JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

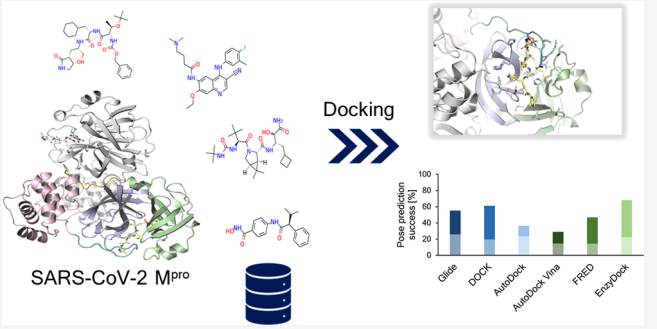
pubs.acs.org/jcim Article

Benchmarking the Ability of Common Docking Programs to Correctly Reproduce and Score Binding Modes in SARS-CoV-2 Protease Mpro

Shani Zev, Keren Raz, Renana Schwartz, Reem Tarabeh, Prashant Kumar Gupta, and Dan T. Major*

Cite This: https://doi.org/10.1021/acs.jcim.1c00263 Read Online

ABSTRACT: The coronavirus SARS-CoV-2 main protease, M^{pro}, is conserved among coronaviruses with no human homolog and has therefore attracted significant attention as an enzyme drug target for COVID-19. The number of studies targeting M^{pro} for in silico screening has grown rapidly, and it would be of great interest to know in advance how well docking methods can reproduce the correct ligand binding modes and rank these correctly. Clearly, current attempts at designing drugs targeting M^{pro} with the aid of computational docking would benefit from a priori knowledge of the ability of docking programs to predict correct binding modes and score these correctly. In the current work, we tested the ability of several leading docking programs, namely, Glide, DOCK, AutoDock, AutoDock Vina, FRED, and EnzyDock, to correctly identify and score the binding mode of M^{pro} ligands in 193 crystal structures. None of the codes were able to correctly identify the crystal structure binding mode (lowest energy pose with root-mean-square deviation < 2 Å) in more than 26% of the cases for noncovalently bound ligands (Glide: top performer), whereas for covalently bound ligands the top score was 45% (EnzyDock). These results suggest that one should perform in silico campaigns of M^{pro} with care and that more comprehensive strategies including ligand free energy perturbation might be necessary in conjunction with virtual screening and docking.



INTRODUCTION

Coronaviruses are positive-stranded RNA viruses that infect humans and animals and cause common and severe respiratory diseases, including severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS).[1,2] These viruses rely heavily on functional polypeptides that are generated by proteolytic cleavage of polyproteins that are translated from viral RNA. The principal coronavirus proteases responsible for this polypeptide formation are mainly protease and papain protease. The coronavirus SARS-CoV-2 main protease, M^{pro} (henceforth denoted simply as M^{pro}), has garnered significant attention in the past year as an enzyme drug target due to the COVID-19 pandemic outbreak.[3] M^{pro} is a druggable target[4,5] as it is conserved among coronaviruses and has no human homolog. The first M^{pro} structures were published in early 2020.[3,6,7] The first crystal structures revealed that the active form of M^{pro} is a homodimer containing two protomers, each composed of three domains (Figure 1A). The active site in M^{pro} is located in a cleft between domains I and II (Figure 1B) and features a noncanonical catalytic Cys−His dyad. The active site is composed of four regions: S1′, S1, S2, and S4, with the catalytic Cys145 located in S1′ and His41 located in S1′ and S2 (Figure 1C). To date, well over 200 three-dimensional holostructures have been resolved at a resolution of 3 Å or better.[3,6−9] In these structures, ligands bind to a variety of binding sites, including covalent and noncovalent binding in the main catalytic site, noncovalent binding in pockets in between the two proteomers, and weakly bound at the protein surface (i.e., in between crystallographic homodimer units). This wealth of structural information makes M^{pro} an attractive benchmarking system for testing the ability of docking programs to correctly identify and rank the correct poses. This becomes particularly important in light of the severity of the COVID-19 pandemic[2] and the significant number of mutant forms of the virus that are rapidly appearing and might render current and future vaccines ineffective. To date, M^{pro} has been studied extensively using ligand docking and screening tools[10−19] and computational enzymology tools, such as hybrid quantum mechanics−molecular mechanics (QM/MM).[20−29]

Received: March 6, 2021

https://doi.org/10.1021/acs.jcim.1c00263
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

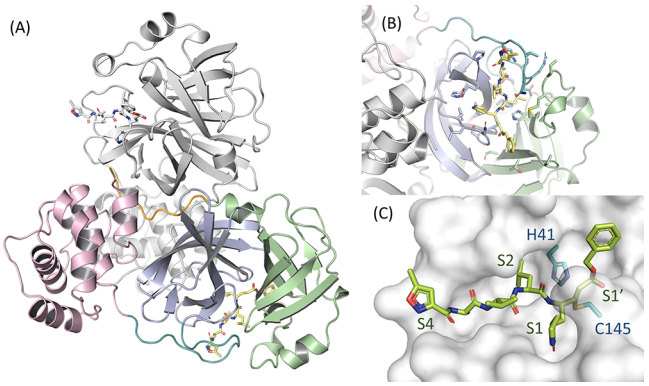


Figure 1. (A) SARS-CoV-2 M^{pro} dimer with bound ligand in each active site (PDB ID: 7BQY). The domains in chain A are colored as follows: N′-finger in bright orange, domain I in pale green, domain II in light blue, loop L3 in light teal, and domain III in light pink. The ligand bound to chain A is colored yellow and appears in stick representation. Both chain B and the ligand bound to it appear in gray. The same color code applies to (B), where the active site is shown in greater detail. (C) Four conserved subsites are presented, along with the Cys 145−His 41 dyad (deep teal stick representation). The covalently bound ligand appears in limon-shade stick representation. The protein appears as a white surface.

The number of studies targeting M^{pro} for in silico screening has grown,[30] and it would be of great interest to know in advance how well docking methods can reproduce the correct ligand binding modes and rank these correctly. Clearly, current attempts at designing drugs targeting M^{pro} with the aid of computational docking are problematic if programs struggle to predict correct binding modes and score these correctly. This is true even if common docking programs have undergone extensive testing since each protein target comes with its own challenges due to the complexity of binding pocket crevices, nature of interactions, and solvent exposure. Therefore, critical evaluation of how well common docking programs perform for M^{pro} is important.

Since the development of the first automated docking program DOCK,[31−35] a multitude of docking software packages have been developed, with different physicochemical approximations and algorithmic details. Popular docking programs in addition to DOCK, include Autodock,[36−38] Autodock Vina,[39] Glide,[40−42] Rosettaligand,[43,44] Gold,[45−47] and CDocker.[48−51] Widely used search algorithms include molecular dynamics (MD), Monte Carlo (MC), and genetic algorithms (GA). Common scoring functions include force field-based energy functions, such as CHARMM,[52−56] AMBER,[57] and OPLS[58,59] (i.e., molecular mechanics, MM), and knowledge-based scoring functions, such as DRUG-SCORE,[60] IT-SCORE,[61] DSX,[62] CHEMSCORE,[63] and SMoG.[64] Specialized docking programs addressing enzymes,[65,66] such as EnzyDock, have also been developed.[67] Current challenges for docking methods include protein flexibility,[68] ligand solvation, and binding-site hydration.[47,69] Thus far, several docking approaches have been employed to screen M^{pro} for potential drugs in virtual screening and drug repurposing campaigns, including Glide,[7,10,13,17,18,70−72] Autodock,[11,13,73,74] Autodock Vina,[11,13,14,19,71,73,75,76] Surflex,[77] PLANT,[78] DockThor,[76] fast pulling of ligands,[14] deep docking,[70] algebraic topology and deep learning,[79] and virtual reality-based docking.[16] However, to the best of our knowledge, no rigorous benchmark study addressing the ability of such docking tools to reproduce and correctly rank known ligand binding modes has been published, in spite of the known inherent challenges in docking.[80−83]

In the current work, we tested the ability of several leading docking programs to correctly identify and score the binding mode of M^{pro} ligands in 193 crystal structures (Figure S1). We tested the following docking programs: Glide,[40−42] DOCK,[31−35] Autodock,[36−38] Autodock Vina,[39] FRED,[84−87] and EnzyDock.[67] The current results suggest that care should be taken in applying docking programs to a challenging protein target such as M^{pro}.

METHODS

Preparing Mpro Structure Database. The available crystal structures of ligand-containing SARS-CoV-2 M^{pro} were downloaded from the RCSB PDB website (March-December 2020).[88] In total, we collected 193 different structures, including covalent and noncovalent ligands (Table S1 and Figure S1). All structures were aligned relative to one reference structure (PDB ID: 5R84) for easy comparison. To perform docking, we separated each protein and ligand into separate files, removing crystal waters, ions, and cosolvents. Missing residues were added using the Modeller homology program.[89] Hydrogens were added using the CHARMM simulation platform (using HBUILD) for the protein structure and using Openbabel for the ligands.[90,91] Visual inspection was also performed. For systems including only one monomer, the complementary unit was generated using the crystallographic information included in the PDB file using CHARMM. Protonation states of His residues were determined based on hydrogen bonding patterns and knowledge of the chemistry catalyzed by M^{pro} (Table S2), and they match the protonation states of key His residues recently published.[23] All docking simulations described below commenced with the CHARMM prepared systems.

Clustering of Ligands and Water Molecules. Chemical descriptors were calculated for all ligands from the 193 PDB files using RDKit libraries in Python. Features thought to be important for ligand binding were chosen. Specifically, we computed the number of rotatable torsions, molecular weight, number of H-bond donors and acceptors, number of aromatic rings, the fraction of carbon atoms in sp^3 hybridization (relative to all carbon atoms in the molecule), and log *P* values. We applied principal component analysis (PCA) using these ligands descriptors, followed by k-means clustering to cluster the ligands into groups. We selected the number of clusters by silhouette analysis of the k-means clustering results. Ligand clustering was performed using Python 3.7. Density-Based Spatial Clustering of Applications with Noise (DBSCAN) clustering was performed to analyze the water molecules from all crystal structures. These water molecules were not included in the docking studies.

Subsite-Binding Pocket Binding. To classify the binding patterns of all the 193 protein−ligand complexes, we categorized the ligands as bound on the surface, at the dimer interface, or in the active site. The latter was characterized according to the subsites S1, S1′, S2, and S4 (Figure 1C, Table S3).[7,77] A ligand is considered to occupy a binding pocket if any ligand atom is within 4.0 Å of any pocket atom and also within 3.0 Å of the geometric center of the pocket (defined as geometric center of all pocket atoms). Moreover, a ligand is considered to be in the proximity of a binding pocket if any ligand atom is within 4.0 Å of any pocket atom and also within

https://doi.org/10.1021/acs.jcim.1c00263
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

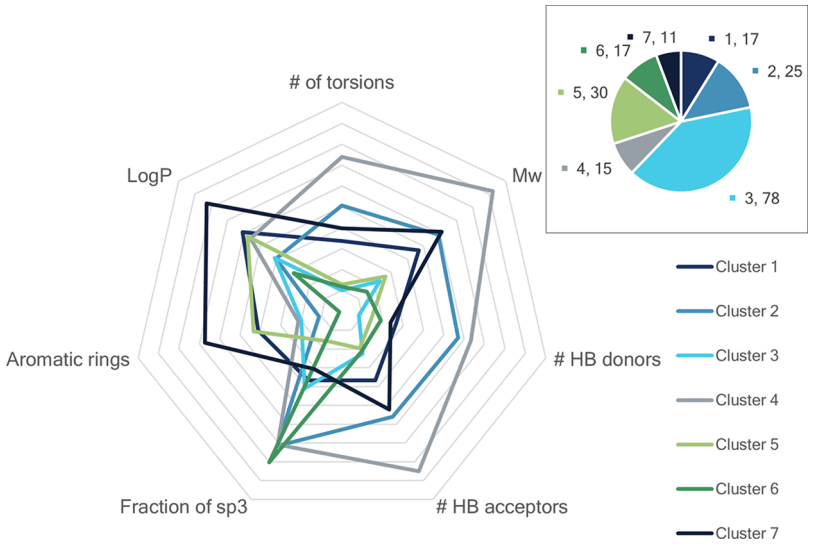


Figure 2. Normalized radar plot showing various features of each cluster centroid for SARS-CoV-2 M^{pro} ligands. Inset: pie plot with the relative fraction of each cluster among all the ligands studied in this work. The first number is the cluster name, and the second is the number of ligands in the cluster.

5.0 Å of the geometric center of the pocket. The criteria were designed to account for both small and bulky ligands and to distinguish between binding poses where ligand groups are well docked inside a subpocket and poses where ligand groups are located at the periphery of a subpocket. Cutoff values are suitable for various nonbonded interactions (e.g., $\pi-\pi$ stacking, hydrogen bond, ionic, and hydrophobic interactions). The final values were obtained by trial and error and validated by means of visual inspection. The subsite-binding pocket occupancy analysis was implemented as a CHARMM[90,91] script.

Docking Protocols. To compare the performance of selected docking programs for use with M^{pro} (search algorithm and scoring function), we performed noncovalent docking using DOCK,[31−35] Autodock Vina,[39] and FRED[84−87] and noncovalent and covalent docking using Glide,[40−42] Autodock,[36−38] and EnzyDock[67] to the systems described above. In all docking simulations described below, the ligand was fully flexible, while the protein was fixed (except for the covalently connected complexes, where the appropriate Cys residue was flexible).

Ligand Docking with Glide. Proteins and ligands were prepared using Schrödinger's Maestro (version 11.4, 2017-4 release) Prep Wiz and LigPrep modules, respectively, with default settings for docking with Glide. All covalent docking simulations were performed using the CovDock module available in Glide.[92] For the noncovalent simulations, the grid was generated using XGlide, which enables creation of different grids in parallel. The grids were centered around the ligand's centroid. The dimensions of the enclosing box and the bounding box were set to 12 × 12 × 12 $Å^3$ and 26 × 26 × 26 $Å^3$, respectively, for all cases. The ligand stereochemistry was kept during all docking simulations. The number of poses written per ligand was set to 10,000. The scaling factors of the vdW radii and the partial atomic charge cutoff were set to the default values 0.80 and 0.15, respectively. Standard precision (SP) mode was chosen for all ligand docking runs. The selection of the best-docked ligand structure among the proposed poses is made based on several model energies implemented with Glide (docking score, Prime energy and E-model energy, and cdock affinity). Solvent effects were incorporated using MMGBSA. All reported energies herein used the docking score function for noncovalent docking and the cdock affinity scoring function for covalent docking as these performed best [i.e., produced the highest number of top-ranking structures with root-mean-square deviation (rmsd) < 2 Å]. In all Glide docking simulations (ligand preparation, protein preparation, grid generation, covalent, and noncovalent docking), the OPLS3 force field[59,93] was used.

Ligand Docking with DOCK (Version 6.9). Proteins and ligands were prepared for docking simulations using the DockPrep option of Chimera v.14.[94] The grid was generated according to the center of mass of the crystal structure ligand with a grid spacing of 0.4 Å. The maximum and minimum radius of the sphere generated was set to 4 and 1.4 Å. All the spheres within 10 Å of the ligand were selected for docking. The box length surrounding the ligand was set to 6 Å, that is, the edge of the box from any atom of the ligand was at least 6 Å away, which easily accommodates all the selected spheres.

Ligand Docking with Autodock (Version 4.2). Proteins and ligand pdbqt files were prepared using standard AutoDock tools (prepare_flexreceptor4.py and prepare_ligand4.py). These files include Cartesian coordinates and Gasteiger atomic charges[95] for each atom. AutoDock employs a united atom method, and thus, no nonpolar hydrogens are present. The center of mass of the crystallographic ligand was used to determine the center of the grid. AutoDock uses one grid box to perform the docking calculations, and the dimensions of this box were set to 37.5 × 37.5 × 37.5 $Å^3$ and the spacing was set to 0.375 Å for all systems. We performed flexible ligand docking into a rigid protein environment using GA, with default settings. For covalent docking,[96] each ligand was prepared with the active Cys residue already present in the input file using AutoDock tools (prepare_receptor4.py and prepare_flexreceptor4.py). For covalent docking, the ligand flexible torsional angles were presampled using MC simulations with CHARMM prior to docking.

Ligand Docking with Autodock Vina. Protein and ligand input pdbqt files were prepared in the same way as for Autodock4.2 (see above). The size of the grid was set to 30.0

https://doi.org/10.1021/acs.jcim.1c00263
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

× 30.0 × 30.0 Å^3, and remaining parameters were set to default values.

Ligand Docking with FRED.[85−87] FRED is one of the docking programs available within the OpenEye scientific library. For the docking process, proteins and ligands were prepared using the graphical user interface "Make Receptor" provided with OpenEye. FRED creates a potential field around the binding site by producing a negative image, which complements the shape of the protein site. This potential field is represented on a contour, which completely surrounds the ligand. OMEGA, an internal program within OpenEye is used to generate an ensemble of conformers for each ligand. A total of 200 different conformers were generated for each ligand for docking, and the 50 lowest energy docked structures were used to select the best pose in terms of lowest rmsd or Chemgauss energy scoring relative to the crystal ligand structure. The proteins and ligands were held fixed during the docking process.

Ligand Docking with EnzyDock. Protein and ligand files were prepared as described at the beginning of Methods. CHARMM topology (RTF) and parameters (PRM) files for the ligands were generated using the CHARMM General Force Field (CGenFF) program.[53,97] For the proteins, CHARMM 36 was used.[52,54,55] The grid was generated according to the center of mass of the crystal structure ligand. The grid was generated with a mesh spacing of 0.25 Å and dimensions of 30.0 Å along each axis. The docking entailed 25 cycles of simulated MC and MD annealing for 25 differently rotated and MC torsion-sampled ligand configurations. Settings for noncovalent and covalent docking were identical.

RESULTS AND DISCUSSION

Ligand Clustering and the Subsite-Binding Pocket. To better understand the nature of the 193 M^{pro} complexes prior to discussing the docking results, we present analyses of the ligands and their binding modes. The ligands were clustered into seven main groups based on their chemical features by PCA and k-means clustering. The features of each cluster were normalized, and the average value for each cluster was calculated (Figure 2). The relative amount of the 193 ligands composing each cluster can be seen in the inserted pie chart (Figure 2).

For instance, cluster 6 is characterized by 17 ligands of low molecular weight and high fraction of carbon atoms with sp^3 hybridization, while cluster 3 is composed of 78 low-molecular-weight ligands that are rather rigid and slightly hydrophobic. Cluster 4 has 15 ligands of high molecular weight, flexible chains with sp^3 carbons, and several hydrogen bond donors and acceptors, while cluster 7 has medium-molecular-weight ligands that are highly hydrophobic with aromatic rings, yet has some hydrogen bond donors and acceptors.

Next, we analyzed the binding modes of the ligand clusters in M^{pro} (Tables 1 and S1). In Table 1 we present the fraction of each cluster that is bound in a specific subpocket of the active site, at the surface, or at the interface between the two monomers. Note that ligands can bind in more than one pocket, and hence, the fractions do not add up to unity for each cluster. Inspection of the binding data clearly shows that ligands of low molecular weight (clusters 1, 3, 5, and 6) tend to bind at the surface of the protein (i.e., clusters 3, 5, and 6 are caught in between crystal units) or at the interface between the homodimer units (cluster 1). Still many low-molecular-weight ligands occupy pockets S1 and S1′ as these are covalently attached to C145. Ligands rich in aromatic rings and correspondingly high log *P* values (cluster 7) tend to occupy all pockets more than average, specifically sites S1 and S2. This is due to favorable $\pi-\pi$ interactions with F140, H163, and H172 located in S1 or H41 and Y54 in S2. Another important observation is that ligands more likely to bind to S4 (which is rarely occupied) belong to clusters 2 and 4, which tend to include large, flexible molecules that are rich in H-bond donors and acceptors.

Table 1. Fraction of Ligands in Each Cluster that are Located in a Specific Active-Site Pocket, on the Surface, or at the Interface between Monomers in the SARS-CoV-2 M^{pro} Crystal Structures

	S1	S1′	S2	S4	surface	dimer interface
1	0.74	0.32	0.62	0.29	0.06	0.18
2	0.84	0.44	0.74	0.50	0.04	0.00
3	0.42	0.35	0.16	0.05	0.22	0.06
4	0.87	0.67	0.80	0.87	0.00	0.00
5	0.57	0.18	0.25	0.10	0.27	0.00
6	0.24	0.44	0.12	0.00	0.29	0.06
7	0.82	0.50	0.77	0.36	0.09	0.09
Total	0.57	0.38	0.37	0.22	0.17	0.05

We also clustered the water molecules in all crystal structures using DBSCAN clustering. Following clustering, we removed all waters that overlap any bound ligand (Figure S2). These waters form a set of active site features that can be included in docking studies (these waters were not included in the current docking study).

Docking of Ligands in M^{pro}. We docked all ligands from 193 crystal M^{pro} structures into their respective crystal protein structure (Table S4). These crystal structures include ligands bound noncovalently to the main binding pocket, surface and dimer interface, as well as covalently attached ligands. In all results below, we present the success rate of different docking programs in reproducing the crystal bound poses. For DOCK, AutoDock Vina, and FRED, the results reflect the noncovalent complexes only.

In Figure 3A we show the overall success of all programs. Glide and EnzyDock reproduce the correct crystal structure pose (rmsd < 2 Å) for over 50% of the structures, with success rates of 64 and 70%, respectively, while for AutoDock, this rate falls to 40%. However, in many cases, even if the correct pose is identified, it is not scored as lowest in energy, and the success rate reduces to 33% (Glide), AutoDock (30%), and EnzyDock (35%). The overall success rates of Glide, AutoDock, and EnzyDock are in part due to the covalent complexes, whose poses are easier to reproduce than the noncovalent ones. If we analyze the success rates of identifying only the covalently bound complexes, Glide, AutoDock, and EnzyDock identify the correct poses 70, 42, and 71% of the cases, while the correct pose is also scored as the best one in 38, 36, and 45% of the cases (Figure 3B). For the noncovalent complexes, Glide, DOCK, AutoDock, AutoDock Vina, FRED, and EnzyDock identify the correct poses in 55, 61, 37, 29, 46, and 68% of the cases, respectively, while these are ranked as the lowest energy poses in 26, 20, 24, 14, 14, and 22% of the cases, respectively (Figure 3C). Finally, if we remove the complexes with surface-bound ligands (i.e., ligands bound in between crystal units), all programs perform significantly better (Figure 3D). The correct poses are identified (and scored as best) as follows (%): Glide 74 (39), DOCK 77 (29),

https://doi.org/10.1021/acs.jcim.1c00263
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

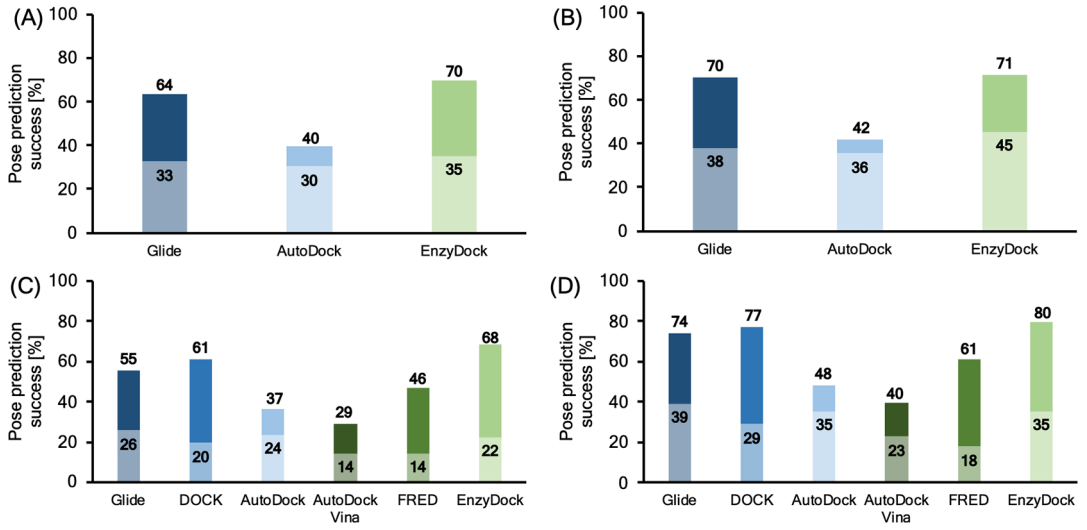


Figure 3. Docking pose prediction success (%) for selected methods for 193 crystal structures of SARS-CoV-2 M^{pro}. For each column, the upper part represents the ability of methods to identify poses with rmsd < 2 Å, while the lower part represents ability of methods to identify a pose with rmsd < 2 Å as the lowest energy pose. (A) All 193 crystal structures, (B) 108 crystal structures with covalently bound ligands, (C) 85 crystal structures with noncovalently bound ligands, (D) 51 crystal structures with noncovalently bound ligands excluding surface bound ligands.

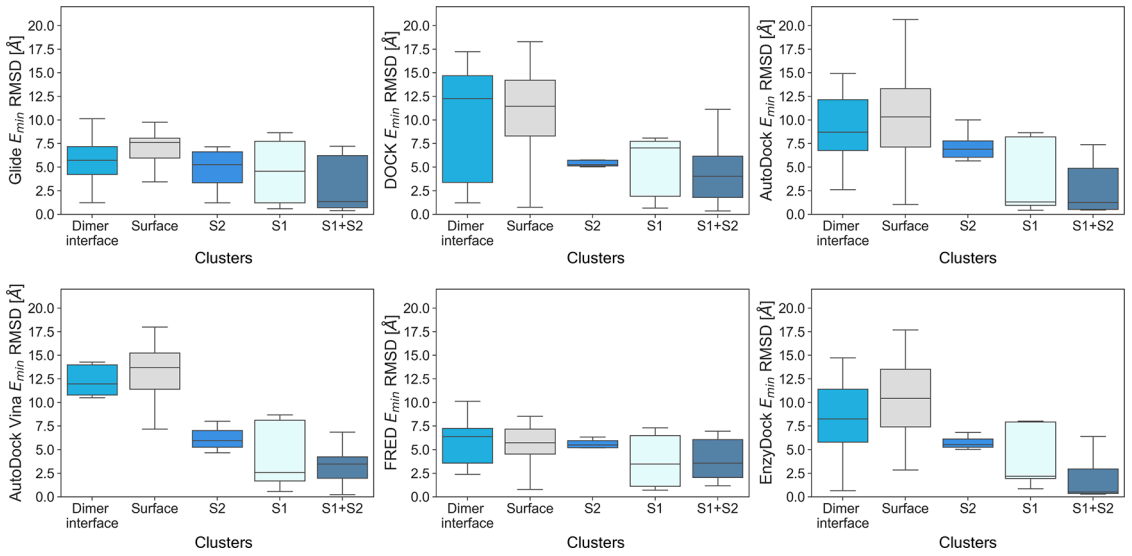


Figure 4. Distributions of rmsd values (Å) for the lowest-scoring pose at different sites in SARS-CoV-2 M^{pro} for selected noncovalent docking methods. The ligands were clustered into groups occupying similar regions, and only clusters with more than seven members are shown.

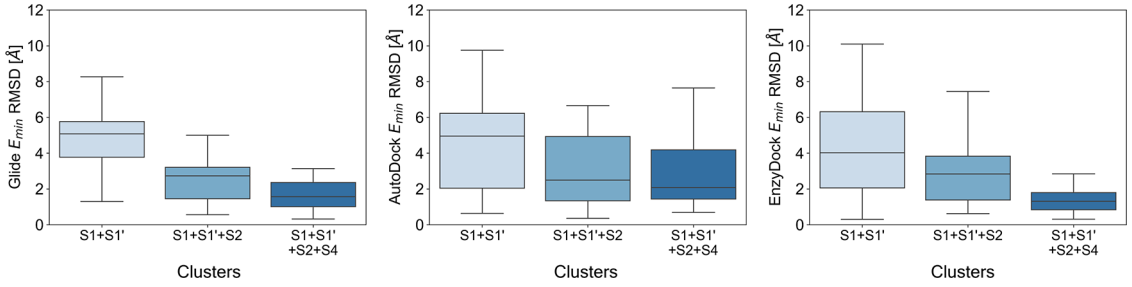


Figure 5. Distributions of rmsd values (Å) for the lowest-scoring pose at different sites in SARS-CoV-2 M^{pro} for selected covalent docking methods. The ligands were clustered into groups occupying similar regions, and only clusters with more than 22 members are shown.

AutoDock 48 (35), AutoDock Vina 40 (23), FRED 61 (18), and EnzyDock 80 (35), respectively.

Next, we analyze how the different programs perform as a function of binding site locations on M^{pro}. In Figures S3 and 4, we present box plots of the best rmsd values and the rmsd values for the lowest scoring pose for noncovalently bound ligands, respectively. All methods struggle with ligands bound at the interface between monomers and on the protein surface,

https://doi.org/10.1021/acs.jcim.1c00263
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

with Glide and FRED producing the best results. Additionally, most methods perform better for ligands bound to more than a single pocket (i.e., S1 + S2), and this trend is particularly clear for Glide and EnzyDock.

Similarly, we analyze how the different covalent docking programs perform as a function of binding site locations on M^{pro}. In Figures S4 and 5, we present box plots of the best rmsd values and the rmsd values for the lowest scoring pose for covalently bound ligands, respectively. Also here, we observe a general trend, where the docking methods perform better for ligands occupying more pockets.

CONCLUSIONS

In wake of the growing threat emerging from the SARS-CoV-2 pandemic, the modeling community has rushed to study a variety of potential pharmaceutical targets. One of these targets, M^{pro}, is particularly attractive as this enzyme has no human analogue and is conserved among coronaviruses. A large number of studies have addressed ligand docking and virtual screening of ligand libraries against M^{pro} in search of promising leads. A prerequisite for such studies is the ability of the docking programs to correctly identify and score ligand poses. Due to the intense efforts by the scientific community, there is already a wealth of structural biology information on M^{pro}, hence enabling comparative studies of different docking approaches against this target. To date, the available crystal structures of M^{pro} include ligands bound covalently and noncovalently to the main catalytic site, surface, and in between the two monomers. Here, we studied several leading docking codes, namely, Glide, DOCK, AutoDock, AutoDock Vina, FRED, and EnzyDock, and evaluate their ability to correctly reproduce and score the crystal structure ligand configuration for 193 M^{pro} crystal structures. None of the codes are able to correctly identify and score the crystal structure in more than 26% of the cases for noncovalently bound ligands (Glide top performer), whereas for covalently bound ligands, the top score was 45% (EnzyDock best performer). Additionally, a general trend, where several of the docking methods (e.g., Glide and EnzyDock) perform better for larger, bulkier ligands occupying more than a single pocket, is observed. All docking methods struggle with prediction of small ligands. In the original crystal structures, many of the smaller ligands are surrounded by numerous explicit water molecules, dimethyl sulfoxide molecules, and Cl^- ions that were removed prior to docking. Thus, these redocking trends may be ascribed to difficulty in accurately scoring docking poses, where a delicate balance between intra- and intermolecular terms and solvation terms must be stricken.

In conclusion, the current results suggest that one should perform docking studies and virtual screening campaigns of M^{pro} with care and that more comprehensive strategies might be necessary. Such strategies might include initial virtual screening (e.g., using FRED or AutoDock Vina) or docking (e.g., Glide or EnzyDock), followed by more rigorous ligand free energy binding calculations[98,99] and in-depth QM/MM studies.[20,24,26,28,29] Inclusion of conserved water molecules, as identified in this study, may also be of help in guiding the docking process. Indeed, MD simulations have pointed to several water molecules, as important for M^{pro}.[11,18,20,24,26,29,77]

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.1c00263.

Chemical structures of all docked ligands, conserved water molecules, additional docking results, all rmsd data, and active-site pocket occupation (PDF)

AUTHOR INFORMATION

Corresponding Author

Dan T. Major − *Department of Chemistry and Institute for Nanotechnology & Advanced Materials, Bar-Ilan University, Ramat-Gan 52900, Israel;* orcid.org/0000-0002-9231-0676; Email: majort@biu.ac.il

Authors

Shani Zev − *Department of Chemistry and Institute for Nanotechnology & Advanced Materials, Bar-Ilan University, Ramat-Gan 52900, Israel*

Keren Raz − *Department of Chemistry and Institute for Nanotechnology & Advanced Materials, Bar-Ilan University, Ramat-Gan 52900, Israel*

Renana Schwartz − *Department of Chemistry and Institute for Nanotechnology & Advanced Materials, Bar-Ilan University, Ramat-Gan 52900, Israel*

Reem Tarabeh − *Department of Chemistry and Institute for Nanotechnology & Advanced Materials, Bar-Ilan University, Ramat-Gan 52900, Israel*

Prashant Kumar Gupta − *Department of Chemistry and Institute for Nanotechnology & Advanced Materials, Bar-Ilan University, Ramat-Gan 52900, Israel;* orcid.org/0000-0002-4792-7538

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.1c00263

Author Contributions

S.Z., K.R., and P.K.G. contributed equally. The docking simulations and system preparations were performed by all authors. The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

All preprepared 193 Mpro systems and accompanying Python and CHARMM scripts are available at https://github.com/shanizev/Benchmarking-SARS-CoV-2. EnzyDock is freely available for noncommercial use on https://github.com/majordt/EnzyDock.

ACKNOWLEDGMENTS

This work was supported by the Israel Ministry of Science, Technology and Space (grant 3-16310) and Israeli Science Foundation (grant # 1683/18).

ABBREVIATIONS

SARS, severe acute respiratory syndrome; SARS-CoV-2, SARS coronavirus 2; MERS, Middle East respiratory syndrome; MD, molecular dynamics; MC, Monte Carlo

REFERENCES

(1) Anand, K.; Ziebuhr, J.; Wadhwani, P.; Mesters, J. R.; Hilgenfeld, R. Coronavirus main proteinase (3CLpro) structure: Basis for design of anti-sars drugs. *Science* **2003**, *300*, 1763−1767.

(2) Liu, C.; Zhou, Q.; Li, Y.; Garner, L. V.; Watkins, S. P.; Carter, L. J.; Smoot, J.; Gregg, A. C.; Daniels, A. D.; Jervey, S.; Albaiu, D. Research and development on therapeutic agents and vaccines for COVID-19 and related human coronavirus diseases. *ACS Cent. Sci.* **2020**, *6*, 315−331.

(3) 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.

(4) Qiao, J.; Li, Y.-S.; Zeng, R.; Liu, F.-L.; Luo, R.-H.; Huang, C.; Wang, Y.-F.; Zhang, J.; Quan, B.; Shen, C.; Mao, X.; Liu, X.; Sun, W.; Yang, W.; Ni, X.; Wang, K.; Xu, L.; Duan, Z.-L.; Zou, Q.-C.; Zhang, H.-L.; Qu, W.; Long, Y.-H. -P.; Li, M.-H.; Yang, R.-C.; Liu, X.; You, J.; Zhou, Y.; Yao, R.; Li, W.-P.; Liu, J.-M.; Chen, P.; Liu, Y.; Lin, G.-F.; Yang, X.; Zou, J.; Li, L.; Hu, Y.; Lu, G.-W.; Li, W.-M.; Wei, Y.-Q.; Zheng, Y.-T.; Lei, J.; Yang, S. SARS-CoV-2 Mpro inhibitors with antiviral activity in a transgenic mouse model. *Science* **2021**, *371*, 1374−1378.

(5) Pelly, S.; Liotta, D. Potent SARS-CoV-2 direct-acting antivirals provide an important complement to COVID-19 vaccines. *ACS Cent. Sci.* **2021**, *7*, 396−399.

(6) Dai, W.; Zhang, B.; Jiang, X.-M.; Su, H.; Li, J.; Zhao, Y.; Xie, X.; Jin, Z.; Peng, J.; Liu, F. Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease. *Science* **2020**, *368*, 1331−1335.

(7) 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.

(8) Kneller, D. W.; Phillips, G.; O'Neill, H. M.; Jedrzejczak, R.; Stols, L.; Langan, P.; Joachimiak, A.; Coates, L.; Kovalevsky, A. Structural plasticity of SARS-CoV-2 3CL M^{pro} active site cavity revealed by room temperature x-ray crystallography. *Nat. Commun.* **2020**, *11*, 3202.

(9) Günther, S.; Reinke, P. Y. A.; Fernández-García, Y.; Lieske, J.; Lane, T. J.; Ginn, H. M.; Koua, F. H. M.; Ehrt, C.; Ewert, W.; Oberthuer, D.; Yefanov, O.; Meier, S.; Lorenzen, K.; Krichel, B.; Kopicki, J. D.; Gelisio, L.; Brehm, W.; Dunkel, I.; Seychell, B.; Gieseler, H.; Norton-Baker, B.; Escudero-Pérez, B.; Domaracky, M.; Saouane, S.; Tolstikova, A.; White, T. A.; Hänle, A.; Groessler, M.; Fleckenstein, H.; Trost, F.; Galchenkova, M.; Gevorkov, Y.; Li, C.; Awel, S.; Peck, A.; Barthelmess, M.; Schluenzen, F.; Lourdu Xavier, P.; Werner, N.; Andaleeb, H.; Ullah, N.; Falke, S.; Srinivasan, V.; França, B. A.; Schwinzer, M.; Brognaro, H.; Rogers, C.; Melo, D.; Zaitseva-Doyle, J. J.; Knoska, J.; Peña-Murillo, G. E.; Mashhour, A. R.; Hennicke, V.; Fischer, P.; Hakanpää, J.; Meyer, J.; Gribbon, P.; Ellinger, B.; Kuzikov, M.; Wolf, M.; Beccari, A. R.; Bourenkov, G.; von Stetten, D.; Pompidor, G.; Bento, I.; Panneerselvam, S.; Karpics, I.; Schneider, T. R.; Garcia-Alai, M. M.; Niebling, S.; Günther, C.; Schmidt, C.; Schubert, R.; Han, H.; Boger, J.; Monteiro, D. C. F.; Zhang, L.; Sun, X.; Pletzer-Zelgert, J.; Wollenhaupt, J.; Feiler, C. G.; Weiss, M. S.; Schulz, E. C.; Mehrabi, P.; Karničar, K.; Usenik, A.; Loboda, J.; Tidow, H.; Chari, A.; Hilgenfeld, R.; Uetrecht, C.; Cox, R.; Zaliani, A.; Beck, T.; Rarey, M.; Günther, S.; Turk, D.; Hinrichs, W.; Chapman, H. N.; Pearson, A. R.; Betzel, C.; Meents, A. X-ray screening identifies active site and allosteric inhibitors of SARS-CoV-2 main protease. *Science* **2021**, *372*, 642−646.

(10) Wang, J. Fast identification of possible drug treatment of coronavirus disease-19 (COVID-19) through computational drug repurposing study. *J. Chem. Inf. Model.* **2020**, *60*, 3277−3286.

(11) 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.

(12) Zhang, L.; Lin, D.; Kusov, Y.; Nian, Y.; Ma, Q.; Wang, J.; von Brunn, A.; Leyssen, P.; Lanko, K.; Neyts, J.; de Wilde, A.; Snijder, E. J.; Liu, H.; Hilgenfeld, R. Alpha-ketoamides as broad-spectrum inhibitors of coronavirus and enterovirus replication: Structure-based design, synthesis, and activity assessment. *J. Med. Chem.* **2020**, *63*, 4562−4578.

(13) 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.

(14) Pham, M. Q.; Vu, K. B.; Pham, T. N. H.; Tran, L. H.; Tung, N. T.; Vu, V. V.; Nguyen, T. H.; Ngo, S. T. Rapid prediction of possible inhibitors for SARS-CoV-2 main protease using docking and fpl simulations. *RSC Adv.* **2020**, *10*, 31991−31996.

(15) Bzówka, M.; Mitusińska, K.; Raczyńska, A.; Samol, A.; Tuszyński, J. A.; Góra, A. Structural and evolutionary analysis indicate that the SARS-CoV-2 M^{pro} is a challenging target for small-molecule inhibitor design. *Int. J. Mol. Sci.* **2020**, *21*, 3099.

(16) Deeks, H. M.; Walters, R. K.; Barnoud, J.; Glowacki, D. R.; Mulholland, A. J. Interactive molecular dynamics in virtual reality is an effective tool for flexible substrate and inhibitor docking to the SARS-CoV-2 main protease. *J. Chem. Inf. Model.* **2020**, *60*, 5803−5814.

(17) Gahlawat, A.; Kumar, N.; Kumar, R.; Sandhu, H.; Singh, I. P.; Singh, S.; Sjostedt, A.; Garg, P. Structure-based virtual screening to discover potential lead molecules for the SARS-CoV-2 main protease. *J. Chem. Inf. Model.* **2020**, *60*, 5781−5793.

(18) Kumar, S.; Sharma, P. P.; Shankar, U.; Kumar, D.; Joshi, S. K.; Pena, L.; Durvasula, R.; Kumar, A.; Kempaiah, P.; Poonam; Rathi, B. Discovery of new hydroxyethylamine analogs against 3CL(pro) protein target of SARS-CoV-2: Molecular docking, molecular dynamics simulation, and structure-activity relationship studies. *J. Chem. Inf. Model.* **2020**, *60*, 5754−5770.

(19) Ngo, S. T.; Quynh Anh Pham, N.; Thi Le, L.; Pham, D. H.; Vu, V. V. Computational determination of potential inhibitors of SARS-CoV-2 main protease. *J. Chem. Inf. Model.* **2020**, *60*, 5771−5780.

(20) Świderek, K.; Moliner, V. Revealing the molecular mechanisms of proteolysis of SARS-CoV-2 M^{pro} by QM/MM computational methods. *Chem. Sci.* **2020**, *11*, 10626−10630.

(21) Wang, H.; He, S.; Deng, W.; Zhang, Y.; Li, G.; Sun, J.; Zhao, W.; Guo, Y.; Yin, Z.; Li, D. Comprehensive insights into the catalytic mechanism of middle east respiratory syndrome 3c-like protease and severe acute respiratory syndrome 3c-like protease. *ACS Catal.* **2020**, *10*, 5871−5890.

(22) Khrenova, M. G.; Tsirelson, V. G.; Nemukhin, A. V. Dynamical properties of enzyme−substrate complexes disclose substrate specificity of the SARS-CoV-2 main protease as characterized by the electron density descriptors. *Phys. Chem. Chem. Phys.* **2020**, *22*, 19069−19079.

(23) Verma, N.; Henderson, J. A.; Shen, J. Proton-coupled conformational activation of sars coronavirus main proteases and opportunity for designing small-molecule broad-spectrum targeted covalent inhibitors. *J. Am. Chem. Soc.* **2020**, *142*, 21883−21890.

(24) Ramos-Guzmań, C. A.; Ruiz-Pernía, J. J.; Tuñoń, I. a. Unraveling the SARS-CoV-2 main protease mechanism using multiscale methods. *ACS Catal.* **2020**, *10*, 12544−12554.

(25) Mondal, D.; Warshel, A. Exploring the mechanism of covalent inhibition: Simulating the binding free energy of α-ketoamide

https://doi.org/10.1021/acs.jcim.1c00263
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

inhibitors of the main protease of sars-cov-2. *Biochemistry* **2020**, *59*, 4601−4608.
(26) Ramos-Guzmán, C. A.; Ruiz-Pernía, J. J.; Tuñón, I. A microscopic description of SARS-CoV-2 main protease inhibition with michael acceptors. Strategies for improving inhibitors design. *Chem. Sci.* **2021**, *12*, 3489−3496.
(27) Jaffrelot-Inizan, T.; Célerse, F.; Adjoua, O.; Ahdab, D. E.; Jolly, L.-H.; Liu, C.; Ren, P.; Montes, M.; Lagarde, N.; Lagardère, L.; Monmarché, P.; Piquemal, J.-P. High-resolution mining of SARS-CoV-2 main protease conformational space: Supercomputer-driven unsupervised adaptive sampling. *Chem. Sci.* **2021**, *12*, 4889−4907.
(28) Awoonor-Williams, E.; Abu-Saleh, A. A.-A. A. Covalent and non-covalent binding free energy calculations for peptidomimetic inhibitors of SARS-CoV-2 main protease. *Phys. Chem. Chem. Phys.* **2021**, *23*, 6746−6757.
(29) Arafet, K.; Serrano-Aparicio, N.; Lodola, A.; Mulholland, A. J.; González, F. V.; Świderek, K.; Moliner, V. Mechanism of inhibition of SARS-CoV-2 M^{pro} by n3 peptidyl michael acceptor explained by QM/MM simulations and design of new derivatives with tunable chemical reactivity. *Chem. Sci.* **2021**, *12*, 1433−1444.
(30) Chodera, J.; Lee, A. A.; London, N.; von Delft, F. Crowdsourcing drug discovery for pandemics. *Nat. Chem.* **2020**, *12*, 581.
(31) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.* **1982**, *161*, 269−288.
(32) Shoichet, B. K.; Kuntz, I. D. Protein docking and complementarity. *J. Mol. Biol.* **1991**, *221*, 327−346.
(33) Jones, G.; Willett, P. Docking small-molecule ligands into active-sites. *Curr. Opin. Biotechnol.* **1995**, *6*, 652−656.
(34) Brooijmans, N.; Kuntz, I. D. Molecular recognition and docking algorithms. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 335−373.
(35) London, N.; Miller, R. M.; Krishnan, S.; Uchida, K.; Irwin, J. J.; Eidam, O.; Gibold, L.; Cimermančič, P.; Bonnet, R.; Shoichet, B. K.; Taunton, J. Covalent docking of large libraries for the discovery of chemical probes. *Nat. Chem. Biol.* **2014**, *10*, 1066−1072.
(36) Goodsell, D. S.; Olson, A. J. Automated docking of substrates to proteins by simulated annealing. *Proteins* **1990**, *8*, 195−202.
(37) Morris, G. M.; Goodsell, D. S.; Huey, R.; Olson, A. J. Distributed automated docking of flexible ligands to proteins: Parallel applications of autodock 2.4. *J. Comput. Aided Mol. Des.* **1996**, *10*, 293−304.
(38) 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.
(39) 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.* **2010**, *31*, 455−461.
(40) 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.
(41) 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.
(42) Repasky, M. P.; Murphy, R. B.; Banks, J. L.; Greenwood, J. R.; Tubert-Brohman, I.; Bhat, S.; Friesner, R. A. Docking performance of the glide program as evaluated on the astex and dud datasets: A complete set of glide sp results and selected results for a new scoring function integrating watermap and glide. *J. Comput. Aided Mol. Des.* **2012**, *26*, 787−799.
(43) Davis, I. W.; Baker, D. Rosettaligand docking with full ligand and receptor flexibility. *J. Mol. Biol.* **2009**, *385*, 381−392.
(44) Davis, I. W.; Raha, K.; Head, M. S.; Baker, D. Blind docking of pharmaceutically relevant compounds using rosettaligand. *Protein Sci.* **2009**, *18*, 1998−2002.
(45) Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R.; Jones, G. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727−748.
(46) Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. Improved protein-ligand docking using GOLD. *Proteins* **2003**, *52*, 609−623.
(47) Verdonk, M. L.; Chessari, G.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Nissink, J. W.; Taylor, R. D.; Taylor, R. Modeling water molecules in protein-ligand docking using GOLD. *J. Med. Chem.* **2005**, *48*, 6504−6515.
(48) Vieth, M.; Hirst, J. D.; Dominy, B. N.; Daigler, H.; Brooks, C. L., III Assessing search strategies for flexible docking. *J. Comput. Chem.* **1998**, *19*, 1623−1631.
(49) Vieth, M.; Hirst, J. D.; Kolinski, A.; Brooks, C. L., III Assessing energy functions for flexible docking. *J. Comput. Chem.* **1998**, *19*, 1612−1622.
(50) Wu, G.; Robertson, D. H.; Brooks, C. L., III; Vieth, M. Detailed analysis of grid-based molecular docking: A case study of CDOCKER-a CHARMm-based md docking algorithm. *J. Comput. Chem.* **2003**, *24*, 1549−1562.
(51) Gagnon, J. K.; Law, S. M.; Brooks, C. L. III Flexible CDOCKER: Development and application of a pseudo-explicit structure-based docking method within CHARMM. *J. Comput. Chem.* **2016**, *37*, 753−762.
(52) MacKerell, A. D.; Bashford, D.; Bellot, M.; Dunbrack, J. R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Kuczera, J.; Yin, D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(53) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D., Jr. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **2010**, *31*, 671−690.
(54) Huang, J.; MacKerell, A. D., Jr. CHARMM36 all-atom additive protein force field: Validation based on comparison to nmr data. *J. Comput. Chem.* **2013**, *34*, 2135−2145.
(55) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; Groot, B. L. d.; Grubmüller, H.; Alexander D. MacKerell, J. Charmm36m: An improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **2017**, *14*, 71−73.
(56) Lin, F.-Y.; Alexander D. MacKerell, J. Improved modeling of halogenated ligand−protein interactions using the drude polarizable and CHARMM additive empirical force fields. *J. Chem. Inf. Model.* **2019**, *59*, 215−228.
(57) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(58) Jorgensen, W. L.; Tirado-Rives, J. The OPLS force field for proteins. Energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* **1988**, *110*, 1657−1666.
(59) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.
(60) Velec, H. F.; Gohlke, H.; Klebe, G. Drugscore(CSD)-knowledge-based scoring function derived from small molecule crystal data with superior recognition rate of near-native ligand poses and better affinity prediction. *J. Med. Chem.* **2005**, *48*, 6296−6303.
(61) Grinter, S. Z.; Yan, C.; Huang, S. Y.; Jiang, L.; Zou, X. Automated large-scale file preparation, docking, and scoring:

https://doi.org/10.1021/acs.jcim.1c00263
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Evaluation of itscore and stscore using the 2012 community structure-activity resource benchmark. *J. Chem. Inf. Model.* **2013**, *53*, 1905−1914.
(62) Neudert, G.; Klebe, G. DSX: A knowledge-based scoring function for the assessment of protein-ligand complexes. *J. Chem. Inf. Model.* **2011**, *51*, 2731−2745.
(63) Moitessier, N.; Englebienne, P.; Lee, D.; Lawandi, J.; Corbeil, C. R. Towards the development of universal, fast and highly accurate docking/scoring methods: A long way to go. *Br. J. Pharmacol.* **2008**, *153*, S7−S26.
(64) Ishchenko, A. V.; Shakhnovich, E. I. SMall Molecule Growth 2001 (SMoG2001): An improved knowledge-based scoring function for protein-ligand interactions. *J. Med. Chem.* **2002**, *45*, 2770−2780.
(65) O'Brien, T. E.; Bertolani, S. J.; Tantillo, D. J.; Siegel, J. B. Mechanistically informed predictions of binding modes for carbocation intermediates of a sesquiterpene synthase reaction. *Chem. Sci.* **2016**, *7*, 4009−4015.
(66) O'Brien, T. E.; Bertolani, S. J.; Zhang, Y.; Siegel, J. B.; Tantillo, D. J. Predicting productive binding modes for substrates and carbocation intermediates in terpene synthases-bornyl diphosphate synthase as a representative case. *ACS Catal.* **2018**, *8*, 3322−3330.
(67) Das, S.; Shimshi, M.; Raz, K.; Nitoker, N.; Mhashal, A. R.; Ansbacher, T.; Major, D. T. EnzyDock: Protein-ligand docking of multiple reactive states along a reaction coordinate in enzymes. *J. Chem. Theory Comput.* **2019**, *15*, 5116−5134.
(68) Kokh, D. B.; Wade, R. C.; Wenzel, W. Receptor flexibility in small-molecule docking calculations. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, *1*, 298−314.
(69) Osterberg, F.; Morris, G. M.; Sanner, M. F.; Olson, A. J.; Goodsell, D. S. Automated docking to multiple target structures: Incorporation of protein mobility and structural water heterogeneity in autodock. *Proteins* **2002**, *46*, 34−40.
(70) 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*, No. e2000028.
(71) Fischer, A.; Sellner, M.; Neranjan, S.; Smiesko, 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.
(72) Achilonu, I.; Iwuchukwu, E. A.; Achilonu, O. J.; Fernandes, M. A.; Sayed, Y. Targeting the SARS-CoV-2 main protease using FDA-approved isavuconazonium, a p2-p3 alpha-ketoamide derivative and pentagastrin: An in-silico drug discovery approach. *J. Mol. Graphics Modell.* **2020**, *101*, 107730.
(73) 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.
(74) Shivanika, C.; Deepak Kumar, S.; Ragunathan, V.; Tiwari, P.; Sumitha, A.; Brindha Devi, P. Molecular docking, validation, dynamics simulations, and pharmacokinetic prediction of natural compounds against the SARS-CoV-2 main-protease. *J. Biomol. Struct. Dyn.* **2020**, 1−27.
(75) Keretsu, S.; Bhujbal, S. P.; Cho, S. J. Rational approach toward COVID-19 main protease inhibitors via molecular docking, molecular dynamics simulation and free energy calculation. *Sci. Rep.* **2020**, *10*, 17716.
(76) Ferraz, W. R.; Gomes, R. A.; Novaes, A. L. S.; Trossini, G. H. G. Ligand and structure-based virtual screening applied to the SARS-CoV-2 main protease: An in silico repurposing study. *Future Med. Chem.* **2020**, *12*, 1815−1828.
(77) Stoddard, S. V.; Stoddard, S. D.; Oelkers, B. K.; Fitts, K.; Whalum, K.; Whalum, K.; Hemphill, A. D.; Manikonda, J.; Martinez, L. M.; Riley, E. G.; Roof, C. M.; Sarwar, N.; Thomas, D. M.; Ulmer, E.; Wallace, F. E.; Pandey, P.; Roy, S. Optimization rules for SARS-CoV-2 Mpro antivirals: Ensemble docking and exploration of the coronavirus protease active site. *Viruses* **2020**, *12*, 942.
(78) Maffucci, I.; Contini, A. In silico drug repurposing for SARS-CoV-2 main proteinase and spike proteins. *J. Proteome Res.* **2020**, *19*, 4637−4648.
(79) Nguyen, D. D.; Gao, K.; Chen, J.; Wang, R.; Wei, G.-W. Unveiling the molecular mechanism of SARS-CoV-2 main protease inhibition from 137 crystal structures using algebraic topology and deep learning. *Chem. Sci.* **2020**, *11*, 12036−12046.
(80) Mortenson, P. N. Diverse, high-quality test set for the validation of protein-ligand native and cross-docking performance. *J. Med. Chem.* **2007**, *50*, 726−741.
(81) 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.
(82) Chen, Y. C. Beware of docking! *Trends Pharmacol. Sci* **2015**, *36*, 78−95.
(83) Fu, D. Y.; Meiler, J. Predictive power of different types of experimental restraints in small molecule docking: A review. *J. Chem. Inf. Model.* **2018**, *58*, 225−233.
(84) Hawkins, P. C. D.; Skillman, A. G.; Nicholls, A. Comparison of shape-matching and docking as virtual screening tools. *J. Med. Chem.* **2007**, *50*, 74−82.
(85) McGann, M. FRED pose prediction and virtual screening accuracy. *J. Chem. Inf. Model.* **2011**, *51*, 578−596.
(86) McGann, M. FRED and hybrid docking performance on standardized datasets. *J. Comput. Aided Mol. Des.* **2012**, *26*, 897−906.
(87) Kelley, B. P.; Brown, S. P.; Warren, G. L.; Muchmore, S. W. POSIT: Flexible shape-guided docking for pose prediction. *J. Chem. Inf. Model.* **2015**, *55*, 1771−1780.
(88) Berman, H. M.; Henrick, K.; Nakamura, H. Announcing the worldwide protein data bank. *Nat. Struct. Biol.* **2003**, *10*, 980.
(89) Šali, A.; Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **1993**, *234*, 779−815.
(90) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM - a program for macro-molecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187−217.
(91) Brooks, B. R.; Brooks, C. L., III; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545−1614.
(92) Zhu, K.; Borrelli, K. W.; Greenwood, J. R.; Day, T.; Abel, R.; Farid, R. S.; Harder, E. Docking covalent inhibitors: A parameter free approach to pose prediction and scoring. *J. Chem. Inf. Model.* **2014**, *54*, 1932−1940.
(93) Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; Kaus, J. W.; Cerutti, D. S.; Krilov, G.; Jorgensen, W. L.; Abel, R.; Friesner, R. A. OPLS3: A force field providing broad coverage of drug-like small molecules and proteins. *J. Chem. Theory Comput.* **2016**, *12*, 281−296.
(94) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605−1612.
(95) Gasteiger, J.; Marsili, M. Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges. *Tetrahedron* **1980**, *36*, 3219−3228.
(96) Bianco, G.; Forli, S.; Goodsell, D. S.; Olson, A. J. Covalent docking using autodock: Two-point attractor and flexible side chain methods. *Protein Sci.* **2016**, *25*, 295−301.
(97) Yu, W.; He, X.; Vanommeslaeghe, K.; Alexander D. MacKerell, J. Extension of the CHARMM general force field to sulfonyl-containing compounds and its utility in biomolecular simulations. *J. Comput. Chem.* **2012**, *33*, 2451−2468.

https://doi.org/10.1021/acs.jcim.1c00263
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

(98) Wang, L.; Berne, B. J.; Friesner, R. A. On achieving high accuracy and reliability in the calculation of relative protein−ligand binding affinities. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 1937−1942.
(99) 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.

https://doi.org/10.1021/acs.jcim.1c00263
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX