A. Lill ^{1,*} , Artur Góra ^{2,*} and Martin Smieško ^{1,*}

¹ Computational Pharmacy, Department of Pharmaceutical Sciences, University of Basel, 4056 Basel, Switzerland; and.fischer@unibas.ch (A.F.); manuel.sellner@unibas.ch (M.S.)

² Tunneling Group, Biotechnology Centre, ul. Krzywoustego 8, Silesian University of Technology, 44-100 Gliwice, Poland; k.mitusinska@tunnelinggroup.pl (K.M.); m.bzowka@tunnelinggroup.pl (M.B.)

* Correspondence: markus.lill@unibas.ch (M.A.L.); a.gora@tunnelinggroup.pl (A.G.); martin.smiesko@unibas.ch (M.S.)

† These authors contributed equally to this work.

Abstract: The pandemic of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) poses a serious global health threat. Since no specific therapeutics are available, researchers around the world screened compounds to inhibit various molecular targets of SARS-CoV-2 including its main protease (M^{Pro}) essential for viral replication. Due to the high urgency of these discovery efforts, off-target binding, which is one of the major reasons for drug-induced toxicity and safety-related drug attrition, was neglected. Here, we used molecular docking, toxicity profiling, and multiple molecular dynamics (MD) protocols to assess the selectivity of 33 reported non-covalent inhibitors of SARS-CoV-2 M^{Pro} against eight proteases and 16 anti-targets. The panel of proteases included SARS-CoV M^{Pro}, cathepsin G, caspase-3, ubiquitin carboxy-terminal hydrolase L1 (UCHL1), thrombin, factor Xa, chymase, and prostaticin. Several of the assessed compounds presented considerable off-target binding towards the panel of proteases, as well as the selected anti-targets. Our results further suggest a high risk of off-target binding to chymase and cathepsin G. Thus, in future discovery projects, experimental selectivity assessment should be directed toward these proteases. A systematic selectivity assessment of SARS-CoV-2 M^{Pro} inhibitors, as we report it, was not previously conducted.

Keywords: coronavirus; SARS; protease; selectivity; structure-based design



Citation: Fischer, A.; Sellner, M.; Mitusińska, K.; Bzówka, M.; Lill, M.A.; Góra, A.; Smieško, M. Computational Selectivity Assessment of Protease Inhibitors against SARS-CoV-2. *Int. J. Mol. Sci.* **2021**, *22*, 2065. <https://doi.org/10.3390/ijms22042065>

Academic Editors: Cristina Belizna, Jan Willem Cohen Tervaert, Yehuda Shoenfeld and Alexander Makatsariya

Received: 4 January 2021

Accepted: 11 February 2021

Published: 19 February 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In late 2019, a novel coronavirus termed SARS-CoV-2 emerged and spread around the world causing coronavirus disease 2019 (COVID-19). Until today, over 62 million cases were reported accounting for over a 1.46 million of fatalities (as of 1 December 2020) [1]. While pharmaceutical interventions primarily remained symptomatic, multiple clinical trials are investigating novel treatments, mainly based on drug repurposing [2,3]. Thus, the treatment of this infection with specific drugs constitutes an urgent and unmet medical need. In the pharmaceutical treatment of viral infections such as human immunodeficiency virus and hepatitis C virus, the inhibition of viral proteases is a successfully applied strategy. Consequently, many computational and experimental efforts were directed toward targeting the main protease (M^{Pro}) of SARS-CoV-2 with small molecules leading to the discovery of multiple promising candidates [4–9]. Due to the high urgency and the competitive scientific field, off-target binding was rarely considered in the latest discovery projects. However, a large share of drug attrition in clinical trials, especially regarding compound safety, can be traced back to low target specificity and off-target binding [10–12]. To avoid the large cost associated with late stage drug failure, early off-target profiling, especially with comparably economical computational methods, offers an attractive strategy [12–14]. On the other hand, binding to multiple targets can be beneficial in specific

cases, such as pan inhibition of the viral proteases of SARS-CoV-2 and SARS-CoV [15]. Similarly, the concurrent inhibition of the coagulation protein factor Xa and SARS-CoV-2 M^{Pro} constitutes another example for a potentially synergistic effect of multi-target binding as COVID-19 infection is associated with life-threatening coagulopathies that can be treated with anticoagulants [16,17]. Computational methods such as molecular docking and specialized molecular dynamics (MD) protocols can be exploited to explore the selectivity of small-molecule compounds, as it was evidenced for various targets. For example, we previously applied docking combined with cosolvent MD simulations and determination of hydration hot-spots to investigate the selectivity of allosteric inhibitors against eight nuclear receptors to support targeted experimental profiling of novel compounds [18]. In general, it was discussed that cosolvent MD simulations can provide valuable insights for the development of potent and selective compounds [19,20]. Although multiple studies focused on the selectivity assessment among kinases using computational methods [21,22], there were only minor efforts to establish selectivity factors for small-molecules that bind to proteases [23,24].

Here, we examined the selectivity of 33 experimentally confirmed non-covalent SARS-CoV-2 M^{Pro} inhibitors against eight different proteases including SARS-CoV M^{Pro}, factor Xa, cathepsin G, caspase-3, prostatic, thrombin, ubiquitin carboxy-terminal hydrolase L1 (UCHL1), and chymase by molecular docking (Table 1). The proteases were selected based on a structural similarity search in the NCBI database, as well as considerations regarding their pharmacological relevance. First, we compared the active sites regarding pharmacophores and electrostatic potential. Furthermore, we performed classical as well as cosolvent MD simulations to identify water and small-molecule hot-spots of the respective active sites offering explanations for compound selectivity. Based on our results, experimental selectivity profiling in future discovery projects can be directed toward targets with an inherent high liability for off-target binding. Up to this date, such a comprehensive evaluation of off-target binding of SARS-CoV-2 M^{Pro} was not previously conducted and is of high importance to support antiviral drug development.

Table 1. Proteins considered in this study.

Protein	Function	Anti-Target ^a	Consequence of Inhibition
SARS-CoV-2 M ^{Pro}	Viral replication	-	antiviral activity
SARS-CoV M ^{Pro}	Viral replication	no	antiviral activity
Caspase-3	Apoptosis	yes	interference with development
Factor Xa	Coagulation	no	prevention of coagulopathies
Cathepsin G	Immune system	yes	interference with immune response
UCHL1	Protein degradation	yes	interference with development and homeostasis
Prostatic	Sodium balance	yes	alters homeostasis
Thrombin	Coagulation	no	prevention of coagulopathies
Chymase	Vasoconstriction	yes	interference with blood pressure

^a Description if the protein is regarded as anti-target.

2. Results and Discussion

2.1. Sequence and Active Site Comparison

We selected eight proteases to assess off-target binding and to structurally compare them to SARS-CoV-2 M^{Pro}. The proteases were selected based on their catalytic residues, sequence and structural similarity, availability of structural information, as well as their pharmacological and physiological relevance. Except for SARS-CoV M^{Pro}, they exhibited different overall folds and showed low global sequence similarity to SARS-CoV-2 M^{Pro} (Table 2). Furthermore, the volumes of the active site cavity (Table S1) of the SARS-CoV-2 M^{Pro} was the smallest among all analyzed proteases, and regarding absolute values, most similar to chymase as opposed to SARS-CoV M^{Pro}. However, when we compared the active sites of all analyzed proteases, they presented a remarkable degree of similarity. The root

mean-square deviation (RMSD) of their catalytic residues did not exceed 2 Å, except for UCHL1 with 2.8 Å. Furthermore, when using FuzCav [25] to determine the similarity of the active site pockets of the panel of proteases, we observed that not only their catalytic residues, but also the complete active sites are similar relative to SARS-CoV-2 M^{Pro} (Supplementary Figure S1). Active sites exceeding a similarity value of 0.16 can be regarded as similar [25]. We also compared the electrostatic potentials of the binding cavities of all analyzed proteases using PIPSA [26] software. The electrostatic potentials were compared quantitatively by calculating the Hodgkin similarity index (Table 2, Figure S2). Using the Hodgkin similarity index it is possible to determine the correlation between the potentials of the analysed proteases (+1 indicates that potentials are identical, 0 indicates that potentials are fully uncorrelated, and -1 indicates that potentials are anti-correlated). Three out of eight proteases (SARS-CoV M^{Pro}, prostaticin, and thrombin) presented a positive correlation, whereas the remaining ones indicated an anti-correlation in relation to the SARS-CoV-2 M^{Pro} binding cavity. In the case of UCHL1, the high anti-correlation was caused by a specific binding site spot with reversed distribution of electrostatic potentials. Prostaticin, thrombin and SARS-CoV M^{Pro} also showed high similarity relative to the active site of SARS-CoV-2 M^{Pro} using FuzCav. Thus, even though the substrate specificity of the panel of proteases is diverse [27], their exceptionally similar active sites compared to SARS-CoV-2 may suggest a potential for off-target binding of small-molecules.

Table 2. Similarity of proteins and active sites.

Protein	Catalytic Residues	Global Identity ^a	AS RMSD (Å) ^b	Site Similarity ^c	Fold	Similarity Index ^d
SARS-CoV M ^{Pro}	H41, C145	96.1%	0.2	0.91	α/β	0.90
Caspase-3	H121, C163	11.6%	1.9	0.67	α/β	-0.74
Factor Xa	H57, D102, S195	11.6%	1.8	0.71	all- β	-0.67
Cathepsin G	H59, D103, S196	14.5%	1.8	0.61	all- β	-0.71
UCHL1	C90, H161, D176	15.7%	2.8	0.74	α/β	-0.98
Prostaticin	H85, D134, S154	13.1%	1.6	0.69	all- β	0.26
Thrombin	H57, D102, S195	12.5%	1.8	0.74	all- β	0.31
Chymase	H45, D89, S182	19.0%	1.9	0.64	all- β	-0.49

^a Sequence identity of the catalytic unit to SARS-CoV-2 M^{Pro}; ^b RMSD between histidine, cysteine/serine, and aspartic acid residues relative to SARS-CoV-2 M^{Pro}; ^c Similarity of the binding sites to SARS-CoV-2 M^{Pro} determined by FuzCav; ^d Hodgkin similarity index of the binding sites comparison relative to the SARS-CoV-2 M^{Pro} determined by PIPSA.

2.2. Hydration and Small-Molecule Hot-Spots

In the next step, we further characterized and compared the selected proteases according to their hydration and small-molecule hot-spots by using different molecular probes including water, acetonitrile, isopropanol, and pyridine. The use of specific functional groups represented by the different organic probes associating with the active sites can be used to fine-tune the selectivity profile of protease inhibitors [18,28]. Similarly, selectively targeting hydration sites occurring in one protein, but not in an anti-target, offers potential to be exploited in structure-based design. It should however be mentioned that whether the displacement of water molecules is favorable or not depends on the thermodynamic profile of the respective hydration site [29], which was not assessed in this work. The comparison of the hot-spots in the vicinity of the active site residues revealed distinct similarities (Figure 1).

For SARS-CoV-2 M^{Pro}, we identified two hydration sites located in the vicinity of H41, as well as a small-molecule hot-spot for acetonitrile molecules at the same location (Figure 2). While we could not detect a hydration site in the vicinity of C141, association of pyridine and isopropanol was detected. In the case of SARS-CoV M^{Pro}, however, four hydration sites could be identified. Potentially, the increased flexibility of SARS-CoV M^{Pro} allowed for increased solvent accessibility in comparison to SARS-CoV-2 M^{Pro} (Table S1) [30]. Furthermore, the increased magnitude of cosolvent densities in SARS-CoV M^{Pro} confirmed this observation, although they mainly occupied the vicinity of H41. Even though one hydration site between the M^{Pro}s overlapped, there were signifi-

cant differences between the small-molecule hot-spots of the SARS-CoV proteases, which have to be accounted for in the design of pan inhibitors against the two coronaviruses (Figure 2). For caspase-3, two unique hydration sites were identified in the active site cavity distant from the catalytic residues, which were not observed in the SARS-CoV-2 M^{Pro}. One of the hydration sites overlapped with the occupancy of multiple organic probes, which sampled a unique region not observed in any other protease besides UCHL1. Three hydration sites were identified in the vicinity of the active site residues in UCHL1. One of the hydration sites was located between three catalytic residues as in caspase-3, but none of the other cysteine proteases. Thus, these proteases are similar, while presenting distinct differences to the M^{Pro}s despite their shared catalytic mechanism using cysteine for the nucleophilic attack of the substrate (Table 2). In the active site of factor Xa, we detected two hydration sites matching the position in prostatic and chymase indicating that they are conserved. As one of the sites (denoted as site B in Figure 1) could not be observed in the SARS-CoV-2 M^{Pro}, it might contribute to ligand specificity. Remarkably, the cosolvent densities among factor Xa, cathepsin G, thrombin, and chymase presented a high degree of similarity, with an additional density for acetonitrile compared to SARS-CoV-2 M^{Pro}. The same region was occupied by pyridine probes. A common density of pyridine in the center of the sites, however, suggested a common preference for hydrophobic or aromatic moieties among the aforementioned enzymes. Comparing the organic probe density of factor Xa to SARS-CoV-2 M^{Pro}, a common preference for acetonitrile on the distal side of the catalytic histidine could be observed. This could guide the placement of an amphipathic moiety in this region to inhibit both proteases with future antivirals. While cathepsin G and thrombin shared one of the most commonly observed hydration sites, they lack the common site observed in the viral M^{Pro}s which indicates that the displacement of this water molecule (denoted as site A in Figure 1) could contribute to selective binding. Both thrombin and chymase presented a high number of hydration sites within their active sites sharing the above-mentioned hydration site B not observed in the viral proteases, as well as the previously discussed acetonitrile density near the backbone of the catalytic histidine residue. Thus, the inherent potential for off-target binding of novel antivirals to these targets is small, similar to caspase-3 and UCHL1. Compared to the volume of its active site cavity, the site of chymase seemed highly accessible to the surrounding solvent. Overall, the highest similarity among the proteases regarding hydration sites and small-molecule binding hot-spots, could be observed for factor Xa and chymase. Differences in hydration site locations identified for individual simulations are most probably related to conformational changes of the proteins, as the volumes of the active sites underlined (Table S1).

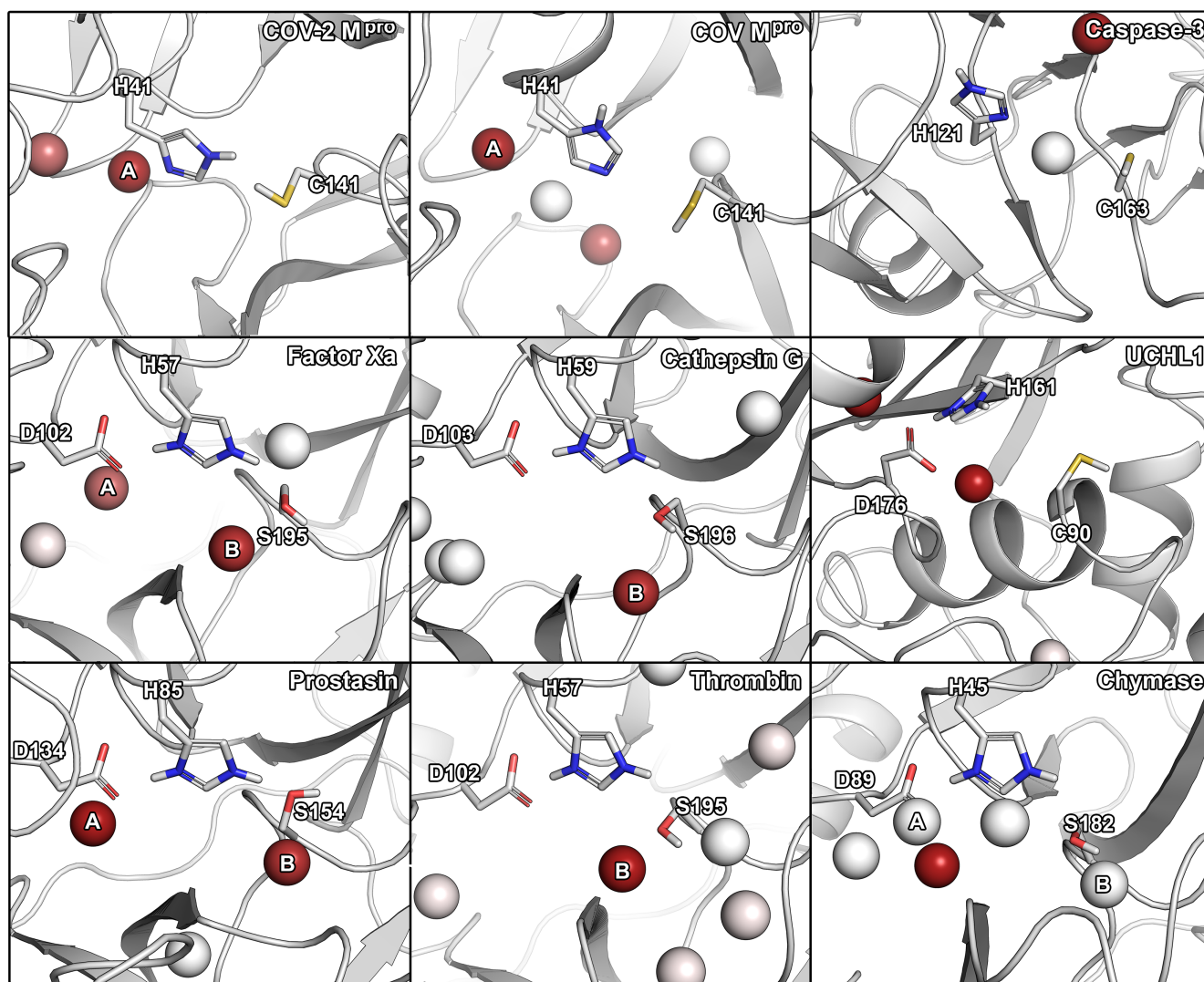


Figure 1. Hydration hot-spots of the selected panel of proteases in relation to the catalytic residues. To allow a direct comparison, the protein structures were aligned according to their catalytic residues. Two most consistently occurring hydration sites are indicated by A and B. The hot-spots are color-coded according to the occupancy of a particular region by identified hydration hot-spots. Hydration hot-spots with highest occupancy are colored in dark red, those with low occupancy in white.

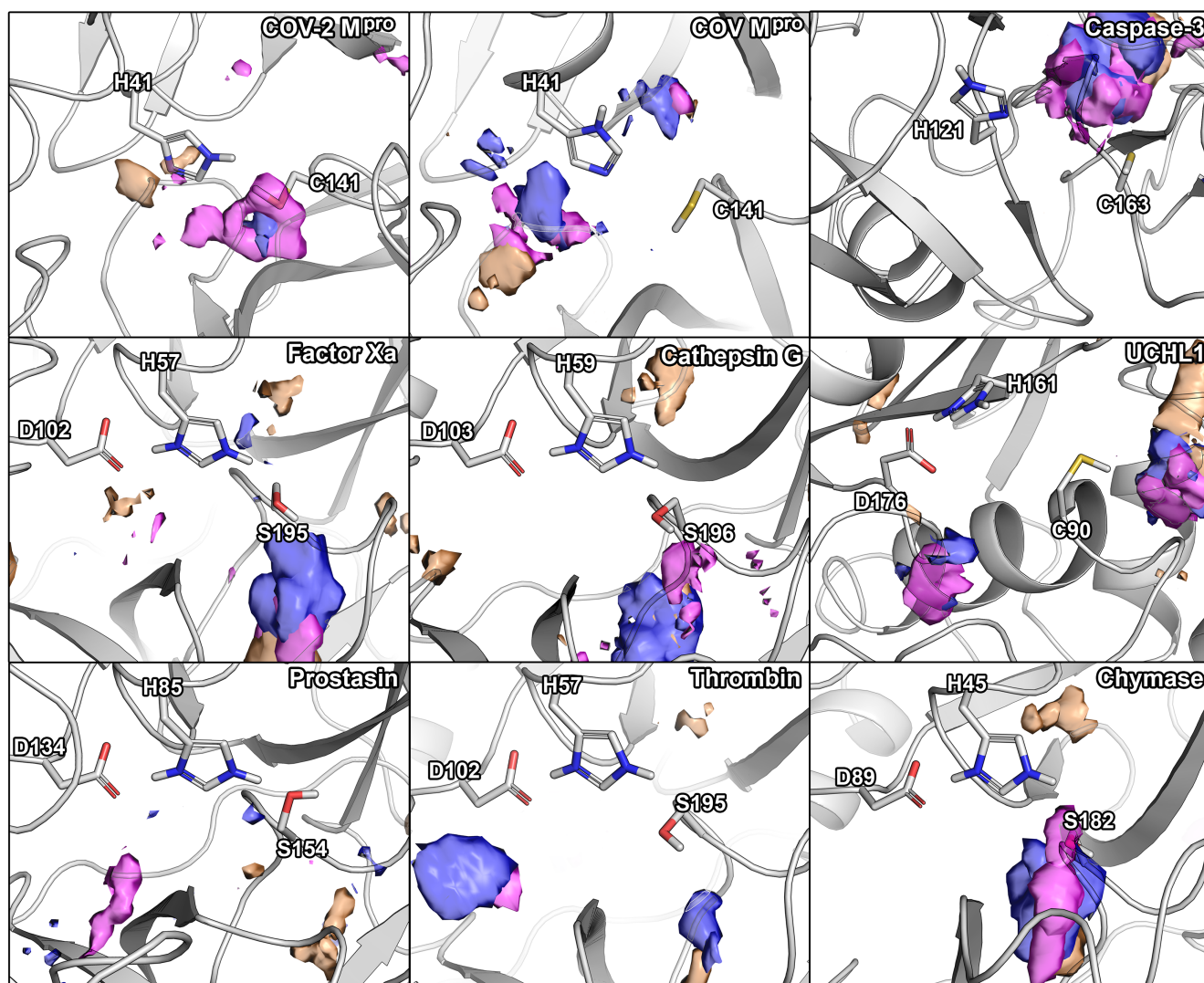


Figure 2. Small-molecule hot-spots of the selected panel of proteases in relation to their catalytic residues. Blue densities correspond to isopropanol, pink densities to pyridine, and orange densities to acetonitrile.

2.3. Protease Selectivity Assessed by Molecular Docking

Binding modes obtained from molecular docking have been widely used to establish selectivity factors toward different targets [13,18,31]. Here, we compiled compound sets comprising experimentally verified ligands of nine proteases to assess the selectivity of recently reported SARS-CoV-2 M^{Pro} inhibitors (Figure 3) by cross-docking them into the respective protein active sites. These known binders were either retrieved from a set of cocrystallized ligands, the PubChem BioAssay database [32], or the literature (Tables S2–S10, Figures S3–S18). First, to ensure accurate pose-prediction of our computational models, we retrieved numerous crystal structures from the Protein Data Bank for each target and cross-docked their cocrystallized ligands, which is a common procedure in virtual screening projects [33]. Based on the obtained RMSD values between predicted and native binding poses, ensembles of protein structures that yielded best-possible pose prediction quality for non-covalent ligands were identified. In the case of unsatisfactory performance of these structures, short MD simulations were performed to enrich the proteins' structural diversity. These procedures were performed for the Glide standard precision (SP) and smina docking protocol, to address known differences among docking programs. This resulted in excellent docking accuracy for most protein systems which ranged from 75% to 100% of cocrystallized ligands being predicted below an RMSD threshold of 2.5 Å.

ligands of the respective enzyme. Again, we determined the ROC AUC metrics with the SARS-CoV-2 compounds regarded as decoys. Except for cathepsin G and chymase, the docking calculations with SARS-CoV-2 M^{Pro} inhibitors as decoys displayed higher ROC AUC values compared to the docking calculations with randomly selected decoy molecules. This indicates an overall low potential for off-target binding based on this metric. However, in addition to the ROC AUC metric, we depicted the scores in histograms for every target (Figure 4A). The docking scores of the SARS-CoV-2 compounds were predicted to be comparably high in magnitude with the majority of compounds only scored slightly better than -6.0 kcal/mol, even when docked to SARS-CoV-2 M^{Pro} itself. This is not surprising as the experimentally measured affinity for those compounds only reached micromolar IC₅₀ values. As already established by the ROC AUC values, cathepsin G presented the highest score overlap between the compound sets suggesting a risk for off-target inhibition of this protease involved in antigen processing. Other proteins with a comparatively high overlap were SARS-CoV M^{Pro}, UCHL1, and chymase. Further, some inhibitors of caspase-3 presented an overlap with the best-scoring SARS-CoV-2 M^{Pro} inhibitors, while the majority of compounds were highly separated. In SARS-CoV M^{Pro}, thrombin, and chymase several SARS-CoV-2 M^{Pro} inhibitors yielded a similar docking score as some of the actives for that target, even though the peaks of the score distribution were well separated. Especially, in the case of thrombin, concurrent binding could benefit COVID-19 patients suffering from coagulopathies such as venous thromboembolism or sepsis-induced coagulopathy [16,17]. In addition to caspase-3, the distribution in factor Xa and prostatic presented a clear separation of the peaks for each compound category indicating a low potential for concurrent binding. In order to obtain more confidence in the results from molecular docking, we used the complexes and subjected them to MD simulations followed by molecular mechanics-generalized Born surface area (MM/GBSA) post-processing. This methodology is considered to be more precise as opposed to docking scores for a multitude of biomolecular systems [34]. Even though the spread of the values was higher using this protocol, the general trends remained highly similar, especially for chymase, cathepsin G, and caspase-3 (Figure S19). There was a slightly higher overlap of the scores for factor Xa and SARS-CoV-M^{Pro}, which would indicate a higher potential for a compound to hit both targets.

Interestingly, when the actives of each target were docked to SARS-CoV-2 M^{Pro}, prostatic inhibitors yielded better docking scores compared to the native ligands (Figure S20). We noticed similar, but less pronounced trends for inhibitors of thrombin, factor Xa, chymase, and caspase-3, while cathepsin G, SARS-CoV M^{Pro}, and UCHL1 compounds presented nearly identical maxima of their docking scores compared to SARS-CoV-2 M^{Pro} inhibitors. In conclusion, the distribution of the scores indicate promiscuity toward chymase, UCHL1, and especially cathepsin G, even though the majority of SARS-CoV-2 M^{Pro} inhibitors presented inferior binding scores towards all assessed proteins. Notably, empirical scoring functions, as they were used in this project, have a known degree of inaccuracy, and thus, the absolute numbers should be regarded only as trends.

To acquire structural insights into the selectivity factors of each protease, we visualized binding modes of ligands presenting either low or high binding affinities for each target. According to these complexes, compounds intended to inhibit SARS-CoV-2 M^{Pro} should present π - π stacking with the catalytic residue H41 as well as high complementarity with the available subpockets of the active site (Figure 4B). To achieve potent and selective interaction with cathepsin G, the binding modes suggest a salt bridge to K192 or H57 as well as a deeply buried hydrophobic moiety to be optimal (Figure 4C). Similarly, ionic interactions, especially if they were buried, seemed to play a role for selective binding toward thrombin (Figure 4D), caspase-3 (Figure 5A), and chymase (Figure 5B). Interestingly, compounds hitting the presumably desired off-target factor Xa also strongly relied on shape complementarity as for SARS-CoV-2 M^{Pro} (Figure 5C), which might explain the concurrent binding to these targets, as we have previously detected in a virtual screening project [4].

Two-dimensional (2D) depictions of all discussed binding modes are presented in the Supplementary Information (Figures S21 and S22).

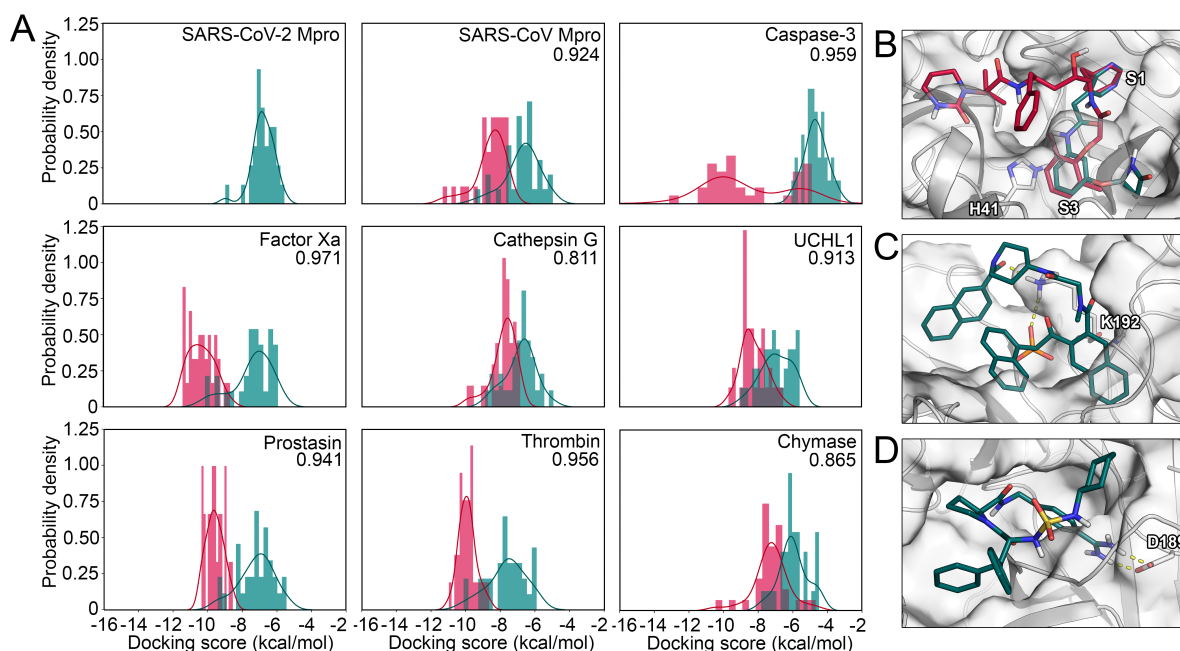


Figure 4. (A) Score distribution of SARS-CoV-2 M^{pro} inhibitors docked to the selected panel of proteases. The compounds designed against SARS-CoV-2 M^{pro} are shown in pine green, while the known actives for the remaining targets are shown in red. ROC AUC values with the SARS-CoV-2 M^{pro} inhibitors regarded as decoys for every target are shown. (B) Binding mode of compound 32 (red) and compound 14 (pine green) toward SARS-CoV-2 M^{pro}. (C) Binding mode of compound 199 toward cathepsin G. (D) Binding mode of compound 177 toward thrombin.

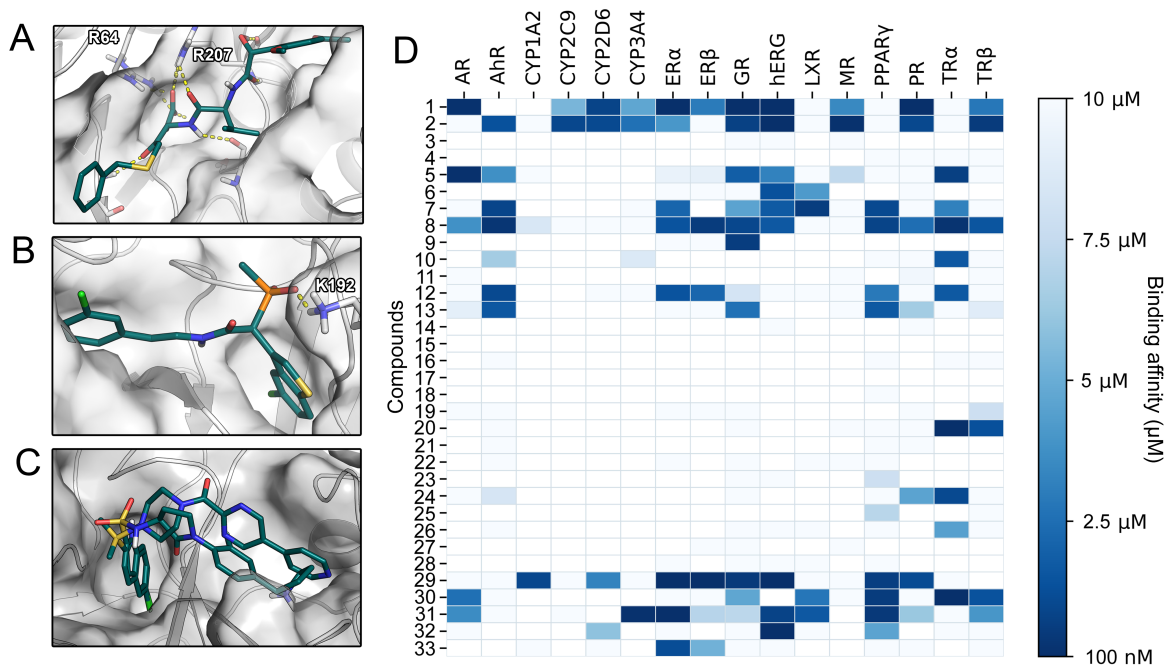


Figure 5. (A) Binding mode of compound 86 toward caspase-3. (B) Binding mode of compound 293 toward chymase. (C) Binding mode of compounds 130 and 137 toward factor Xa. (D) SARS-CoV-2 M^{pro} inhibitors examined with VTL. The predicted binding affinities of the assessed compounds 1-33 for 16 anti-targets are given.

2.4. Toxicity Profiling

As mentioned in the introduction, affinity toward anti-targets is frequently responsible for drug attrition [12]. To establish toxicologically relevant binding profiles of drug candidates our lab has developed the VirtualToxLab (VTL) evaluating their interaction with 16 anti-targets relevant for endocrine disruption, cardiac adverse effects, and extensive or undesired metabolism [13]. Besides estimates for binding affinities against the anti-targets, the VTL provides a parameter referred to as toxic potential serving as a consensus readout for potential undesired effects of the respective compound. SARS-CoV-2 M^{Pro} inhibitors with toxic potential significantly higher than 0.5 (Table S13) included compounds **1** ((*R*)-bepiridil), **2** ((*S*)-bepiridil), **29**, **31** (nelfinavir), and **32** (lopinavir). While compounds **1** and **2** were predicted to interact with multiple nuclear receptors, the hERG channel, and various cytochromes, compounds **29**, **31**, and **32** presented affinity toward a more narrow spectrum of anti-targets (Figure 5D). Compound **29**, for example, almost exclusively bound to nuclear receptors, resulting in a high estimated risk for endocrine disruption [13]. A common feature of compounds **29**, **31**, and **32** was their prediction as hERG binders. The hERG potassium channel is one of the most frequently tested anti-targets in drug development due to its involvement in fatal arrhythmias [12,13,35]. The results regarding the HIV protease inhibitors nelfinavir and lopinavir included in our study confirmed the predictive power of the VTL protocol, as in vitro experiments evidenced their hERG inhibition [35]. Thus, bepiridil (compounds **1** and **2**) displaying a strong interaction for hERG might be at risk to cause cardiac arrhythmia. A large share of the reported SARS-CoV-2 M^{Pro} inhibitors have comparably low molecular weights with 26 of 33 reported compounds below 300 g/mol (Figure S23). Since such fragment-like compounds frequently display low specificity [36], the expansion of these scaffolds might generally decrease their potential for off-target binding and at the same time improve their moderate potency.

2.5. Selectivity from Different Perspectives

We analyzed the selectivity of nine proteases from different perspectives including sequence and active site similarity, the location of hydration hot-spots and preference for certain chemical probes, as well as molecular docking and toxicological profiling. At the first sight, the low sequence similarity among the proteases (Table 1), the different cleavage sites, as well as the different volumes of the active sites may suggest a low risk of off-target binding. However, all investigated off-target proteins showed a considerable active site similarity based on 3D fingerprints and the positioning of catalytic residues (Table 2). Based on these parameters, prostatic and factor Xa were the most similar proteases compared to SARS-CoV-2 M^{Pro}. In six analyzed proteases, a hydration site indicated by water hot-spot was identified near the histidine (denoted as site A). Only for cathepsin G, caspase-3, and thrombin, we could not identify this hydration site. Thus, the displacement of this water molecule would not add any ligand selectivity in this regard. Further, for five of the nine proteases, another hydration site was identified, near the serine/cysteine residues in factor Xa, cathepsin G, prostatic, thrombin, and chymase. Regarding the observed cosolvent densities, we detected a density of acetonitrile in factor Xa, cathepsin G, UCHL1, thrombin, and chymase, but not SARS-CoV-2 M^{Pro}, where this region was explored by pyridine. Distinct placement of pharmacophores matching these differences in density could be exploited to improve inhibitor selectivity. The similarity of the overall densities in factor Xa, cathepsin G, and chymase coupled to the dissimilarity to SARS-CoV-2 M^{Pro}, indicated low potential for off-target binding toward these proteases. A common density of pyridine in the center of multiple sites including the one of SARS-CoV-2 M^{Pro}, however, showed the preference of an aromatic or hydrophobic moiety in this region. Caspase-3 and UCHL1 presented the most unique densities indicating a low potential for off-target binding of inhibitors targeting the remaining proteases.

The docking results of 33 non-covalent SARS-CoV-2 M^{Pro} suggested an overall high potential for binding to factor Xa, thrombin, and cathepsin G. On the other hand, the distribution of the docking scores indicated a low potential of the 33 compounds for binding

toward UCHL1 and caspase-3 (Figure 4A). As the individual enzymes offered different structural factors relevant for potent ligand-protein interaction, their consideration might improve the design of selective inhibitors. Regarding individual compounds, we identified four compounds which had the highest binding affinity toward more than a half of the analyzed proteases: nelfinavir (31) [4,37,38], lopinavir (32) [39–41], pimoziide (33) [42], and baicalein (30) [43] (Figure S24). Similarly, the aforementioned compounds, as well as both stereoisomers of beiperidil (1-2) were predicted to interact with a large panel of known anti-targets. Especially, interactions with the hERG potassium channel, as it was also observed in laboratory experiments, raised safety concerns for several compounds (Figure 5D). In this regard, nelfinavir (32) was not only predicted to interact with other proteases such as factor Xa, but also towards the hERG channel indicating low selectivity of this compound. Interestingly, when we focused on the compound with the highest predicted binding affinity toward a particular enzyme, we could distinguish three groups: one in which nelfinavir binds best (including SARS-CoV M^{Pro}, UCHL1, thrombin, and chymase), second in which pimoziide binds best (including prostaticin, factor Xa, and caspase-3), and third in which baicalein binds best (including SARS-CoV-2 M^{Pro} and cathepsin). A closer look into the first group of enzymes revealed that none of them share the same cleavage site, moreover both SARS-CoV and SARS-CoV-2 M^{Pro}s bind best to different compounds (nelfinavir and baicalein, respectively). As we highlighted selectivity from different perspectives, the different metrics are inherently not always consistent for a single target. To conclude, our predictions indicate the highest potential for off-target binding of SARS-CoV-2 M^{Pro} inhibitors for factor Xa, SARS-CoV M^{Pro}, and cathepsin G. Low potential was determined for prostaticin, thrombin, and to the largest degree, for caspase-3 (Table 3).

Table 3. Conclusions for selectivity of SARS-CoV-2 M^{Pro} inhibitors.

Protein	Docking	Cosolvents	Hydration	Site Similarity ^a
SARS-CoV M ^{Pro}	**	*	**	***
Caspase-3	none	none	none	*
Factor Xa	**	**	*	*
Cathepsin G	***	*	none	*
UCHL1	***	*	*	none
Prostaticin	**	none	*	**
Thrombin	**	none	*	**
Chymase	**	*	*	*

The potential for off-target binding of SARS-CoV-2 M^{Pro} inhibitors against the selected panel of proteases based on molecular docking, cosolvent MD simulations, and hydration site analysis. The potential was defined according to asterisks from none to three. ^a Consensus of binding site similarity determined with FuzCav and PIPSA.

3. Materials and Methods

3.1. Selection of Proteases

The panel of proteases for this work were selected using the VAST+ tool [44] by using the SARS-CoV-2 M^{Pro} structure (PDB ID: 6Y2E) as input. The VAST+ protocol determines similar macromolecules to a query structure by computing the superposition of three dimensional protein structures relying purely on geometric measures. The results were filtered to only match only human proteases. The next criterion was the reaction mechanism of the selected protease, to ensure a representation of both cysteine and serine proteases in our study. Finally, we examined the availability of crystal structures with cocrystallized ligands leading to the selection of proteases listed in Table 1. The preprocessing of the structures is given in the Supplementary Materials.

3.2. Similarity of Proteins and Active Sites

The sequence identity was determined using FASTA sequences derived from the UniProt database [45] (Table S14). In the case of both SARS-CoV-2 and SARS-CoV M^{Pro}, as well as factor Xa and thrombin, we truncated the sequences to cover the entry in the

respective crystal structures limiting the analysis to the catalytic unit. The sequences were aligned with the ClustalW algorithm [46] in the UGENE suite (v34.0) [47]. The sequence identity was computed based on matches of the respective protein to SARS-CoV-2 M^{PRO} in respect to the length of the sequence.

The active sites of the proteases were aligned in PyMOL using the pair_fit command. Each protease was aligned to the reference structure: the SARS-CoV-2 M^{PRO} (PDB ID: 6Y2E), fitting the protease catalytic histidine residue with the H41 of the SARS-CoV-2 M^{PRO}, protease catalytic serine or cysteine residue with the C145 of the SARS-CoV-2 M^{PRO}, and the protease aspartic acid residue with the catalytic water molecule (ID 582) of the SARS-CoV-2 M^{PRO}. The RMSD values were computed according to the superimposition of the catalytic residues in PyMOL.

To determine the similarity of the active sites of the considered off-targets to SARS-CoV-2 M^{PRO}, we used the FuzCav [25] routine. This routine computes the similarity based on fingerprints for each binding site incorporating pharmacophoric properties from the coordinates of surrounding α -carbon atoms. As input, we selected residues in 5 Å around the cocrystallized ligands in the structures.

To compare the electrostatic interaction properties of the binding cavities, we used the PIPSA [26] software. First, we preprocessed the structures using PDB2PQR tool [48], and we calculated the Adaptive Poisson-Boltzmann Solver (APBS) electrostatics potentials [49] setting the grid spacing to 0.6. Then, we calculated the similarity matrix (Hodgkin index) of the binding cavities from APBS grids. The binding pocket was set as a sphere with a radius of 12.5 Å around the geometric centre of the catalytic amino acids after superposition.

3.3. Cosolvent MD Simulations

The cosolvent MD simulations were conducted with the Mixed Solvent MD workflow of the Desmond (v2019-1) simulation engine [50] with acetonitrile, isopropanol, and pyridine as probe molecules as they are water-miscible and feature a low potential for aggregation. The concentration of the probe molecules was selected at 5% (by volume) and, if required for system setup, the water buffer parameter was increased from 12.0 to 15.0, as described in the documentation of the workflow. From the above-mentioned protein structures, monomers were retained to reduce the computational cost of the simulations. Furthermore, the ligands were removed from the structures to sample the respective binding sites. The simulations were performed with the default specifications at a temperature of 300 K and the OPLS_2005 force field in an NPT ensemble. After an equilibration of 15 ns, production runs of each probe were individually executed for 5 ns with 10 replica simulations resulting in a cumulative simulation time of 600 ns per protein.

3.4. Classical MD Simulations

The H++ server [51] was used to protonate all proteins structures listed in Table S15 using standard parameters at pH 7.4. The missing 4-amino-acids-long loop of the 1Q2W SARS-CoV M^{PRO} model was added using the corresponding loop of the 6LU7 model (from SARS-CoV-2 M^{PRO}), and its quality was confirmed by comparison with another crystal structure of SARS-CoV M^{PRO} (PDB ID: 2H2Z). Counter ions were added to neutralize the systems as shown in Table S15. Water molecules were placed using the combination of 3D-RISM [52] and the Placevent algorithm [53]. AMBER 18 LEaP [54] was used to immerse models in a truncated octahedral box with 12 Å radius of TIP3P water molecules and prepare the systems for simulation using the ff14SB force field [55]. The number of added water molecules is shown in Table S15. The PMEMD CUDA package of AMBER 18 software [54] was used to run 10 replicas of 50 ns for each system. The starting geometry for each system was kept, but the initial vectors were randomly assigned to enrich conformational sampling. The minimization procedure consisted of 2000 steps, involving 1000 steepest descent steps followed by 1000 steps of conjugate gradient energy minimization, with decreasing constraints on the protein backbone (500, 125 and 25 kcal·mol⁻¹·Å⁻²) and a conjugate gradient minimization with no constraints. Next,

the systems were gradually heated from 0 to 300 K over 20 ps using a Langevin thermostat with a collision frequency of 1.0 ps^{-1} in periodic boundary conditions with constant volume. Equilibration stage was run using the periodic boundary conditions with constant pressure for 1 ns (10 ns in the case of caspase-3 and factor Xa structures to ensure proper equilibration) with 1 fs step using Langevin dynamics with a frequency collision of 1 ps^{-1} to maintain temperature. Production stage was run for 50 ns with a 2 fs time step using Langevin dynamics with a collision frequency of 1 ps^{-1} to maintain constant temperature. Long-range electrostatic interactions were treated using the particle mesh Ewald method with a non-bonded cut-off of 10 \AA and the SHAKE algorithm. The coordinates were saved at an interval of 1 ps. The computation of the maximum available volume (MAV) is given in the Supplementary Information.

3.5. Water Molecules Tracking, Hot-Spots Identification

AQUA-DUCT 1.0 software [56] was used to track water molecules for all proteases in each simulation replica. Tracking of water molecules was conducted in two specific regions: the *Object*, which represents the cavity of a particular interest, and the *Scope* representing the whole macromolecule. The *Object* was defined as a 4 \AA sphere around the centroid of the active site residues of each protein (catalytic residues listed in Table 2), and the *Scope* was defined as the interior of a convex hull of α -carbon atoms in all structures. AQUA-DUCT was also used for identification of hot-spots, defined as the regions of the highest density of traced molecules within the protein interior. AQUA-DUCT is able to calculate hot-spots using two types of data: (i) using only the pathways of those molecules that entered the *Object* to calculate local hot-spots, and (ii) using the pathways of all molecules that entered the *Scope* region to calculate the global hot-spots. Both types of hot-spots were calculated for each of the simulation replica, and then simplified using the `hs_gsimplifier.py` script. The `hs_simplifier.py` script was used to analyze the positions of all identified hot-spots and grouped those hot-spots which were located within a radius of 3 \AA in the case of global hot-spots, and 2 \AA in the case of local hot-spots. Then it provided the information about the particular simulation replica in which the hot-spot was identified. The information was kept and the simplified hot-spots are colour-coded according to their occupation. Those which were the most common were colored dark red, and those which were rare are white.

3.6. Molecular Docking and Validation

To ensure a high accuracy in pose prediction, we determined a fitting structural ensemble for each target that was able reproduce to binding modes of a maximal portion of non-covalent cocrystallized ligands. We evaluated the Glide standard-precision (SP) [57] as well as the `smina` [58] docking protocol throughout this study. While default setting were retained for Glide including the grid generation, an exhaustiveness of 16, a cubic search space with a side length of 21 \AA , as well as a random seed of 42 was configured for `smina`. The centroids for the respective search spaces were determined by computing the mass center of the cocrystallized ligand. The RMSD between the docked pose and the respective cocrystallized ligand was computed using the `rmsd.py` script that comes with Maestro after protein structure alignment. In the case of unsatisfying pose prediction, we created structural ensembles (Tables S16–S18) by clustering representative structures of MD simulations as detailed in our previous work [4]. For each protein, the docking protocol combined with the structural ensemble correctly reproducing the highest number of cocrystallized ligands was selected to assess the potential to discriminate random decoy compounds from known binders. Actives for each target were collected from various sources including crystal structures, the literature, as well as the PubChem database [59]. The respective decoy compounds were generated using the novel DUDE-Z web server [60] with SMILES strings as input. Since LigPrep frequently generated multiple plausible protonation states and stereoisomers of the decoys, we chose the best docking score against the ensemble of each isomer. Together with the results from the known binders, the docking scores were submitted to the Screening Explorer web server [61] to determine enrichment metrics including

the maximal reachable enrichment as well as the ROC AUC. The following selectivity assessment was conducted by using SARS-CoV-2 M^{PRO} inhibitors as decoy compounds combined with known binders as actives. The ROC AUC metric was again obtained from the Screening Explorer web server and the scores were compared in a histogram computed using the Matplotlib [62] python library. To compare the absolute values of the docking scores among the different proteins, smina was used to rescore the poses obtained from Glide SP docking in the case of SARS-CoV-2 M^{PRO}, caspase-3, and chymase. Lastly, we aimed on deducing the structural factors for selective binding by visual inspection of the binding modes.

3.7. MD and MM/GBSA Post-Processing

To obtain more confidence in the results from docking, we post-processed the docking poses with the MM/GBSA protocol. Using Desmond (v2019-1), we conducted 2 ns simulations of all 576 ligand-protein complexes with different targets. We used the OPLS_2005 force field in an NPT ensemble with a temperature of 310 K maintained by the Nose-Hoover thermostat and atmospheric pressure maintained by the Martyna-Tobias-Klein barostat. The orthorhombic periodic boundary system was solvated with TIP3P water molecules. Long-range interactions were treated with the u-series algorithm [63] and short-range interactions were cut off at 9 Å, while bonds to hydrogen atoms were constrained with the M-SHAKE algorithm. The default relaxation protocol in Desmond was applied before the production phase. Atomic coordinates were deposited in an interval of 20 ps and the thermal_mmgbsa.py script that comes with Maestro (v2019-4) was applied to obtain binding free energies of the last 10 frames of the simulations, which were averaged thereafter.

4. Conclusions

Due to the current COVID-19 pandemic and the lack of specific therapeutics, many small molecules that inhibit SARS-CoV-2 M^{PRO} have been proposed. This work aims to support further development of these compounds in order to avoid safety issues due to off-target binding, which is one of the major reasons for late stage drug attrition. We addressed the concern of selectivity and off-target binding of 33 published, experimentally confirmed SARS-CoV-2 M^{PRO} inhibitors by predicting their affinity toward eight different proteases and profiling their active sites regarding hydration site and small-molecule hot-spots. Even though the selected off-target proteins presented a low global sequence identity to SARS-CoV-2 M^{PRO}, their binding sites were considerably similar. This similarity could explain the predicted affinities of the SARS-CoV-2 M^{PRO} inhibitors, which presented a considerable overlap with actives against chymase, UCHL1, and cathepsin G. Interestingly, inhibitors of prostatic displayed higher predicted binding affinities to SARS-CoV-2 M^{PRO} than its native inhibitors. Refining the SARS-CoV-2 M^{PRO} inhibitors may be necessary to achieve higher affinities toward their designated target, as well as to improve selectivity and thereby decrease off-target binding. Around one third of the investigated compounds presented medium to high potential for endocrine disruption, altered drug metabolism, or cardiac adverse events based on the prediction of binding affinities towards 16 well established anti-targets. Our work showed that, while there are many proposed SARS-CoV-2 M^{PRO} inhibitors, they generally exhibit poor selectivity and may cause pharmacological undesired effects by off-target binding. Even though the panel of proteases share a comparably low sequence identity and different substrate specificity, enzymes such as cathepsin G, factor Xa, as well as UCHL1 could be relevant off-targets for novel antivirals. If experimental testing and compound optimization efforts will be guided to achieve selectivity over the suggested anti-targets, novel antivirals could have an improved safety profile.

Supplementary Materials: The following are available online at <https://www.mdpi.com/1422-0067/22/4/2065/s1>, Figures S1 and S2: Active site similarity, Table S1: Maximal available volumes, Figures S3–S18: Structures of actives for each protease, Tables S2–S10: Origin of active compounds, Figure S19: Binding free energies obtained by MM/GBSA calculations, Figure S20: Score distribution of SARS-CoV-2 actives against every target, Figures S21 and S22: 2D depiction of discussed binding

modes, Tables S11 and S12: metrics of docking protocol validation, Table S13: Toxic potential of SARS-CoV-2 actives, Figure S23: Molecular weight of SARS-CoV-2 actives, Figure S24: Comparison of docking scores, Supporting text: Model preparation, Supporting text: MAV calculation, Table S14: UniProt identifiers for all proteases, Table S15: Ions and water molecules for conventional MD, Tables S16–S18: Crystal structures considered in docking. References [64–66] are cited in the Supplementary Materials.

Author Contributions: Conceptualization, A.F.; methodology, A.F.; formal analysis and investigation, A.F., M.S. (Manuel Sellner), M.B., K.M.; writing—original draft preparation, A.F., K.M., M.B.; writing—review and editing, A.F., M.S. (Manuel Sellner), K.M., M.B., M.A.L., A.G., M.S. (Martin Smieško); visualization, A.F., K.M., M.B.; supervision, M.A.L., A.G., M.S. (Martin Smieško); project administration, A.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Raw data obtained from molecular docking, VTL, MM/GBSA, co-solvent MD, and conventional MD are available in a public repository on GitHub under https://github.com/mmodbasel/SARS-CoV2_selectivity, accessed on 4 January 2021.

Acknowledgments: We gratefully acknowledge the support of NVIDIA Corporation with the donation of a Titan Xp GPU used for this research. This research was supported in part by PL-Grid Infrastructure.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. World Health Organisation. *Novel Coronavirus (2019-nCoV) Situation Reports*; WHO: Geneva, Switzerland, 2020.
2. Kalil, A.C. Treating COVID-19—Off-Label Drug Use, Compassionate Use, and Randomized Clinical Trials During Pandemics. *JAMA* **2020**, *323*, 1897–1898. [[CrossRef](#)]
3. Chen, N.; Zhou, M.; Dong, X.; Qu, J.; Gong, F.; Han, Y.; Qiu, Y.; Wang, J.; Liu, Y.; Wei, Y.; et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: A descriptive study. *Lancet* **2020**, *395*, 507–513. [[CrossRef](#)]
4. Fischer, A.; Sellner, M.; Neranjan, S.; Smieško, M.; Lill, M.A. Potential inhibitors for novel coronavirus protease identified by virtual screening of 606 million compounds. *Int. J. Mol. Sci.* **2020**, *21*, 3626. [[CrossRef](#)]
5. Dai, W.; Zhang, B.; Jiang, X.M.; Su, H.; Li, J.; Zhao, Y.; Xie, X.; Jin, Z.; Peng, J.; Liu, F.; et al. Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease. *Science* **2020**, *368*, 1331–1335. [[CrossRef](#)] [[PubMed](#)]
6. Jin, Z.; Du, X.; Xu, Y.; Deng, Y.; Liu, M.; Zhao, Y.; Zhang, B.; Li, X.; Zhang, L.; Peng, C.; et al. Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature* **2020**, *582*, 289–293. [[CrossRef](#)]
7. He, J.; Hu, L.; Huang, X.; Wang, C.; Zhang, Z.; Wang, Y.; Zhang, D.; Ye, W. Potential of coronavirus 3C-like protease inhibitors for the development of new anti-SARS-CoV-2 drugs: Insights from structures of protease and inhibitors. *Int. J. Antimicrob. Agents* **2020**, 106055. [[CrossRef](#)]
8. Ma, C.; Sacco, M.D.; Hurst, B.; Townsend, J.A.; Hu, Y.; Szeto, T.; Zhang, X.; Tarbet, B.; Marty, M.T.; Chen, Y.; et al. Boceprevir, GC-376, and calpain inhibitors II, XII inhibit SARS-CoV-2 viral replication by targeting the viral main protease. *Cell Res.* **2020**, *30*, 678–692. [[CrossRef](#)]
9. Kandeel, M.; Al-Nazawi, M. Virtual screening and repurposing of FDA approved drugs against COVID-19 main protease. *Life Sci.* **2020**, *251*, 117627. [[CrossRef](#)] [[PubMed](#)]
10. Van Vleet, T.R.; Liguori, M.J.; Lynch, J.J.; Rao, M.; Warder, S. Screening Strategies and Methods for Better Off-Target Liability Prediction and Identification of Small-Molecule Pharmaceuticals. *SLAS Discov.* **2019**, *24*, 1–24. [[CrossRef](#)]
11. Bowes, J.; Brown, A.J.; Hamon, J.; Jarolimek, W.; Sridhar, A.; Waldron, G.; Whitebread, S. Reducing safety-related drug attrition: The use of in vitro pharmacological profiling. *Nat. Rev. Drug Discov.* **2012**, *11*, 909–922. [[CrossRef](#)] [[PubMed](#)]
12. Waring, M.J.; Arrowsmith, J.; Leach, A.R.; Leeson, P.D.; Mandrell, S.; Owen, R.M.; Pairaudeau, G.; Pennie, W.D.; Pickett, S.D.; Wang, J.; et al. An analysis of the attrition of drug candidates from four major pharmaceutical companies. *Nat. Rev. Drug Discov.* **2015**, *14*, 475–486. [[CrossRef](#)] [[PubMed](#)]
13. Vedani, A.; Dobler, M.; Hu, Z.; Smieško, M. OpenVirtualToxLab-A platform for generating and exchanging in silico toxicity data. *Toxicol. Lett.* **2015**, *232*, 519–532. [[CrossRef](#)] [[PubMed](#)]
14. Rao, M.S.; Gupta, R.; Liguori, M.J.; Hu, M.; Huang, X.; Mantena, S.R.; Mittelstadt, S.W.; Blomme, E.A.G.; Van Vleet, T.R. Novel Computational Approach to Predict Off-Target Interactions for Small Molecules. *Front. Big Data* **2019**, *2*, 1–17. [[CrossRef](#)]
15. Xia, S.; Liu, M.; Wang, C.; Xu, W.; Lan, Q.; Feng, S.; Qi, F.; Bao, L.; Du, L.; Liu, S.; et al. Inhibition of SARS-CoV-2 (previously 2019-nCoV) infection by a highly potent pan-coronavirus fusion inhibitor targeting its spike protein that harbors a high capacity to mediate membrane fusion. *Cell Res.* **2020**, *30*, 343–355. [[CrossRef](#)]

16. Tang, N.; Bai, H.; Chen, X.; Gong, J.; Li, D.; Sun, Z. Anticoagulant treatment is associated with decreased mortality in severe coronavirus disease 2019 patients with coagulopathy. *J. Thromb. Haemost.* **2020**, 1094–1099. [[CrossRef](#)] [[PubMed](#)]
17. Connors, J.M.; Levy, J.H. COVID-19 and its implications for thrombosis and anticoagulation. *Blood* **2020**, *135*, 2033–2040. [[CrossRef](#)] [[PubMed](#)]
18. Fischer, A.; Smieško, M. Allosteric binding sites on nuclear receptors: Focus on drug efficacy and selectivity. *Int. J. Mol. Sci.* **2020**, *21*, 534. [[CrossRef](#)] [[PubMed](#)]
19. Ghanakota, P.; Carlson, H.A. Driving Structure-Based Drug Discovery through Cosolvent Molecular Dynamics. *J. Med. Chem.* **2016**, *59*, 10383–10399. [[CrossRef](#)]
20. Kimura, S.R.; Hu, H.P.; Ruvinsky, A.M.; Sherman, W.; Favia, A.D. Deciphering Cryptic Binding Sites on Proteins by Mixed-Solvent Molecular Dynamics. *J. Chem. Inf. Model.* **2017**, *57*, 1388–1401. [[CrossRef](#)] [[PubMed](#)]
21. Keretsu, S.; Bhujbal, S.P.; Joo Cho, S. Computational study of paroxetine-like inhibitors reveals new molecular insight to inhibit GRK2 with selectivity over ROCK1. *Sci. Rep.* **2019**, *9*, 1–14. [[CrossRef](#)]
22. Ferguson, F.M.; Gray, N.S. Kinase inhibitors: The road ahead. *Nat. Rev. Drug Discov.* **2018**, *17*, 353–376. [[CrossRef](#)]
23. Chaudhury, S.; Gray, J.J. Identification of Structural Mechanisms of HIV-1 Protease Specificity Using Computational Peptide Docking: Implications for Drug Resistance. *Structure* **2009**, *17*, 1636–1648. [[CrossRef](#)] [[PubMed](#)]
24. Boyd, S.E.; Garcia de la Banda, M.; Pike, R.N.; Whisstock, J.C.; Rudy, G.B. PoPS: A computational tool for modeling and predicting protease specificity. *Proc. IEEE Comp. Syst. Bioinform. Conf.* **2004**, 372–381. [[CrossRef](#)]
25. Weill, N.; Rognan, D. Alignment-free ultra-high-throughput comparison of druggable protein-ligand binding sites. *J. Chem. Inf. Model.* **2010**, *50*, 123–135. [[CrossRef](#)] [[PubMed](#)]
26. Blomberg, N.; Gabdoulline, R.R.; Nilges, M.; Wade, R.C. Classification of protein sequences by homology modeling and quantitative analysis of electrostatic similarity. *Protein. Struct. Funct. Genet.* **1999**, *37*, 379–387. [[CrossRef](#)]
27. Perona, J.J.; Craik, C.S. Structural basis of substrate specificity in the serine proteases. *Protein. Sci.* **1995**, *4*, 337–360. [[CrossRef](#)]
28. Tan, Y.S.; Spring, D.R.; Abell, C.; Verma, C.S. The Application of Ligand-Mapping Molecular Dynamics Simulations to the Rational Design of Peptidic Modulators of Protein-Protein Interactions. *J. Chem. Ther. Comput.* **2015**, *11*, 3199–3210. [[CrossRef](#)]
29. Yang, Y.; Hu, B.; Lill, M.A. *WATsite2.0 with PyMOL Plugin: Hydration Site Prediction and Visualization BT—Protein Function Prediction: Methods and Protocols*; Springer: New York, NY, USA, 2017; pp. 123–134. [10](#). [[CrossRef](#)]
30. Bzówka, M.; Mitusińska, K.; Raczyńska, A.; Samol, A.; Tuszyński, J.A.; Góra, A. Structural and evolutionary analysis indicate that the sars-COV-2 mpro is a challenging target for small-molecule inhibitor design. *Int. J. Mol. Sci.* **2020**, *21*, 3099. [[CrossRef](#)] [[PubMed](#)]
31. Ortiz, A.R.; Gomez-Puertas, P.; Leo-Macias, A.; Lopez-Romero, P.; Lopez-Viñas, E.; Morreale, A.; Murcia, M.; Wang, K. Computational approaches to model ligand selectivity in drug design. *Curr. Top. Med. Chem.* **2006**, *6*, 41–55. [[CrossRef](#)]
32. Wang, Y.; Bryant, S.H.; Cheng, T.; Wang, J.; Gindulyte, A.; Shoemaker, B.A.; Thiessen, P.A.; He, S.; Zhang, J. PubChem BioAssay: 2017 update. *Nucleic Acids Res.* **2017**, *45*, D955–D963. [[CrossRef](#)]
33. Kumar, A.; Zhang, K.Y.J. A cross docking pipeline for improving pose prediction and virtual screening performance. *J. Comput.-Aided Mol. Des.* **2018**, *32*, 163–173. [[CrossRef](#)]
34. Wang, E.; Sun, H.; Wang, J.; Wang, Z.; Liu, H.; Zhang, J.Z.; Hou, T. End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design. *Chem. Rev.* **2019**, *119*, 9478–9508. [[CrossRef](#)] [[PubMed](#)]
35. Anson, B.D.; Weaver, J.G.R.; Ackerman, M.J.; Akinsete, O.; Henry, K.; January, C.T.; Badley, A.D. Blockade of HERG channels by HIV protease inhibitors. *Lancet (Lond. UK)* **2005**, *365*, 682–686. [[CrossRef](#)]
36. Chen, Y.; Shoichet, B.K. Molecular docking and ligand specificity in fragment-based inhibitor discovery. *Nat. Chem. Biol.* **2009**, *5*, 358–364. [[CrossRef](#)] [[PubMed](#)]
37. Huynh, T.; Wang, H.; Luan, B. In Silico Exploration of the Molecular Mechanism of Clinically Oriented Drugs for Possibly Inhibiting SARS-CoV-2's Main Protease. *J. Phys. Chem. Lett.* **2020**, *11*, 4413–4420. [[CrossRef](#)]
38. Yamamoto, N.; Matsuyama, S.; Hoshino, T.; Yamamoto, N. Nelfinavir inhibits replication of severe acute respiratory syndrome coronavirus 2 in vitro. *bioRxiv* **2020**. [[CrossRef](#)]
39. Liu, X.; Wang, X.J. Potential inhibitors against 2019-nCoV coronavirus M protease from clinically approved medicines. *J. Genet. Genom.* **2020**, *47*, 119–121. [[CrossRef](#)]
40. Kumar, Y.; Singh, H.; Patel, C.N. In silico prediction of potential inhibitors for the main protease of SARS-CoV-2 using molecular docking and dynamics simulation based drug-repurposing. *J. Infect. Public Health* **2020**, *13*, 1210–1223. [[CrossRef](#)] [[PubMed](#)]
41. Narkhede, R.; Cheke, R.; Ambhore, J.; Shinde, S. The Molecular Docking Study of Potential Drug Candidates Showing Anti-COVID-19 Activity by Exploring of Therapeutic Targets of SARS-CoV-2. *Eurasian J. Med. Oncol.* **2020**, *4*, 185–195. [[CrossRef](#)]
42. Vatansever, E.; Yang, K.; Kratch, K.; Drelich, A.; Cho, C.; Mellott, D.; Xu, S.; Tseng, C.; Liu, W. Bepridil is potent against SARS-CoV-2 In Vitro. *bioRxiv* **2020**. [[CrossRef](#)]
43. Eleftheriou, P.; Amanatidou, D.; Petrou, A.; Geronikaki, A. In Silico Evaluation of the Effectivity of Approved Protease Inhibitors against the Main Protease of the Novel SARS-CoV-2 Virus. *Molecules* **2020**, *25*, 2529. [[CrossRef](#)] [[PubMed](#)]
44. Madej, T.; Lanczycki, C.J.; Zhang, D.; Thiessen, P.A.; Geer, R.C.; Marchler-Bauer, A.; Bryant, S.H. MMDB and VAST+: Tracking structural similarities between macromolecular complexes. *Nucleic Acids Res.* **2014**. [[CrossRef](#)]
45. Consortium, T.U. UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res.* **2019**, *47*, D506–D515. [[CrossRef](#)]

46. Thompson, J.D.; Higgins, D.G.; Gibson, T.J. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **1994**, *22*, 4673–4680. [[CrossRef](#)]
47. Okonechnikov, K.; Golosova, O.; Fursov, M.; Varlamov, A.; Vaskin, Y.; Efremov, I.; German Grehov, O.G.; Kandrov, D.; Rasputin, K.; Syabro, M.; et al. Unipro UGENE: A unified bioinformatics toolkit. *Bioinformatics* **2012**, *28*, 1166–1167. [[CrossRef](#)] [[PubMed](#)]
48. Dolinsky, T.J.; Nielsen, J.E.; McCammon, J.A.; Baker, N.A. PDB2PQR: An automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.* **2004**, *32*, W665–W667. [[CrossRef](#)]
49. Baker, N.A.; Sept, D.; Joseph, S.; Holst, M.J.; McCammon, J.A. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 10037–10041. [[CrossRef](#)] [[PubMed](#)]
50. Bowers, K.; Chow, E.; Xu, H.; Dror, R.; Eastwood, M.; Gregersen, B.; Klepeis, J.; Kolossvary, I.; Moraes, M.; Sacerdoti, F.; et al. Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. In Proceedings of the ACM/IEEE SC 2006 Conference (SC'06), Tampa, FL, USA, 11–17 November 2006; p. 43. [[CrossRef](#)]
51. Anandakrishnan, R.; Aguilar, B.; Onufriev, A.V. H++ 3.0: Automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res.* **2012**, *40*, W537–W541. [[CrossRef](#)] [[PubMed](#)]
52. Luchko, T.; Gusarov, S.; Roe, D.R.; Simmerling, C.; Case, D.A.; Tuszynski, J.; Kovalenko, A. Three-Dimensional Molecular Theory of Solvation Coupled with Molecular Dynamics in Amber. *J. Chem. Ther. Comput.* **2010**, *6*, 607–624. [[CrossRef](#)]
53. Sindhikara, D.J.; Yoshida, N.; Hirata, F. Placevent: An algorithm for prediction of explicit solvent atom distribution-Application to HIV-1 protease and F-ATP synthase. *J. Comput. Chem.* **2012**, *33*, 1536–1543. [[CrossRef](#)] [[PubMed](#)]
54. Case, D.A.; Walker, R.C.; Cheatham, T.E.; Simmerling, C.; Roitberg, A.; Merz, K.M.; Luo, R.; Darden, T. *Amber 18*; University of California: San Francisco, CA, USA, 2018.
55. Maier, J.A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K.E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Comput.* **2015**, *11*. [[CrossRef](#)] [[PubMed](#)]
56. Magdziarz, T.; Mitusińska, K.; Bzówka, M.; Raczyńska, A.; Stańczak, A.; Banas, M.; Bagrowska, W.; Góra, A. AQUA-DUCT 1.0: Structural and functional analysis of macromolecules from an intramolecular voids perspective. *Bioinformatics* **2020**, *36*, 2599–2601. [[CrossRef](#)]
57. Halgren, T.A.; Murphy, R.B.; Friesner, R.A.; Beard, H.S.; Frye, L.L.; Pollard, W.T.; Banks, J.L. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J. Med. Chem.* **2004**, *47*, 1750–1759. [[CrossRef](#)] [[PubMed](#)]
58. Koes, D.R.; Baumgartner, M.P.; Camacho, C.J. Lessons learned in empirical scoring with smina from the CSAR 2011 benchmarking exercise. *J. Chem. Inf. Model.* **2013**, *53*, 1893–1904. [[CrossRef](#)]
59. Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B.A.; Thiessen, P.A.; Yu, B.; et al. PubChem 2019 update: Improved access to chemical data. *Nucleic Acids Res.* **2019**, *47*, D1102–D1109. [[CrossRef](#)] [[PubMed](#)]
60. Mysinger, M.M.; Carchia, M.; Irwin, J.J.; Shoichet, B.K. Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. *J. Med. Chem.* **2012**, *55*, 6582–6594. [[CrossRef](#)] [[PubMed](#)]
61. Empereur-Mot, C.; Zagury, J.F.; Montes, M. Screening Explorer—An Interactive Tool for the Analysis of Screening Results. *J. Chem. Inf. Model.* **2016**, *56*, 2281–2286. [[CrossRef](#)]
62. Hunter, J.D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95. [[CrossRef](#)]
63. Shaw, D.E.; Grossman, J.P.; Bank, J.A.; Batson, B.; Butts, J.A.; Chao, J.C.; Deneroff, M.M.; Dror, R.O.; Even, A.; Fenton, C.H.; et al. Anton 2: Raising the Bar for Performance and Programmability in a Special-Purpose Molecular Dynamics Supercomputer. In Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, SC, Austin TX, USA, 15 November 2015; pp. 41–53. [[CrossRef](#)]
64. Madhavi Sastry, G.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments. *J. Comput.-Aided Mol. Des.* **2013**, *27*, 221–234. [[CrossRef](#)]
65. Schrödinger, LCC. *Maestro Small-Molecule Drug Discovery Suite 2019-3*; Schrödinger, LCC: New York, NY, USA, 2019.
66. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242. [[CrossRef](#)] [[PubMed](#)]