JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim Perspective

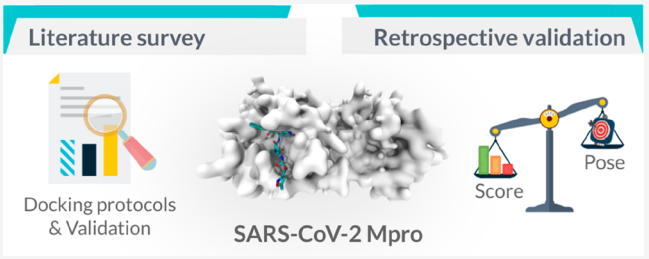
Strengths and Weaknesses of Docking Simulations in the SARS-CoV-2 Era: the Main Protease (Mpro) Case Study

Manuel A. Llanos, Melisa E. Gantner, Santiago Rodriguez, Lucas N. Alberca, Carolina L. Bellera, Alan Talevi, and Luciana Gavernet*

Cite This: https://doi.org/10.1021/acs.jcim.1c00404 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The scientific community is working against the clock to arrive at therapeutic interventions to treat patients with COVID-19. Among the strategies for drug discovery, virtual screening approaches have the capacity to search potential hits within millions of chemical structures in days, with the appropriate computing infrastructure. In this article, we first analyzed the published research targeting the inhibition of the main protease (Mpro), one of the most studied targets of SARS-CoV-2, by docking-based methods. An alarming finding was the lack of an adequate validation of the docking protocols (i.e., pose prediction and virtual screening accuracy) before applying them in virtual screening campaigns. The performance of the docking protocols was tested at some level in 57.7% of the 168 investigations analyzed. However, we found only three examples of a complete retrospective analysis of the scoring functions to quantify the virtual screening accuracy of the methods. Moreover, only two publications reported some experimental evaluation of the proposed hits until preparing this manuscript. All of these findings led us to carry out a retrospective performance validation of three different docking protocols, through the analysis of their pose prediction and screening accuracy. Surprisingly, we found that even though all tested docking protocols have a good pose prediction, their screening accuracy is quite limited as they fail to correctly rank a test set of compounds. These results highlight the importance of conducting an adequate validation of the docking protocols before carrying out virtual screening campaigns, and to experimentally confirm the predictions made by the models before drawing bold conclusions. Finally, successful structure-based drug discovery investigations published during the redaction of this manuscript allow us to propose the inclusion of target flexibility and consensus scoring as alternatives to improve the accuracy of the methods.



INTRODUCTION

By the end of 2019, a new type of coronavirus, the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), was detected in humans for the first time, and it was identified as being responsible for the disease known as COVID-19.[1] Nowadays, the virus has spread all over the world. The scientific community has joined forces to arrive at therapeutic interventions, which include tools for immunization (such as vaccines or hyperimmune globulin), the repositioning of known drugs (such as Hydroxychloroquine, Chloroquine, Remdesivir, and Oseltamivir), and the development of novel antiviral compounds.[2,3] For the last two strategies, *in silico* drug discovery tools have become an interesting option, particularly if we consider the current capacity of the computers to analyze millions of structures in short periods of time, even days when the necessary computing power is available. In addition to the advantage of speeding up the search process, the use of computational approaches allows the researchers to continue working at home (at least partially) with remote access to their computational facilities. This fact is opportune considering that, in many parts of the globe, total or partial isolation continues to be one of the government's main measures to prevent the spread of the virus.

Molecular docking (or just "docking") consists of simulating the interaction between two partners, usually a small molecule (such as a drug candidate) and a biological target.[4,5] The docking software combines a search algorithm to explore different conformations and orientations of the ligand into a predefined binding site, using a scoring function to estimate the binding energy associated with each generated pose.[5] As a result, the program yields a ranking of the best solutions, that is, the poses with the lowest binding energy.

Docking calculations can be applied to chemical databases of small molecules, in order to rank them according to the predicted binding energy. This strategy is known as target- or

Received: April 9, 2021

https://doi.org/10.1021/acs.jcim.1c00404
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

structure-based virtual screening and can be helpful in selecting the best candidates to be synthesized, acquired, or evaluated in wet experiments. Moreover, it provides initial hypotheses about the binding mode of the compounds, which can be a useful starting point for hit optimization.

One of the main limitations of the method relates to how it models the conformational changes in the protein triggered upon binding of a small molecule. The inclusion of full flexibility would imply searching into the whole protein−ligand complex's conformational landscape, which requires significant computational resources and thus limits the application of highly flexible docking protocols for virtual screening purposes. However, strategies to allow protein flexibility at some level can be incorporated, such as considering a few residue side chains as flexible during the simulation, or docking each ligand on several 3D structures of the target (ensemble docking).[4−8] Moreover, postprocessing approaches such as flexible minimization of the complexes or even more sophisticated simulations like molecular dynamics can be applied to model protein flexibility.[5,9] Additionally, the performance of docking software is highly dependent on the biological system under study. Miscalculations could originate from the lack of information about the experimental 3D structure of the target (or about the location of the active site), or from a deficiency of the search algorithms to adequately explore the conformational space of the ligands. Other concerns are the limitations of the scoring functions to assign the lowest score to the binding conformation among all the poses explored or to rank the most active compounds at the top positions in a virtual screening campaign. The latter often results in a high false positive rate.[10−13] To gain confidence about the capacities of the docking programs and protocols to analyze the specific molecular system under study, docking performances can be measured and compared. Two different types of test are usually applied: one to evaluate the ability of the software to find the correct binding pose (which is known as pose prediction or *docking power*), and the other one to analyze its ability to identify the active compounds by its docking score (which is known as virtual screening accuracy or *screening power*).[14−17]

To test the pose prediction, experimental information about the binding mode of a ligand into the target is essential. The compound is first removed from the experimental complex and then docked into the macromolecule to regenerate the system (*redocking*). Additionally, the ligand can be removed from the complex and then docked into another 3D structure of the target, if available (*cross-docking*). The root mean squared deviation (RMSD) of the position of the ligand's atoms is a commonly used metric to assess the agreement between predicted and experimental binding poses.[17]

The screening accuracy is used to measure the aptitude of the scoring function to retrieve active structures among a set of compounds. A test set with known binders among a pool of nonbinders (inactive compounds, decoys, or random structures) is necessary for this purpose, and enrichment metrics, like the area under the receiver operating characteristic curves (AUCROC), can be used to analyze the capacity of the software to discriminate binders from nonbinders.[17,18]

The confidence in docking predictions is increased by using *in silico* validation protocols prior to submitting emerging hits to experimental confirmation. Throughout the following sections, we will first analyze the published research in the field of drug discovery of active compounds to treat COVID-19 by means of docking-based methods. We will focus on the publications available online that involve the search for inhibitors of the SARS-CoV-2 main protease (Mpro, also called 3CLpro), one of the most studied targets of SARS-CoV-2 and other coronaviruses. Second, based on the alarming findings during the literature analysis, we will carry out an evaluation of the performance of three different docking protocols for the Mpro-inhibitors system, through the analysis of their pose predictions and virtual screening accuracies under different conditions.

Main Protease and Its Inhibitors. SARS-CoV-2 Mpro is a cysteine protease with a crucial role in the viral life cycle. It catalyzes the cleavage of the polyproteins translated from the viral RNA in the replication process of the virus.[19,20] Furthermore, the enzyme recognizes the cleavage sites into the polypeptide after a Gln residue. At most sites, the sequence is Leu−Gln−(Ser, Ala, Gly), which is not observed in human proteases.[20] So, inhibiting its activity would block the viral replication, and the inhibitors are unexpected to be toxic for the host (at least at the target level), making it a very promising target.

The enzyme is catalytically active as a homodimer, and each subunit has 306 residues constituting three domains (Figure 1).[20] Domains I and II share the antiparallel six-stranded β-barrel structure, whereas domain III consists of five α-helices arranged into a largely antiparallel globular cluster and connected with domain II by a loop of 13 amino acids (residues from 185 to 198). The binding site involves the Cys145−His41 catalytic dyad and is placed at a cleft between domains I and II (Figure 1). As is usual in proteases, there are

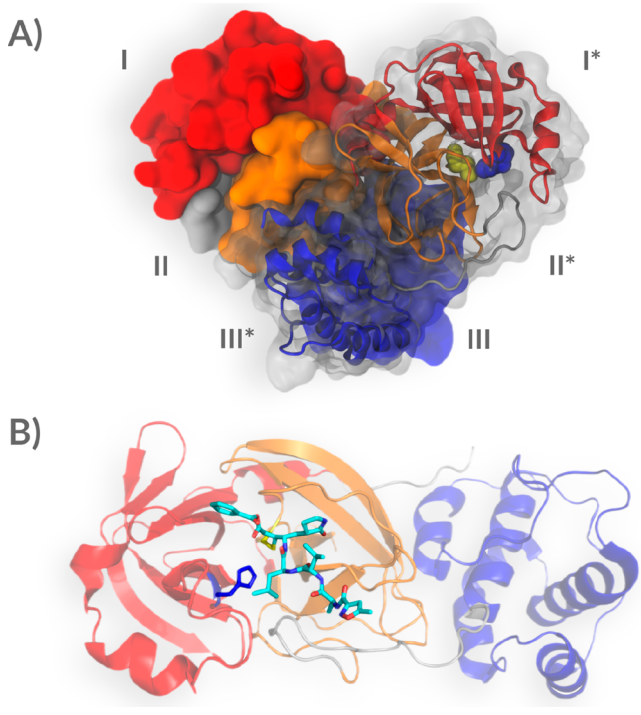


Figure 1. Crystal structure of Mpro taken from the Protein Data Bank (PDB ID: 6LU7). (A) The three different domains for each monomer are shown (I−III/I*−III*). The catalytic dyad is depicted as blue (His41) and yellow (Cys145) spheres. (B) The covalent inhibitor N3 into SARS-CoV-2 Mpro binding site. N3 is depicted as cyan sticks, whereas catalytic His41 and Cys145 are colored in blue and yellow, respectively.

https://doi.org/10.1021/acs.jcim.1c00404
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

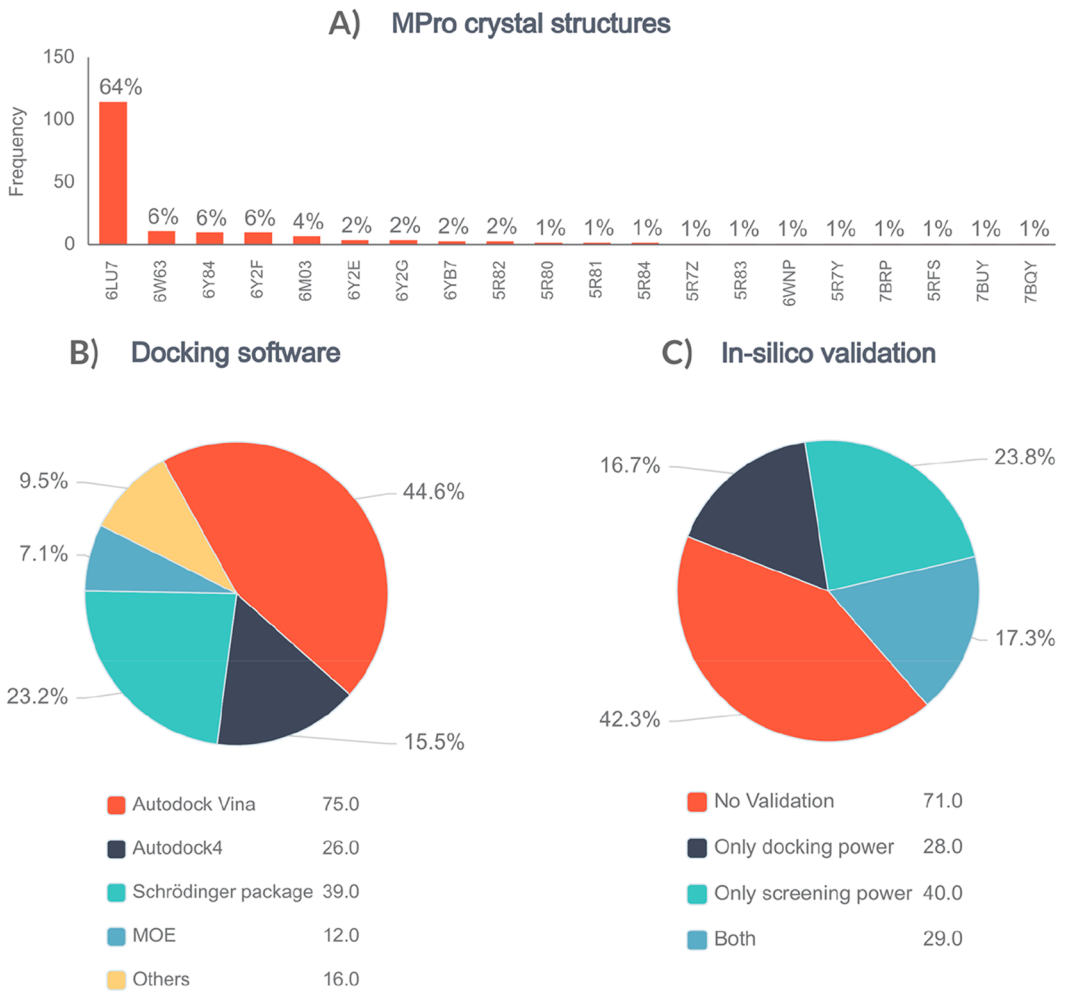


Figure 2. (A) PDB IDs of the Mpro crystal structures, sorted by the frequency of use in the analyzed investigations. (B) Percentage distribution of the docking programs used. Autodock Vina includes QuickVina, Smina, and YASARA. (C) Percentage distribution of the validated and nonvalidated screening reports.

other important binding subsites near the cleavage region of the substrate named S5 to S1 and S1′, corresponding to the positions of its P5 to P1 and P1′ residues, respectively (the cleavage point of the substrate is the bond between amino acids P1 and P1′).[20]

Regarding the discovery of inhibitors, some structure-based strategies targeting the dimer interface were explored to find dimerization inhibitors,[21] but most approaches involved the search of small molecules that interact with the previously described substrate-binding regions. Many structures of Mpro in its apo (no ligand-bound) or holo conformations (bound with inhibitors or molecular fragments) have been deposited in the Protein Data Bank (PDB), giving valuable experimental support for the discovery of novel modulators. Most of them were solved by X-ray diffraction, but there was one valuable exception elucidated by Kovalevsky et al. (PDB ID: 7JUN) by means of neutron difraction.[22] This is a powerful technique for unraveling the protonation states of amino acids, information that is lacking in standard X-ray experiments.[23−26] Furthermore, the diffraction data were collected at room temperature, so these protonation states were determined closer to physiological conditions than in the case of X-ray cryo-crystallographic measurements.[22]

Literature Analysis of Virtual Screening Campaigns. Considering the increasing amount of structural information available for Mpro, a widely chosen strategy to discover novel inhibitors is to perform docking-based virtual screening campaigns targeting this viral protease. During the year 2020, computational researchers around the globe carried out numerous investigations in this regard, which were translated into published articles. We analyzed their research by means of a literature survey, focused on original articles indexed in the PubMed and Scopus databases[27,28] until August 30th, 2020. We used the following keywords as search criteria (occurrence in the title/abstract of the publications): (docking OR *in silico* OR virtual screening) AND (Mpro OR main protease OR 3CLpro) AND (COVID OR COVID-19 OR SARS-CoV-2). After manual curation of the 194 resulting original articles, we retained 168 for further analysis, as they actually reported the use of a docking-based virtual screen to find potential inhibitors of Mpro of SARS-CoV-2. The list of articles is provided as Supporting Information (Table S1). Figure 2 shows the main information compiled after our survey.

Mpro Structures. We found that most docking protocols (96.4%) used experimental structures of SARS-CoV-2 Mpro as targets (instead of homology models), thanks to the early availability of crystallographic data. The coordinates were

https://doi.org/10.1021/acs.jcim.1c00404
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

taken from PDB entries (Figure 2A), and the main structure used was the one elucidated by Yang et al. (PDB ID: 6LU7).[29] It was the first 3D structure of the enzyme published, and it includes a peptide-like inhibitor (N3) in its binding site, covalently bound to Cys145 (Figure 1). It is worth mentioning that some *in silico* screens were conducted on previously elucidated structures of SARS-CoV (PDB IDs: 2GTB, 1Q2W, 5N5O, and 4MDS), which have a high sequence identity (96%) with the SARS-CoV-2 protease.[30]

Docking Software and Performance Evaluation. Regarding docking software, the most popular was Autodock Vina, used in 44.6% of the articles (including QuickVina, Smina, and YASARA; Figure 2B). Autodock Vina is a widely known docking program, and it is freely available for noncommercial use.[31] Software like the Schrödinger package (Glide; 23.2%), Autodock4 (15.5%), and the Molecular Operating Environment platform (MOE; 7.1%) were also chosen in several campaigns.[32−36]

The performance of the programs on the Mpro system was examined prior to virtual screening in 57.7% of the analyzed investigations (Figure 2C). Among them, 29.9% tested both the pose prediction and the virtual screening accuracy, while the rest verified only one of them (28.9% pose prediction and 41.2% virtual screening accuracy). It is worth noting here that the existence of a virtual screening accuracy analysis was counted not only in the investigations where a test set was built but also in those where the comparison of the score values was carried out in some way. The literature analysis reveals that among the cases with virtual screening accuracy calculations, in 95.6% of the investigations the authors use the score of a known active compound (or a small set of active structures) as a reference value to select the hits from the virtual screening. And surprisingly, only the remaining 4.4% of the cases used test sets (with active and inactive/decoy/random compounds) to evaluate the performance in a retrospective screen (based on ROC curves or another related metric), prior to the prospective campaign.[37−39]

A model with adequate pose prediction provides higher confidence about the predicted binding mode of the hits, which can be helpful to discard false positives by visual inspection, as well as to aid the potency optimization of the hits after confirming their biological activity. On the other hand, good virtual screening accuracy supports the selection of top-ranked compounds as potential hits to be experimentally evaluated. Both types of evaluations can aid in the selection of the software and docking conditions to find true hits in virtual screening campaigns and should be indispensable when applying these kinds of methodologies for drug discovery. Prior benchmarking studies on the performance of docking programs across a diversity of modeled systems show that, while the ligand binding poses that reproduce the experimental conformations can be identified in most cases, the ranks of the binding affinities are usually not well represented by the rank of scores, which vary substantially across protein families.[40−42] This reaffirms the utter importance of validating not only the pose prediction but also the virtual screening accuracy if molecular docking is applied as a part of a virtual screening workflow, when possible.

Chemical Databases. In addition to examining what software and conditions the researchers used to find active compounds, we analyzed what universe of structures they considered for this task. The 41.1% of the screens were performed on data sets of approved drugs, like ZINC-FDA, superDRUG2, or DrugBank databases (in some cases drugs in clinical trials were also included).[43−46] The strategy of finding new medical uses for existing drugs is known as drug repurposing and offers advantages in terms of cost and speed of the drug development pipeline.[47] Repurposed drugs have proven safety in the context of their original indication, so they are less likely to fail in the initial stages of the clinical trials (for the new therapy) than *de novo* drugs.[48] These advantages explain why the repurposing approach has been extensively applied in COVID-19-related projects, given the urgent need for therapeutic solutions.[3] So far, Remdesivir is the only example of a repositioned drug approved to treat COVID-19, though other drugs, including repurposed ones, have been approved for emergency use.[3,49,50]

Natural products are another commonly explored source of COVID-19 drug candidates, as observed in 39.9% of the investigations. Researchers looked for hits in focused libraries of molecules occurring in specific extracts, as well as in larger sets of natural products, such as the Traditional Chinese Medicine data set; the Indian Medicinal Plants, Phytochemistry And Therapeutics (IMPPAT) database; the Marine Natural Products database; the Selleck Natural Products library; or the Natural Products Atlas database.[51−55] The search of active compounds of natural origin provides access to diverse and complex scaffolds, which are sometimes challenging for scaling up to massive production. However, once a hit is confirmed, simpler related structures can be proposed to maintain (or increase) the biological activity. One suitable example of the simplification strategy is chloroquine (a drug primarily used to treat malaria), which resulted from the structural analysis of the quinine molecule isolated from the extracts of the Cinchona tree.[56]

According to our literature analysis, known drugs and natural products are the most frequent chemical space explored to find inhibitors of Mpro. We have grouped the remaining screened data sets under the term "General," which includes large libraries with commercially available chemical compounds such as ZINC, PubChem, ChEMBL, and Molport, as well as in-house compiled libraries.[43,57−59] It is worth mentioning that in several investigations additional criteria were applied before the docking simulations, such as ADMETox filters, similarity searches, and other ligand-based methodologies.[60−66] By doing so, the computational cost to perform a target-based virtual screening can be drastically reduced, either by filtering out compounds that are not expected to display an adequate pharmacokinetic profile or by exclusively docking structures that are more likely to be active according to independent (and cost-efficient) models.

Molecular Dynamics Simulations and Rescoring Functions. Molecular dynamics (MD) is a computational tool used to simulate the behavior of a system as a function of time.[67−69] In the context of drug discovery, it provides information about the protein−ligand interactions at the atomic level and includes flexibility features for the whole system, which are closely connected to the current models of the binding event (such as the induced fit and conformational selection theories).[5] MD can be used in conjunction with docking simulations as sampling engines, giving access to multiple conformations of the target.[9,70]

In addition to the representation of protein flexibility, one of the major limitations of docking-based methodologies continues to be the scoring function.[9,70] The docking scores assigned to each compound are particularly important in

https://doi.org/10.1021/acs.jcim.1c00404
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

virtual screening campaigns, since it is the numerical variable used to distinguish binders from nonbinders in the screened library. The quantification of solvent-mediated effects and entropic contributions by the scoring functions are two of the current challenges in docking simulations. MD-based binding free energy calculations can greatly improve the accuracy of the predictions, since these methods can partially account for entropy and solvation effects in the calculations.[9] Another possible strategy is to apply machine learning-based methods and/or entropy corrections, which can enhance hit recovery in structure-based virtual screening.[71−74]

Back to SARS-CoV-2 Mpro virtual screening, we observed that about half of the analyzed investigations invested efforts in MD simulations to generate different conformations of the target, to understand the binding modes of the candidates or to refine the binding energy initially predicted by docking protocols. It is worth mentioning that in 20.2% of the cases, the authors reported the use of more than one scoring function to select the candidates in an attempt to improve the accuracy in virtual screening. In this regard, the most popular software suite was the Schrödinger package, which allows the carrying out a sequential screening based on scoring functions and search algorithms, gaining precision in each round of hit selection: HTVS (high-throughput virtual screening) mode for millions of compounds, SP (standard precision) mode for tens to hundreds of thousands ligands, and XP (extra precision) mode for extensive sampling and scoring of a small set of structures.[32−34]

During the redaction of this paper, an interesting study was published, which incorporated target flexibility in docking-based virtual screening on SARS-CoV-2 targets.[75] Acharya and co-workers applied an ensemble docking protocol for the virtual screening campaigns and docked the compounds on multiple receptor conformations as a way to incorporate target flexibility. They employed temperature replica-exchange molecular dynamics (T-REMD) to generate the conformations. In this method, multiple copies of the target at different temperatures are simulated, with periodic exchanging of the coordinates between the copies. Even though the authors expressed that this is a preliminary publication, they already noted that the T-REMD method was able to generate a diverse ensemble of conformations (especially in the loop regions) for the 24 systems tested (where Mpro was included in different protonation states and in its monomereric or dimeric form). Additionally, their docking experiments with Vina, MOE, and Autodock-GPU on Mpro retrieved known active compounds with enrichment rates ranging from 7% to 14%, which are higher than the rates obtained experimentally (5.7%).[75] Undoubtedly, valuable conclusions would emerge in future papers of these authors with the complete analysis of the data collected. Another approach in this regard is DINC−COVID, a user-friendly docking interface developed by Hall-Swan et al. published as a preprint in January 2021.[76] This computational tool allows ensemble docking with three SARS-CoV-2 proteins, including the Mpro. Three distinct ensembles for each SARS-CoV-2 protein are available, including data from crystal structures or from MD simulations with different force fields (CHARMM or GROMOS). Moreover, all generated binding modes are rescored and ranked using three scoring functions (i.e., Vina, Vinardo, and AutoDock4).

CONSENSUS SCORING

Our literature survey also yielded investigations with approaches that jointly apply more than one scoring function to rank the compounds and/or include energy minimization stages of the docking complexes to improve binding energy estimation.[37,77] The combination of multiple scoring functions for binding affinity estimation or hit selection in virtual screening campaigns is known as consensus scoring.[78,79] Classical consensus approaches select compounds in the intersection of the set of best scores for each docking program; even more complex combinations of screening scores are also popular.[80,81] Many studies support the idea that combining results from different docking programs to get a final rank or score for each molecule leads to a higher success rate in virtual screening campaigns,[82−86] and this concept was explored by the authors for the Mpro target, even though no comparison was reported between the performance of single docking scores and the combination of multiple protocols.[37,77]

Experimental Confirmation. Undoubtedly, the biochemical and biological evaluation of the hits represents a fundamental stage in the virtual screening process. In addition to the obvious outcome of finding active compounds, it provides valuable feedback to improve the docking models by including more data. Surprisingly, most of the hits proposed in the revised investigations are yet to be confirmed experimentally. Only two (1.2%) publications mentioned results of biological tests, and in one of them the experimental results were not disclosed (the authors refer to them as unpublished data).[87] The other investigation was conducted by Vatansever et al., who focused the virtual search of Mpro inhibitors on a selected group of FDA/EMA-approved small molecules.[88] They found six known drugs with IC50 values below 100 μM against the protease, and one of them, Bepridil, showed a complete inhibition of the cytopathogenic effects in Vero E6 cells induced by the SARS-CoV-2 virus.[71] As mentioned before, finding hits is the main purpose of the screening. However, a proper validation of the methods requires a statistical analysis to estimate the minimum sample size required for a given effect size, significance level, and statistical power, before drawing any conclusion.

Docking Protocols Performance Evaluation. The analysis of the published data reveals that it is not yet possible to know the capabilities of target-based virtual screening in the Mpro system due to the absence of experimental confirmation about the predicted inhibitory activity of the hits in most of the investigations. Therefore, we conducted a retrospective analysis of three docking protocols using the available structural information (for the determination of the pose prediction) and the reported biological data on Mpro inhibition assays (for the assessment of the virtual screening accuracy). These validations allowed us to gain more information about the strengths and weaknesses of the protocols on this particular system.

Methodology. All structures of SARS-CoV-2 Mpro released until October 2020 were analyzed and manually curated. The data were retrieved from https://covid-19.bioreproducibility.org/, a database of carefully validated COVID-19 related structures.[89] When the experimental structure had been partially or fully refined, we used the corrected structure; otherwise the original one was downloaded from the RCSB PDB database. Experimental complexes with inhibitors covalently bound to the catalytic Cys145 were

https://doi.org/10.1021/acs.jcim.1c00404
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

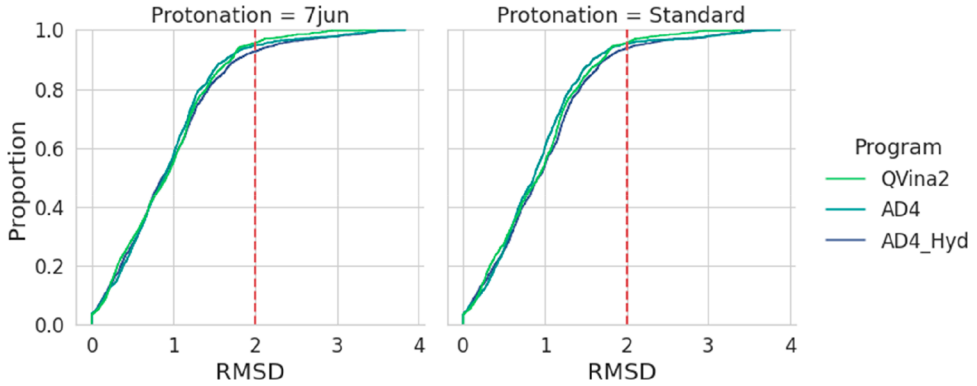


Figure 3. Pose prediction evaluation. Cumulative distribution plots of RMSD values obtained from redocking and cross-docking experiments for both protonation schemes, grouped by docking program.

not included as the docking protocols tested were constructed based on noncovalent interactions. Structures with missing segments, mutations, or modified residues were also discarded. After splitting multimeric complexes into individual chains, we obtained 52 curated structures: 23 in the apo and 29 in the holo form. Only those complexes with ligands bound to the catalytic site were considered.

Missing atoms were repaired with Modeller,[90] and hydrogens were added with Reduce software in the presence of water molecules and ligands, allowing rotations and flips for NQH residues.[91] All structures were superimposed using Modeller's salign function and converted to pdbqt format using the prepare_receptor4 script distributed in AutodockTools.[35,90] Ligand preparation included protonation at pH 7.4 using OpenBabel, conversion to pdbqt with the prepare_ligand4 script, and coordinates randomization before docking simulations.[92]

Three different docking protocols were evaluated, based on two docking programs: Quick Vina 2 (QVina2), a faster version of the classical Autodock Vina employing the same scoring function,[93] and Autodock4-GPU (AD4-GPU), the GPU accelerated version of Autodock4 software.[94] For Autodock4, a modified protocol was also tested, in which discrete displaceable waters are placed around the ligand before docking simulations, aiming to predict the position of waters mediating the ligand interactions with the target, i.e., hydrated docking (AD4-Hyd).[95]

In all protocols, the docking site was defined as a cubic box of 20 × 20 × 20 Å enclosing all crystallized ligands. For QVina2, the exhaustiveness parameter was set to 32, and three poses per ligand were generated. Regarding AD4-GPU, affinity maps were calculated over the same box using Autogrid with the default spacing of 0.375 Å. The number of energy evaluations and the local-search algorithm was set on-the-fly for each ligand based on a built-in heuristic, and the automatic stop criterion based on energy convergence was turned on. A total of 200 docking runs for each ligand were carried out. All other parameters were set to default values.

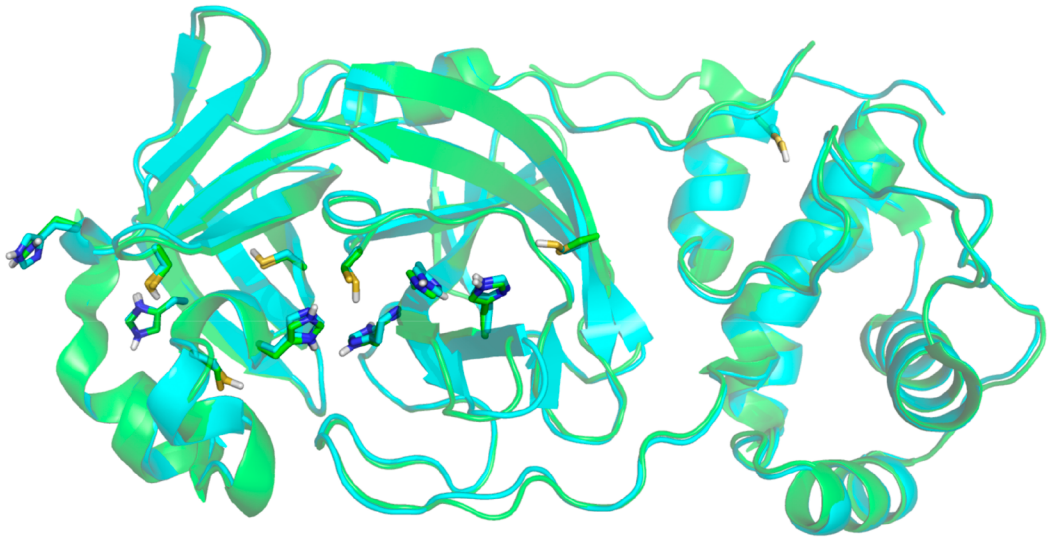
As a postprocessing approach, complexes generated with QVina2 were rescored with three different scoring functions implemented in Smina software.[96] The rescoring was preceded by a minimization stage considering residue side chains within 3 Å of the docked ligand as flexible. We have previously used this postprocessing strategy to improve the virtual screening accuracy of docking models, with only a small increase in computational cost.[97]

For the pose prediction assessment, the symmetry-corrected RMSD (Å) between predicted and experimental poses was calculated with RDKit.[98] For the virtual screening accuracy, a data set of compounds evaluated against the SARS-CoV-2 Mpro was compiled from the literature. Compounds with IC50 > 10 μM were classified as active, while those with IC50 > 20 μM, or %INH (20 μM) < 80%, or %INH (10 μM) < 50% were considered inactive. Covalent inhibitors were not included. A total of 816 compounds (53 active and 763 inactive, Table S2) were retrieved. Because docking this whole data set on every Mpro crystal, under each docking protocol and protonation scheme, would imply a huge computational cost, we created a smaller core data set including only those compounds published in peer-reviewed articles and excluding preprints. The refined data set comprises nine active and 52 inactive compounds (Table S3). The virtual screening accuracy evaluation first involved docking the core data set on every protein structure, with all protonation schemes using the three docking protocols described before. Then, the best performing combinations were re-evaluated using the whole data set to better estimate the screening performance. The AUCROC was used to measure the performance of the models as binary classifiers.

Pose Predictions. The results of redocking and cross-docking experiments are shown in Figure 3, and the data plotted are provided as Supporting Information. After docking all crystallized ligands on all protein structures, the RMSD to the experimental binding mode was calculated and grouped by structure and by docking protocol. In most combinations, the RMSD values are lower than 2 Å, which provides confidence about the ability of docking programs to reproduce the experimental binding conformation. Regarding the docking programs, we calculated the mean RMSD for each docking protocol, considering the whole set of crystal structures, and we found that AD4_GPU (0.955 ± 0.658 Å) and AD4_hyd (0.970 ± 0.655 Å) outperform QVina2 (1.575 ± 1.203 Å) in terms of pose prediction.

It is worth mentioning that among docked ligands there are some small fragments with few rotatable bonds, so the identification of the correct binding mode becomes less demanding for the search algorithms than for inhibitors with more degrees of conformational freedom. Additionally, these X-ray diffraction data sets were all phased with the deposited structure of Mpro with the N3 inhibitor (PDB ID: 6LU7), which fixes the final conformation of the protein in the complexes with the fragments and, therefore, influences the

https://doi.org/10.1021/acs.jcim.1c00404
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX



	16	22	38	41	44	64	80	85	117	128	145	156	160	163	164	172	246	265	300
7JUN	CYS	CYM	CYM	HIP	CYM	HIP	HIP	CYS	CYS	CYM	CYM	CYS	CYS	HID	HIP	HIE	HIE	CYS	CYS
6LU7	CYS	CYS	CYS	HID	CYS	HIE	HID	CYS	CYS	CYS	CYS	CYS	CYS	HIE	HIE	HIE	HIE	CYS	CYS

Figure 4. Comparison between the protonation scheme of His and Cys residues in the neutron diffraction structure 7JUN (green) and the X-ray structure 6LU7 protonated with Reduce software (cyan). The latter is the most frequently chosen structure for docking simulations according to our literature survey. All His and Cys residues in the structure are listed in the bottom table, and those diverging between the structures are highlighted as sticks in the superimposed structures. CYS refers to a standard protonated cysteine residue, while CYM refers to the deprotonated anionic form. For histidines, HID refers to a delta-protonated imidazole ring, HIE is the epsilon-protonated, and HIP indicates a fully protonated positively charged histidine. Only polar hydrogens are shown.

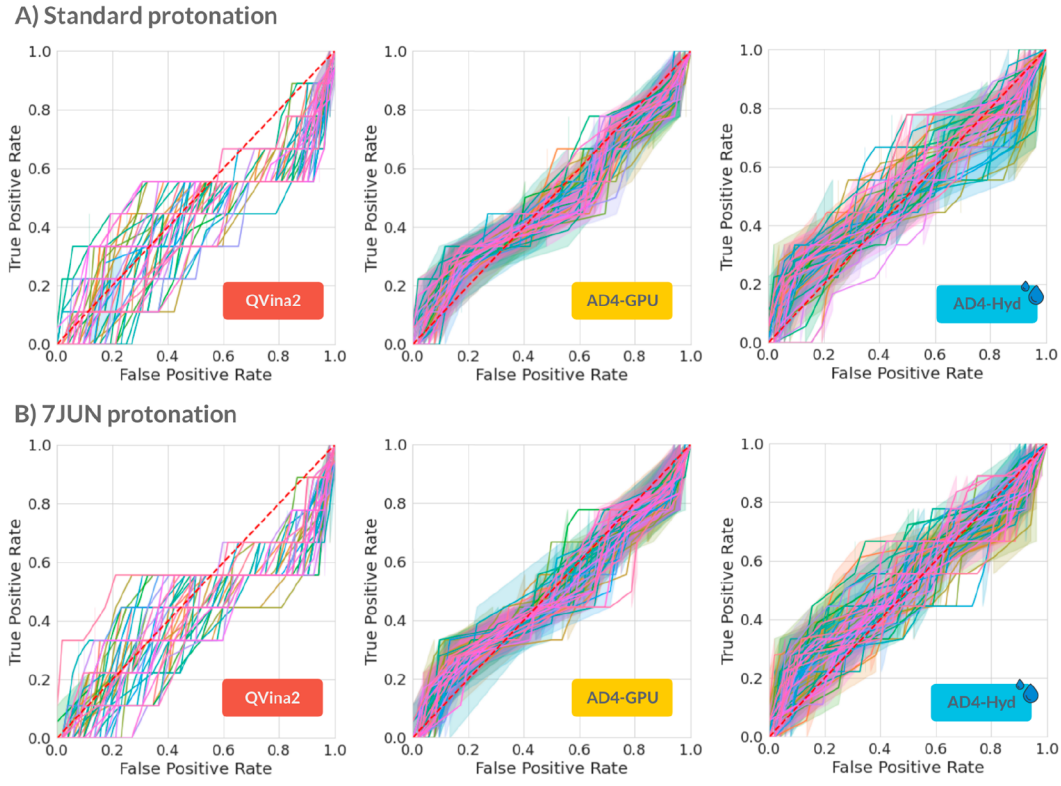


Figure 5. ROC curves obtained after docking the core data set on all Mpro structures, grouped by docking protocol. Each curve was drawn from a different structure, and the 95% confidence interval obtained by bootstrapping is shown as a light shade. (A) Structures protonated by Reduce with default parameters. (B) Structures reprotonated according to the neutron diffraction crystal structure 7JUN.

cross-docking simulation results.[99] The protonation state of titratable residues is a critical choice to make before performing any kind of biomolecular simulation. As mentioned before, the low electron density of hydrogen in X-ray

https://doi.org/10.1021/acs.jcim.1c00404
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

crystallography experiments hampers the direct determination of protonation states. The recently solved neutron diffraction structure of Mpro (PDB ID: 7JUN) brings valuable information in this regard, such as the protonation state of histidine and cysteine residues.[100] However, these protonation states are quite particular and may differ from those automatically assigned by standard bioinformatic software.[100] As shown in Figure 4, most of these differences lie in the vicinity of the catalytic site. For example, the program Reduce protonated the sulfur atom of Cys145 and showed His41 as HID, whereas in the experimental structure 7JUN the thiol function is deprotonated and His41 is shown in the HIP form (fully protonated). Hence, in order to investigate the effect of different His and Cys protomers on docking performance, we changed the protonation state of these residues in all structures to match those in the experimental neutron diffraction model.

Figure 3 shows the results obtained with the standard protonation assigned by Reduce software and the reprotonated structures according to the neutron diffraction crystal structure 7JUN. The results showed a similar good performance in both systems, possibly because, as they are noncovalent inhibitors, the most relevant interactions occur in the region of the active site but not necessarily with the catalytic dyad.

The protonation states of the catalytic dyad of SARS-CoV-2 Mpro as well other important His residues in its close proximity (such as His163, His164, and His172) have been studied due to their influence on the conformational stability of the apo and inhibitor-bound proteins. Pavlova and co-workers used MD simulations to analyze the stability of 12 viable protonation states of the protein in the apo and holo forms (the last one with two different inhibitors). They underlined the complexity of the system and concluded that the most structurally stable protonation states vary in a ligand-dependent manner.[100] During the preparation of this manuscript, Kneller and co-workers solved another neutron diffraction structure of Mpro in complex with the α-ketoamide inhibitor telaprevir.[101] By direct comparison with the previously solved unliganded structure 7JUN, they found that inhibitor binding to the Mpro induced changes in the protonation states of key histidine residues in the active site (His41, His163, and His164).

Virtual Screening Accuracy. To test the screening accuracy of the programs, the compounds of the core data set were docked into every Mpro experimental structure. Then, they were ranked by the docking score and the AUCROCs were calculated to evaluate the classification power of the models. A value equal to 1 represents an ideal performance while an AUCROC of 0.5 means that, on average, the ranked compounds do not differ from a randomly ordered list.

As shown in Figure 5, all tested structures and docking protocols showed a poor capacity to correctly rank the compounds in the core data set according to its predicted affinity. In practice, this means that top-ranked compounds from a virtual screening campaign may not have higher chances to be true hits than randomly selected molecules.

Both protonation schemes perform similarly as evidenced by the ROC curves obtained. However, even though it seems to have a small impact on docking performance under the docking protocols and conditions tested, the protonation state of titratable residues will heavily impact the results of further MD simulations.[100] Regarding the postprocessing of the complex by flexible side-chain minimization and rescoring, we found no clear improvement in the screening performance compared to the original docking results.

For each docking protocol and protonation scheme, the best performing combination in terms of screening accuracy was re-evaluated using the whole data set of 816 compounds, in order to confirm the results with a bigger and more diverse set of structures. However, the results were equally disappointing, with AUCROC values even lower than before. These results confirm the trend observed with the core data set, discarding the possibility that the bad performance observed before was due to the small size of the test set. A remarkable limitation of the available data is that the inhibitory mechanism of the reported inhibitors is seldomly characterized. Typically, literature reports inform the activity of a given compound at a given concentration (or, at best, the IC50), but the type of inhibition (e.g., competitive, noncompetitive, uncompetitive, etc.) has been rarely investigated at the time of the report, which limits the accuracy of any retrospective screen based on such data owing to the uncertainty on the actual binding site. The screening accuracy results for all docking protocols and data sets are provided as Supporting Information, including virtual screening specific metrics such as BEDROC (alpha = 20), RIE (alpha = 20), and EF1%.

In order to evaluate the consensus between the ranks generated by different docking protocols, we counted the number of coincidences found in the top 20% of the ranking. For each docking protocol, the best performing Mpro structure in terms of AUCROC was chosen, considering the standard protonation scheme assigned by Reduce software. Ten compounds were ranked in the top 20% by at least two programs, and among them, only four could be found in all three rankings. Unsurprisingly, we found very few active molecules in this zone, as reflected by the poor enrichment metrics obtained.

FUTURE DIRECTIONS

It is important to note that while we processed the results of our literature search and wrote this paper, successful docking-based investigations including experimental evaluations were published albeit without *in silico* validation. For example, Ghahremanpour and co-workers ran independent virtual screening campaigns with four docking protocols to minimize the bias of the scoring functions (constructed to reproduce a finite set of experimental ligand-binding affinities) in a library of 2000 approved oral drugs.[102] Then, they focused on the structures that ranked among the top 10% percent in at least three out of the four runs. After visual inspection, 17 compounds were tested against Mpro, and four of them showed IC50 values below 20 μM (Manidipine, Boceprevir, Lercanidipine, and Bedaquiline). Perampanel was also identified as a weak Mpro inhibitor in this investigation, but only a rough IC50 of 100−250 μM was estimated due to interferences in the fluorescence experiment. However, this structure caught the author's attention for the design of noncovalent and nonpeptidic Mpro inhibitors. Starting from the docking poses of Perampanel, they proposed structural modifications to optimize the interactions with the target, using free-energy perturbation (FEP) calculations to guide the optimization process.[103] The computational analyses in conjunction with the experimental elucidation of multiple crystal structures of the complexes led to optimized inhibitors of Mpro with nanomolar potency. Another example is the work of Gupta and co-workers, who found three candidates for hit

https://doi.org/10.1021/acs.jcim.1c00404
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

optimization: Cobicistat, Cangrelor, and Denufosol (IC50 of ~6.7 μM, 0.9 mM, and 1.3 mM, respectively).[104] These results were obtained from docking-based virtual screens of three databases (DrugBank, ZINC-FDA, and Spec database).[104,105] The authors used the Glide module of Schrodinger as the main docking software, but other programs were also considered for the comparison of the poses. Moreover, they ran MD simulations to better estimate the binding energies.

The aforementioned examples show the successful application of target-based protocols for the identification of Mpro inhibitors, particularly if more than one docking program is used and protein flexibility is taken into account. Furthermore, we propose that the inclusion of validation tests as a selection criterion for the protocols would save time and increase the chances of success in the screening campaigns, since less accurate protocols would be discarded early.

CONCLUSIONS

A major challenge in docking simulations is to find the combination of search algorithms and scoring functions that accurately predicts the binding mode of a protein−ligand complex and correctly quantifies its binding affinity. Although it is possible to make inferences about the characteristics of different programs by analyzing how they were built and what algorithms they use, the most direct way to test their capacities is to confront them with the specific system under study. In the context of virtual screening, the virtual screening accuracy and the pose predictions are critical tests: they evaluate the capacity of the software to identify potential hits (by the docking score) and to propose explanations about the binding interactions, as well as to guide further structure optimization efforts (by the binding pose).

One of the most promising targets under study to tackle the SARS-CoV-2 virus is Mpro, an essential protease for the virus with structural characteristics not observed in human isoforms. The analysis of 168 publications associated with the search for Mpro inhibitors through docking-based virtual screening, and the evaluation of the docking performance shown here, allowed us to assess the strengths and weaknesses of the methodology in this particular system.

In most of the analyzed investigations, the authors assumed a direct relationship between top-ranked molecules by docking simulations and *in vitro* inhibitory potency. However, the retrospective validation of docking methodologies applied to the SARS-CoV-2 Mpro system shown here reveals that these assumptions should be, at least, revised.

In order to improve the predictive power of a model, it is necessary to have well-established protocols and robust metrics to measure it. There are many factors that can influence the performance of a docking model, and many alternatives which can be explored to overcome its limitations. Some of them were explored in this article, like considering alternative protonation states for critical residues, the flexible side-chain minimization, consensus scoring, rescoring of the docking complexes with different scoring functions, or the inclusion of solvation effects through specific protocols like Autodock hydrated.

Complementary approaches, like MD simulations, can be explored in order to refine the results from docking simulations and discard potential false positives. Moreover, MD provides valuable information about ligand−receptor interactions explicitly considering solvent-mediated effects, which can be exploited during the potency optimization of a hit.

The good pose prediction of all docking protocols tested supports its application in virtual screening campaigns, to generate initial hypotheses about the binding mode of the screened compounds. However, the binding affinity predicted for each pose may not be reliable, as suggested by the poor screening accuracy obtained, so additional criteria should be considered for the final selection of the hits. Finally, the experimental evaluation of the candidates must be included in the investigations as a final selection criterion, to advance in the definition of the optimal requirements for the activity at the molecular level and to improve the computational models with reliable information.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.1c00404.

The articles analyzed in the literature survey are listed in Table S1; the compounds evaluated against the SARS-CoV-2 Mpro compiled from the literature are listed in Table S2; and the compounds that comprise the core data set are listed in Table S3 (PDF)

Results of all redocking and cross-docking experiments (XLSX)

Screening accuracy results for all docking protocols and data sets (XLSX)

AUTHOR INFORMATION

Corresponding Author

Luciana Gavernet − *Laboratory of Bioactive Research and Development (LIDeB), Department of Biological Sciences, Faculty of Exact Sciences, National University of La Plata (UNLP), Buenos Aires, Argentina;* orcid.org/0000-0001-9285-3788; Email: lgavernet@gmail.com

Authors

Manuel A. Llanos − *Laboratory of Bioactive Research and Development (LIDeB), Department of Biological Sciences, Faculty of Exact Sciences, National University of La Plata (UNLP), Buenos Aires, Argentina*

Melisa E. Gantner − *Laboratory of Bioactive Research and Development (LIDeB), Department of Biological Sciences, Faculty of Exact Sciences, National University of La Plata (UNLP), Buenos Aires, Argentina;* orcid.org/0000-0001-7491-4268

Santiago Rodriguez − *Laboratory of Bioactive Research and Development (LIDeB), Department of Biological Sciences, Faculty of Exact Sciences, National University of La Plata (UNLP), Buenos Aires, Argentina*

Lucas N. Alberca − *Laboratory of Bioactive Research and Development (LIDeB), Department of Biological Sciences, Faculty of Exact Sciences, National University of La Plata (UNLP), Buenos Aires, Argentina*

Carolina L. Bellera − *Laboratory of Bioactive Research and Development (LIDeB), Department of Biological Sciences, Faculty of Exact Sciences, National University of La Plata (UNLP), Buenos Aires, Argentina;* orcid.org/0000-0002-1237-4929

Alan Talevi − *Laboratory of Bioactive Research and Development (LIDeB), Department of Biological Sciences, Faculty of Exact Sciences, National University of La Plata*

https://doi.org/10.1021/acs.jcim.1c00404
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

(UNLP), Buenos Aires, Argentina; orcid.org/0000-0003-3178-826X

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.1c00404

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors are funded by UNLP (Incentivos UNLP) and Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT), via grants PICT 2017-0643 and PICT 2016-0165. This work was also supported by the URGENCE nouveau coronavirus fundraising campaign of the Institut Pasteur. The authors gratefully acknowledge the support of NVIDIA Corporation with the donation of the Quadro M5000 and the Titan Xp GPUs used for this research.

■ 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.; Zhang, Y.-Z. A New Coronavirus Associated with Human Respiratory Disease in China. *Nature* **2020**, *579*, 265−269.

(2) Osman, E. E. A.; Toogood, P. L.; Neamati, N. COVID-19: Living through Another Pandemic. *ACS Infect. Dis.* **2020**, *6*, 1548−1552.

(3) Bellera, C. L.; Llanos, M.; Gantner, M. E.; Rodriguez, S.; Gavernet, L.; Comini, M.; Talevi, A. Can Drug Repurposing Strategies Be the Solution to the COVID-19 Crisis? *Expert Opin. Drug Discovery* **2021**, *16*, 605−612.

(4) Sousa, S. F.; Ribeiro, A. J. M.; Coimbra, J. T. S.; Neves, R. P. P.; Martins, S. A.; Moorthy, N. S. H. N.; Fernandes, P. A.; Ramos, M. J. Protein-Ligand Docking in the New Millennium - A Retrospective of 10 Years in the Field. *Curr. Med. Chem.* **2013**, *20*, 2296−2314.

(5) Salmaso, V.; Moro, S. Bridging Molecular Docking to Molecular Dynamics in Exploring Ligand-Protein Recognition Process: An Overview. *Front. Pharmacol.* **2018**, *9*, 923.

(6) Repasky, M. P.; Shelley, M.; Friesner, R. A. Flexible Ligand Docking with Glide. *Curr. Protoc. Bioinforma.* **2007**, Chapter 8, Unit 8.12. DOI: 10.1002/0471250953.bi0812s18.

(7) Lemmon, G.; Meiler, J. Rosetta Ligand Docking with Flexible XML Protocols. In *Computational Drug Discovery and Design*; Baron, R., Ed.; Methods in Molecular Biology; Springer: New York, 2012; pp 143−155. DOI: 10.1007/978-1-61779-465-0_10.

(8) B-Rao, C.; Subramanian, J.; Sharma, S. D. Managing Protein Flexibility in Docking and Its Applications. *Drug Discovery Today* **2009**, *14*, 394−400.

(9) Guterres, H.; Im, W. Improving Protein-Ligand Docking Results with High-Throughput Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2020**, *60*, 2189−2198.

(10) Khanjiwala, Z.; Khale, A.; Prabhu, A. Docking Structurally Similar Analogues: Dealing with the False-Positive. *J. Mol. Graphics Modell.* **2019**, *93*, 107451.

(11) Adeshina, Y. O.; Deeds, E. J.; Karanicolas, J. Machine Learning Classification Can Reduce False Positives in Structure-Based Virtual Screening. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117*, 18477−18488.

(12) Gushchina, I. V.; Polenova, A. M.; Suplatov, D. A.; Švedas, V. K.; Nilov, D. K. VsFilt: A Tool to Improve Virtual Screening by Structural Filtration of Docking Poses. *J. Chem. Inf. Model.* **2020**, *60*, 3692−3696.

(13) Ye, W.-L.; Shen, C.; Xiong, G.-L.; Ding, J.-J.; Lu, A.-P.; Hou, T.-J.; Cao, D.-S. Improving Docking-Based Virtual Screening Ability by Integrating Multiple Energy Auxiliary Terms from Molecular Docking Scoring. *J. Chem. Inf. Model.* **2020**, *60*, 4216−4230.

(14) Cross, J. B.; Thompson, D. C.; Rai, B. K.; Baber, J. C.; Fan, K. Y.; Hu, Y.; Humblet, C. Comparison of Several Molecular Docking Programs: Pose Prediction and Virtual Screening Accuracy. *J. Chem. Inf. Model.* **2009**, *49*, 1455−1474.

(15) Hevener, K. E.; Zhao, W.; Ball, D. M.; Babaoglu, K.; Qi, J.; White, S. W.; Lee, R. E. Validation of Molecular Docking Programs for Virtual Screening against Dihydropteroate Synthase. *J. Chem. Inf. Model.* **2009**, *49*, 444−460.

(16) Wang, C.; Zhang, Y. Improving Scoring-Docking-Screening Powers of Protein-Ligand Scoring Functions Using Random Forest. *J. Comput. Chem.* **2017**, *38*, 169−177.

(17) Li, Y.; Su, M.; Liu, Z.; Li, J.; Liu, J.; Han, L.; Wang, R. Assessing Protein-Ligand Interaction Scoring Functions with the CASF-2013 Benchmark. *Nat. Protoc.* **2018**, *13*, 666−680.

(18) Rizzi, A.; Fioni, A. Virtual Screening Using PLS Discriminant Analysis and ROC Curve Approach: An Application Study on PDE4 Inhibitors. *J. Chem. Inf. Model.* **2008**, *48*, 1686−1692.

(19) Ullrich, S.; Nitsche, C. The SARS-CoV-2 Main Protease as Drug Target. *Bioorg. Med. Chem. Lett.* **2020**, *30*, 127377.

(20) Zhang, L.; Lin, D.; Sun, X.; Curth, U.; Drosten, C.; Sauerhering, L.; Becker, S.; Rox, K.; Hilgenfeld, R. Crystal Structure of SARS-CoV-2 Main Protease Provides a Basis for Design of Improved α-Ketoamide Inhibitors. *Science* **2020**, *368*, 409−412.

(21) Goyal, B.; Goyal, D. Targeting the Dimerization of the Main Protease of Coronaviruses: A Potential Broad-Spectrum Therapeutic Strategy. *ACS Comb. Sci.* **2020**, *22*, 297−305.

(22) Kneller, D. W.; Phillips, G.; Weiss, K. L.; Pant, S.; Zhang, Q.; O'Neill, H. M.; Coates, L.; Kovalevsky, A. Unusual Zwitterionic Catalytic Site of SARS-CoV-2 Main Protease Revealed by Neutron Crystallography. *J. Biol. Chem.* **2020**, *295*, 17365−17373.

(23) Bax, B.; Chung, C.; Edge, C. Getting the Chemistry Right: Protonation, Tautomers and the Importance of H Atoms in Biological Chemistry. *Acta Crystallogr. Sect. Struct. Biol.* **2017**, *73*, 131−140.

(24) Gardberg, A. S.; Del Castillo, A. R.; Weiss, K. L.; Meilleur, F.; Blakeley, M. P.; Myles, D. A. A. Unambiguous Determination of H-Atom Positions: Comparing Results from Neutron and High-Resolution X-Ray Crystallography. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 558−567.

(25) Golden, E. A.; Vrielink, A. Looking for Hydrogen Atoms: Neutron Crystallography Provides Novel Insights Into Protein Structure and Function. *Aust. J. Chem.* **2014**, *67*, 1751−1762.

(26) Oksanen, E.; Chen, J. C.-H.; Fisher, S. Z. Neutron Crystallography for the Study of Hydrogen Bonds in Macromolecules. *Molecules* **2017**, *22*, 596.

(27) Agarwala, R.; Barrett, T.; Beck, J.; Benson, D. A; Bollin, C.; Bolton, E.; Bourexis, D.; Brister, J R.; Bryant, S. H; Canese, K.; Cavanaugh, M.; Charowhas, C.; Clark, K.; Dondoshansky, I.; Feolo, M.; Fitzpatrick, L.; Funk, K.; Geer, L. Y; Gorelenkov, V.; Graeff, A.; Hlavina, W.; Holmes, B.; Johnson, M.; Kattman, B.; Khotomlianski, V.; Kimchi, A.; Kimelman, M.; Kimura, M.; Kitts, P.; Klimke, W.; Kotliarov, A.; Krasnov, S.; Kuznetsov, A.; Landrum, M. J; Landsman, D.; Lathrop, S.; Lee, J. M; Leubsdorf, C.; Lu, Z.; Madden, T. L; Marchler-Bauer, A.; Malheiro, A.; Meric, P.; Karsch-Mizrachi, I.; Mnev, A.; Murphy, T.; Orris, R.; Ostell, J.; O'Sullivan, C.; Palanigobu, V.; Panchenko, A. R; Phan, L.; Pierov, B.; Pruitt, K. D; Rodarmer, K.; Sayers, E. W; Schneider, V.; Schoch, C. L; Schuler, G. D; Sherry, S. T; Siyan, K.; Soboleva, A.; Soussov, V.; Starchenko, G.; Tatusova, T. A; Thibaud-Nissen, F.; Todorov, K.; Trawick, B. W; Vakatov, D.; Ward, M.; Yaschenko, E.; Zasypkin, A.; Zbicz, K. Database Resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **2018**, *46*, D8−D13.

(28) Burnham, J. F. Scopus Database: A Review. *Biomed. Digit. Libr.* **2006**, *3*, 1.

(29) 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.; Yang, H. Structure of M pro from SARS-CoV-2 and Discovery of Its Inhibitors. *Nature* **2020**, *582*, 289−293.

(30) Yoshino, R.; Yasuo, N.; Sekijima, M. Identification of Key Interactions between SARS-CoV-2 Main Protease and Inhibitor Drug Candidates. *Sci. Rep.* **2020**, *10*, 12493.
(31) Trott, O.; Olson, A. J. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J. Comput. Chem.* **2009**, *31*, 455−461.
(32) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes. *J. Med. Chem.* **2006**, *49*, 6177−6196.
(33) 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.; Shenkin, P. S. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004**, *47*, 1739−1749.
(34) 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.
(35) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J. Comput. Chem.* **2009**, *30*, 2785−2791.
(36) Chemical Computing Group ULC. Molecular Operating Environment (MOE) | MOEsaic | PSILO.https://www.chemcomp.com/Products.htm (accessed Dec 26, 2020).
(37) Gimeno, A.; Mestres-Truyol, J.; Ojeda-Montes, M. J.; Macip, G.; Saldivar-Espinoza, B.; Cereto-Massagué, A.; Pujadas, G.; Garcia-Vallvé, S. Prediction of Novel Inhibitors of the Main Protease (M-pro) of SARS-CoV-2 through Consensus Docking and Drug Reposition. *Int. J. Mol. Sci.* **2020**, *21*, 3793.
(38) Ton, A.-T.; Gentile, F.; Hsing, M.; Ban, F.; Cherkasov, A. Rapid Identification of Potential Inhibitors of SARS-CoV-2 Main Protease by Deep Docking of 1.3 Billion Compounds. *Mol. Inf.* **2020**, *39*, 2000028.
(39) Choudhury, C. Fragment Tailoring Strategy to Design Novel Chemical Entities as Potential Binders of Novel Corona Virus Main Protease. *J. Biomol. Struct. Dyn.* **2021**, *39*, 3733−3746.
(40) Wang, Z.; Sun, H.; Yao, X.; Li, D.; Xu, L.; Li, Y.; Tian, S.; Hou, T. Comprehensive Evaluation of Ten Docking Programs on a Diverse Set of Protein-Ligand Complexes: The Prediction Accuracy of Sampling Power and Scoring Power. *Phys. Chem. Chem. Phys.* **2016**, *18*, 12964−12975.
(41) Li, Y.; Han, L.; Liu, Z.; Wang, R. Comparative Assessment of Scoring Functions on an Updated Benchmark: 2. Evaluation Methods and General Results. *J. Chem. Inf. Model.* **2014**, *54*, 1717−1736.
(42) Plewczynski, D.; Łaźniewski, M.; Augustyniak, R.; Ginalski, K. Can We Trust Docking Results? Evaluation of Seven Commonly Used Programs on PDBbind Database. *J. Comput. Chem.* **2011**, *32*, 742−755.
(43) Sterling, T.; Irwin, J. J. ZINC 15 - Ligand Discovery for Everyone. *J. Chem. Inf. Model.* **2015**, *55*, 2324−2337.
(44) Siramshetty, V. B.; Eckert, O. A.; Gohlke, B.-O.; Goede, A.; Chen, Q.; Devarakonda, P.; Preissner, S.; Preissner, R. SuperDRUG2: A One Stop Resource for Approved/Marketed Drugs. *Nucleic Acids Res.* **2018**, *46*, D1137−D1143.
(45) Goede, A.; Dunkel, M.; Mester, N.; Frommel, C.; Preissner, R. SuperDrug: A Conformational Drug Database. *Bioinformatics* **2005**, *21*, 1751−1753.
(46) Wishart, D. S.; Feunang, Y. D.; Guo, A. C.; Lo, E. J.; Marcu, A.; Grant, J. R.; Sajed, T.; Johnson, D.; Li, C.; Sayeeda, Z.; Assempour, N.; Iynkkaran, I.; Liu, Y.; Maciejewski, A.; Gale, N.; Wilson, A.; Chin, L.; Cummings, R.; Le, D.; Pon, A.; Knox, C.; Wilson, M. DrugBank 5.0: A Major Update to the DrugBank Database for 2018. *Nucleic Acids Res.* **2018**, *46*, D1074−D1082.
(47) Therapeutic Drug Repurposing, Repositioning and Rescue Part II: Business Review. *Drug Discovery World* **2015**, *16*, 57−72.
(48) Oprea, T. I.; Overington, J. P. Computational and Practical Aspects of Drug Repositioning. *Drug Repurposing Rescue Repositioning* **2015**, *1*, 28−35.
(49) FDA Approves First Treatment for COVID-19. https://www.fda.gov/news-events/press-announcements/fda-approves-first-treatment-covid-19 (accessed Dec 26, 2020).
(50) Wang, Y.; Zhang, D.; Du, G.; Du, R.; Zhao, J.; Jin, Y.; Fu, S.; Gao, L.; Cheng, Z.; Lu, Q.; Hu, Y.; Luo, G.; Wang, K.; Lu, Y.; Li, H.; Wang, S.; Ruan, S.; Yang, C.; Mei, C.; Wang, Y.; Ding, D.; Wu, F.; Tang, X.; Ye, X.; Ye, Y.; Liu, B.; Yang, J.; Yin, W.; Wang, A.; Fan, G.; Zhou, F.; Liu, Z.; Gu, X.; Xu, J.; Shang, L.; Zhang, Y.; Cao, L.; Guo, T.; Wan, Y.; Qin, H.; Jiang, Y.; Jaki, T.; Hayden, F. G.; Horby, P. W.; Cao, B.; Wang, C. Remdesivir in Adults with Severe COVID-19: A Randomised, Double-Blind, Placebo-Controlled, Multicentre Trial. *Lancet* **2020**, *395*, 1569−1578.
(51) Chen, C. Y.-C. TCM Database@Taiwan: The World's Largest Traditional Chinese Medicine Database for Drug Screening in Silico. *PLoS One* **2011**, *6*, No. e15939.
(52) Mohanraj, K.; Karthikeyan, B. S.; Vivek-Ananth, R. P.; Chand, R. P. B.; Aparna, S. R.; Mangalapandi, P.; Samal, A. IMPPAT: A Curated Database of I Ndian M Edicinal P Lants, P Hytochemistry A Nd T Herapeutics. *Sci. Rep.* **2018**, *8*, 4329.
(53) Gentile, D.; Patamia, V.; Scala, A.; Sciortino, M. T.; Piperno, A.; Rescifina, A. Putative Inhibitors of SARS-CoV-2 Main Protease from A Library of Marine Natural Products: A Virtual Screening and Molecular Modeling Study. *Mar. Drugs* **2020**, *18*, 225.
(54) Joshi, T.; Joshi, T.; Pundir, H.; Sharma, P.; Mathpal, S.; Chandra, S. Predictive Modeling by Deep Learning, Virtual Screening and Molecular Dynamics Study of Natural Compounds against SARS-CoV-2 Main Protease. *J. Biomol. Struct. Dyn.* **2020**, 1−19.
(55) van Santen, J. A.; Jacob, G.; Singh, A. L.; Aniebok, V.; Balunas, M. J.; Bunsko, D.; Neto, F. C.; Castaño-Espriu, L.; Chang, C.; Clark, T. N.; Cleary Little, J. L.; Delgadillo, D. A.; Dorrestein, P. C.; Duncan, K. R.; Egan, J. M.; Galey, M. M.; Haeckl, F. P. J.; Hua, A.; Hughes, A. H.; Iskakova, D.; Khadilkar, A.; Lee, J.-H.; Lee, S.; LeGrow, N.; Liu, D. Y.; Macho, J. M.; McCaughey, C. S.; Medema, M. H.; Neupane, R. P.; O'Donnell, T. J.; Paula, J. S.; Sanchez, L. M.; Shaikh, A. F.; Soldatou, S.; Terlouw, B. R.; Tran, T. A.; Valentine, M.; van der Hooft, J. J. J.; Vo, D. A.; Wang, M.; Wilson, D.; Zink, K. E.; Linington, R. G. The Natural Products Atlas: An Open Access Knowledge Base for Microbial Natural Products Discovery. *ACS Cent. Sci.* **2019**, *5*, 1824−1833.
(56) Krafts, K.; Hempelmann, E.; Skórska-Stania, A. From Methylene Blue to Chloroquine: A Brief Review of the Development of an Antimalarial Therapy. *Parasitol. Res.* **2012**, *111*, 1−6.
(57) Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B. A.; Thiessen, P. A.; Yu, B.; Zaslavsky, L.; Zhang, J.; Bolton, E. E. PubChem 2019 Update: Improved Access to Chemical Data. *Nucleic Acids Res.* **2019**, *47*, D1102−D1109.
(58) Mendez, D.; Gaulton, A.; Bento, A. P.; Chambers, J.; De Veij, M.; Félix, E.; Magariños, M. P.; Mosquera, J. F.; Mutowo, P.; Nowotka, M.; Gordillo-Marañón, M.; Hunter, F.; Junco, L.; Mugumbate, G.; Rodriguez-Lopez, M.; Atkinson, F.; Bosc, N.; Radoux, C. J.; Segura-Cabrera, A.; Hersey, A.; Leach, A. R. ChEMBL: Towards Direct Deposition of Bioassay Data. *Nucleic Acids Res.* **2019**, *47*, D930−D940.
(59) Easy compound ordering service - MolPort. https://www.molport.com/shop/index (accessed Dec 26, 2020).
(60) Kapusta, K.; Kar, S.; Collins, J. T.; Franklin, L. M.; Kolodziejczyk, W.; Leszczynski, J.; Hill, G. A. Protein Reliability Analysis and Virtual Screening of Natural Inhibitors for SARS-CoV-2 Main Protease (Mpro) through Docking, Molecular Mechanic & Dynamic, and ADMET Profiling. *J. Biomol. Struct. Dyn.* **2020**, *0*, 1−18.
(61) Ghaleb, A.; Aouidate, A.; Ayouchia, H. B. E.; Aarjane, M.; Anane, H.; Stiriba, S.-E. In Silico Molecular Investigations of Pyridine N-Oxide Compounds as Potential Inhibitors of SARS-CoV-2:3D QSAR, Molecular Docking Modeling, and ADMET Screening. *J. Biomol. Struct. Dyn.* **2020**, 1−11.

https://doi.org/10.1021/acs.jcim.1c00404
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

(62) Bahadur Gurung, A.; Ajmal Ali, M.; Lee, J.; Abul Farah, M.; Mashay Al-Anazi, K. Structure-Based Virtual Screening of Phytochemicals and Repurposing of FDA Approved Antiviral Drugs Unravels Lead Molecules as Potential Inhibitors of Coronavirus 3C-like Protease Enzyme. *J. King Saud Univ., Sci.* **2020**, *32*, 2845−2853.
(63) Mohammad, T.; Shamsi, A.; Anwar, S.; Umair, M.; Hussain, A.; Rehman, M. T.; AlAjmi, M. F.; Islam, A.; Hassan, M.. I. Identification of High-Affinity Inhibitors of SARS-CoV-2 Main Protease: Towards the Development of Effective COVID-19 Therapy. *Virus Res.* **2020**, *288*, 198102.
(64) Chakraborti, S.; Bheemireddy, S.; Srinivasan, N. Repurposing Drugs against the Main Protease of SARS-CoV-2: Mechanism-Based Insights Supported by Available Laboratory and Clinical Data. *Mol. Omics* **2020**, *16*, 474−491.
(65) Panda, P. K.; Arul, M. N.; Patel, P.; Verma, S. K.; Luo, W.; Rubahn, H.-G.; Mishra, Y. K.; Suar, M.; Ahuja, R. Structure-Based Drug Designing and Immunoinformatics Approach for SARS-CoV-2. *Sci. Adv.* **2020**, *6*, No. eabb8097.
(66) 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.
(67) Hansson, T.; Oostenbrink, C.; van Gunsteren, W. Molecular Dynamics Simulations. *Curr. Opin. Struct. Biol.* **2002**, *12*, 190−196.
(68) Karplus, M.; McCammon, J. A. Molecular Dynamics Simulations of Biomolecules. *Nat. Struct. Biol.* **2002**, *9*, 646−652.
(69) Norberg, J.; Nilsson, L. Advances in Biomolecular Simulations: Methodology and Recent Applications. *Q. Rev. Biophys.* **2003**, *36*, 257−306.
(70) Alonso, H.; Bliznyuk, A. A.; Gready, J. E. Combining Docking and Molecular Dynamic Simulations in Drug Design. *Med. Res. Rev.* **2006**, *26*, 531−568.
(71) Lu, J.; Hou, X.; Wang, C.; Zhang, Y. Incorporating Explicit Water Molecules and Ligand Conformation Stability in Machine-Learning Scoring Functions. *J. Chem. Inf. Model.* **2019**, *59*, 4540−4549.
(72) Wójcikowski, M.; Ballester, P. J.; Siedlecki, P. Performance of Machine-Learning Scoring Functions in Structure-Based Virtual Screening. *Sci. Rep.* **2017**, *7*, 46710.
(73) Yasuo, N.; Sekijima, M. Improved Method of Structure-Based Virtual Screening via Interaction-Energy-Based Learning. *J. Chem. Inf. Model.* **2019**, *59*, 1050−1061.
(74) Ruvinsky, A. M. Role of Binding Entropy in the Refinement of Protein-Ligand Docking Predictions: Analysis Based on the Use of 11 Scoring Functions: Role of Binding Entropy in the Refinement of Protein-Ligand Docking Predictions. *J. Comput. Chem.* **2007**, *28*, 1364−1372.
(75) Acharya, A.; Agarwal, R.; Baker, M. B.; Baudry, J.; Bhowmik, D.; Boehm, S.; Byler, K. G.; Chen, S. Y.; Coates, L.; Cooper, C. J.; Demerdash, O.; Daidone, I.; Eblen, J. D.; Ellingson, S.; Forli, S.; Glaser, J.; Gumbart, J. C.; Gunnels, J.; Hernandez, O.; Irle, S.; Kneller, D. W.; Kovalevsky, A.; Larkin, J.; Lawrence, T. J.; LeGrand, S.; Liu, S.-H.; Mitchell, J. C.; Park, G.; Parks, J. M.; Pavlova, A.; Petridis, L.; Poole, D.; Pouchard, L.; Ramanathan, A.; Rogers, D. M.; Santos-Martins, D.; Scheinberg, A.; Sedova, A.; Shen, Y.; Smith, J. C.; Smith, M. D.; Soto, C.; Tsaris, A.; Thavappiragasam, M.; Tillack, A. F.; Vermaas, J. V.; Vuong, V. Q.; Yin, J.; Yoo, S.; Zahran, M.; Zanetti-Polzi, L. Supercomputer-Based Ensemble Docking Drug Discovery Pipeline with Application to Covid-19. *J. Chem. Inf. Model.* **2020**, *60*, 5832−5852.
(76) Hall-Swan, S.; Antunes, D. A.; Devaurs, D.; Rigo, M. M.; Kavraki, L. E.; Zanatta, G. DINC-COVID: A Webserver for Ensemble Docking with Flexible SARS-CoV-2 Proteins. *bioRxiv* **2021**. DOI: 10.1101/2021.01.21.427315.
(77) Cavasotto, C. N.; Di Filippo, J. I. In Silico Drug Repurposing for COVID-19: Targeting SARS-CoV-2 Proteins through Docking and Consensus Ranking. *Mol. Inf.* **2021**, *40*, 2000115.
(78) Wang, R.; Wang, S. How Does Consensus Scoring Work for Virtual Library Screening? An Idealized Computer Experiment. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1422−1426.
(79) Clark, R. D.; Strizhev, A.; Leonard, J. M.; Blake, J. F.; Matthew, J. B. Consensus Scoring for Ligand/Protein Interactions. *J. Mol. Graphics Modell.* **2002**, *20*, 281−295.
(80) Feher, M. Consensus Scoring for Protein-Ligand Interactions. *Drug Discovery Today* **2006**, *11*, 421−428.
(81) Palacio-Rodríguez, K.; Lans, I.; Cavasotto, C. N.; Cossio, P. Exponential Consensus Ranking Improves the Outcome in Docking and Receptor Ensemble Docking. *Sci. Rep.* **2019**, *9*, 5142.
(82) Ericksen, S. S.; Wu, H.; Zhang, H.; Michael, L. A.; Newton, M. A.; Hoffmann, F. M.; Wildman, S. A. Machine Learning Consensus Scoring Improves Performance Across Targets in Structure-Based Virtual Screening. *J. Chem. Inf. Model.* **2017**, *57*, 1579−1590.
(83) Charifson, P. S.; Corkery, J. J.; Murcko, M. a.; Walters, W. P. Consensus Scoring: A Method for Obtaining Improved Hit Rates from Docking Databases of Three-Dimensional Structures into Proteins. *J. Med. Chem.* **1999**, *42*, 5100−5109.
(84) Oda, A.; Tsuchida, K.; Takakura, T.; Yamaotsu, N.; Hirono, S. Comparison of Consensus Scoring Strategies for Evaluating Computational Models of Protein-Ligand Complexes. *J. Chem. Inf. Model.* **2006**, *46*, 380−391.
(85) Huang, N.; Shoichet, B. K.; Irwin, J. J. Benchmarking Sets for Molecular Docking. *J. Med. Chem.* **2006**, *49*, 6789−6801.
(86) Kukol, A. Consensus Virtual Screening Approaches to Predict Protein Ligands. *Eur. J. Med. Chem.* **2011**, *46*, 4661−4664.
(87) Zhang, Z.; Wu, W.; Hou, J.; Zhang, L.; Li, F.; Gao, L.; Wu, X.; Shi, J.; Zhang, R.; Long, H.; Lei, M.; Wu, W.; Guo, D.; Chen, K.; Hofmann, L. A.; Ci, Z. Active Constituents and Mechanisms of Respiratory Detox Shot, a Traditional Chinese Medicine Prescription, for COVID-19 Control and Prevention: Network-Molecular Docking-LC-MSE Analysis. *J. Integr. Med.* **2020**, *18*, 229−241.
(88) Vatansever, E. C.; Yang, K. S.; Drelich, A. K.; Kratch, K. C.; Cho, C.-C.; Kempaiah, K. R.; Hsu, J. C.; Mellott, D. M.; Xu, S.; Tseng, C.-T. K.; Liu, W. R. Bepridil Is Potent against SARS-CoV-2 In Vitro. *bioRxiv* **2020**, DOI: 10.1101/2020.05.23.112235.
(89) Wlodawer, A.; Dauter, Z.; Shabalin, I. G.; Gilski, M.; Brzezinski, D.; Kowiel, M.; Minor, W.; Rupp, B.; Jaskolski, M. Ligand-Centered Assessment of SARS-CoV-2 Drug Target Models in the Protein Data Bank. *FEBS J.* **2020**, *287*, 3703−3718.
(90) Webb, B.; Sali, A. Comparative Protein Structure Modeling Using MODELLER. In *Current Protocols in Bioinformatics*; John Wiley & Sons, Inc.: Hoboken, NJ, 2016; p 5.6.1−5.6.37.
(91) Word, J. M.; Lovell, S. C.; Richardson, J. S.; Richardson, D. C. Asparagine and Glutamine: Using Hydrogen Atom Contacts in the Choice of Side-Chain Amide Orientation. *J. Mol. Biol.* **1999**, *285*, 1735−1747.
(92) O'Boyle, N. M.; Banck, M.; James, C. a.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An Open Chemical Toolbox. *J. Cheminf.* **2011**, *3*, 33.
(93) Alhossary, A.; Handoko, S. D.; Mu, Y.; Kwoh, C.-K. Fast, Accurate, and Reliable Molecular Docking with QuickVina 2. *Bioinformatics* **2015**, *31*, 2214−2216.
(94) Santos-Martins, D.; Solis-Vasquez, L.; Tillack, A. F.; Sanner, M. F.; Koch, A.; Forli, S. Accelerating AutoDock4 with GPUs and Gradient-Based Local Search. *J. Chem. Theory Comput.* **2021**, *17*, 1060−1073.
(95) Forli, S.; Olson, A. J. A Force Field with Discrete Displaceable Waters and Desolvation Entropy for Hydrated Ligand Docking. *J. Med. Chem.* **2012**, *55*, 623−638.
(96) 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.
(97) Castro, M. A.; Llanos, M. A.; Rodenak-Kladniew, B. E.; Gavernet, L.; Galle, M. E.; Crespo, R. Citrus Reticulata Peel Oil as an Antiatherogenic Agent: Hypolipogenic Effect in Hepatic Cells, Lipid Storage Decrease in Foam Cells, and Prevention of LDL Oxidation. *Nutr., Metab. Cardiovasc. Dis.* **2020**, *30*, 1590−1599.

https://doi.org/10.1021/acs.jcim.1c00404
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

(98) RDKit. https://www.rdkit.org/ (accessed Dec 27, 2020).
(99) Douangamath, A.; Fearon, D.; Gehrtz, P.; Krojer, T.; Lukacik, P.; Owen, C. D.; Resnick, E.; Strain-Damerell, C.; Aimon, A.; Ábrányi-Balogh, P.; Brandão-Neto, J.; Carbery, A.; Davison, G.; Dias, A.; Downes, T. D.; Dunnett, L.; Fairhead, M.; Firth, J. D.; Jones, S. P.; Keeley, A.; Keserü, G. M.; Klein, H. F.; Martin, M. P.; Noble, M. E. M.; O'Brien, P.; Powell, A.; Reddi, R. N.; Skyner, R.; Snee, M.; Waring, M. J.; Wild, C.; London, N.; von Delft, F.; Walsh, M. A. Crystallographic and Electrophilic Fragment Screening of the SARS-CoV-2 Main Protease. *Nat. Commun.* **2020**, *11*, 5047.
(100) Pavlova, A.; Lynch, D. L.; Daidone, I.; Zanetti-Polzi, L.; Smith, M. D.; Chipot, C.; Kneller, D. W.; Kovalevsky, A.; Coates, L.; Golosov, A. A.; Dickson, C. J.; Velez-Vega, C.; Duca, J. S.; Vermaas, J. V.; Pang, Y. T.; Acharya, A.; Parks, J. M.; Smith, J. C.; Gumbart, J. C. Inhibitor Binding Influences the Protonation States of Histidines in SARS-CoV-2 Main Protease. *Chem. Sci.* **2021**, *12*, 1513−1527.
(101) Kneller, D. W.; Phillips, G.; Weiss, K. L.; Zhang, Q.; Coates, L.; Kovalevsky, A. Direct Observation of Protonation State Modulation in SARS-CoV-2 Main Protease upon Inhibitor Binding with Neutron Crystallography. *J. Med. Chem.* **2021**, *64*, 4991−5000.
(102) Ghahremanpour, M. M.; Tirado-Rives, J.; Deshmukh, M.; Ippolito, J. A.; Zhang, C.-H.; Cabeza de Vaca, I.; Liosi, M.-E.; Anderson, K. S.; Jorgensen, W. L. Identification of 14 Known Drugs as Inhibitors of the Main Protease of SARS-CoV-2. *ACS Med. Chem. Lett.* **2020**, *11*, 2526−2533.
(103) Zhang, C.-H.; Stone, E. A.; Deshmukh, M.; Ippolito, J. A.; Ghahremanpour, M. M.; Tirado-Rives, J.; Spasov, K. A.; Zhang, S.; Takeo, Y.; Kudalkar, S. N.; Liang, Z.; Isaacs, F.; Lindenbach, B.; Miller, S. J.; Anderson, K. S.; Jorgensen, W. L. Potent Noncovalent Inhibitors of the Main Protease of SARS-CoV-2 from Molecular Sculpting of the Drug Perampanel Guided by Free Energy Perturbation Calculations. *ACS Cent. Sci.* **2021**, *7*, 467−475.
(104) Gupta, A.; Rani, C.; Pant, P.; Vijayan, V.; Vikram, N.; Kaur, P.; Singh, T. P.; Sharma, S.; Sharma, P. Structure-Based Virtual Screening and Biochemical Validation to Discover a Potential Inhibitor of the SARS-CoV-2 Main Protease. *ACS Omega* **2020**, *5*, 33151−33161.
(105) Sharma, P.; Vijayan, V.; Pant, P.; Sharma, M.; Vikram, N.; Kaur, P.; Singh, T. P.; Sharma, S. Identification of Potential Drug Candidates to Combat COVID-19: A Structural Study Using the Main Protease (Mpro) of SARS-CoV-2. *J. Biomol. Struct. Dyn.* **2020**, *0*, 1−11.

https://doi.org/10.1021/acs.jcim.1c00404
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX