JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim Article

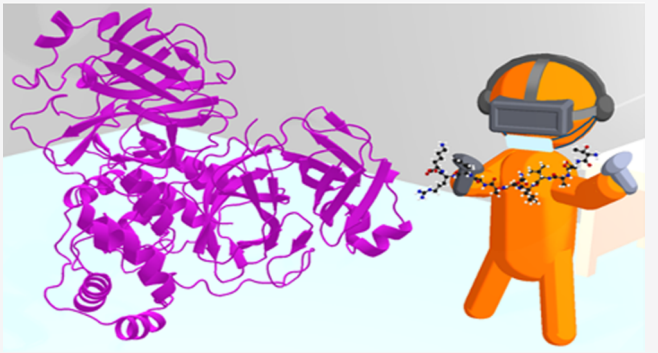
Interactive Molecular Dynamics in Virtual Reality Is an Effective Tool for Flexible Substrate and Inhibitor Docking to the SARS-CoV-2 Main Protease

Helen M. Deeks,[¶] Rebecca K. Walters,[¶] Jonathan Barnoud, David R. Glowacki,* and Adrian J. Mulholland*

Cite This: https://dx.doi.org/10.1021/acs.jcim.0c01030 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The main protease (Mpro) of the SARS-CoV-2 virus is one focus of drug development efforts for COVID-19. Here, we show that interactive molecular dynamics in virtual reality (iMD-VR) is a useful and effective tool for creating Mpro complexes. We make these tools and models freely available. iMD-VR provides an immersive environment in which users can interact with MD simulations and so build protein complexes in a physically rigorous and flexible way. Recently, we have demonstrated that iMD-VR is an effective method for interactive, flexible docking of small molecule drugs into their protein targets (Deeks et al. *PLoS One* **2020**, *15*, e0228461). Here, we apply this approach to both an Mpro inhibitor and an oligopeptide substrate, using experimentally determined crystal structures. For the oligopeptide, we test against a crystallographic structure of the original SARS Mpro. Docking with iMD-VR gives models in agreement with experimentally observed (crystal) structures. The docked structures are also tested in MD simulations and found to be stable. Different protocols for iMD-VR docking are explored, e.g., with and without restraints on protein backbone, and we provide recommendations for its use. We find that it is important for the user to focus on forming binding interactions, such as hydrogen bonds, and not to rely on using simple metrics (such as RMSD), in order to create realistic, stable complexes. We also test the use of apo (uncomplexed) crystal structures for docking and find that they can give good results. This is because of the flexibility and dynamic response allowed by the physically rigorous, atomically detailed simulation approach of iMD-VR. We make our models (and interactive simulations) freely available. The software framework that we use, Narupa, is open source, and uses commodity VR hardware, so these tools are readily accessible to the wider research community working on Mpro (and other COVID-19 targets). These should be widely useful in drug development, in education applications, e.g., on viral enzyme structure and function, and in scientific communication more generally.



1. INTRODUCTION

COVID-19 has rapidly spread across the world, rising to the level of a global pandemic within only a few months. This potentially lethal disease is caused by the SARS-CoV-2 coronavirus. Such coronaviruses pose a particular threat of emerging diseases as they can spread between different organisms[1,2] and lethally infect the respiratory, gastrointestinal, and central nervous systems.[3] There has been a rapid and widespread response to the COVID-19 pandemic from the global biomedical research community to understand the disease and attempt to identify potential vaccines and treatments. Antiviral drugs may have a useful role to play, particularly in the absence of an effective vaccine,[4] for this and for potential future coronavirus pandemics.

One promising drug target is the SARS-CoV-2 main protease (Mpro), also known as 3-chymotrypsin-like protease (3CLpro). Following the original SARS (severe acute respiratory syndrome) epidemic in 2003, the structure of the original SARS-CoV Mpro (which shares 96% sequence identity with SARS-CoV-2 Mpro)[5] was determined and has been used for structure-based drug discovery efforts, including computational molecular "docking".[6−8] The Mpro of both the original SARS and SARS-CoV-2 is a homodimeric protein,[9] responsible for the cleavage of polypeptides translated from viral RNA into smaller nonstructural proteins, using a catalytic Cys145/His41 dyad present in each subunit of the dimer.[8] Both the SARS-

Special Issue: COVID19 - Computational Chemists Meet the Moment

Received: September 1, 2020

CoV and SARS-CoV-2 Mpro are believed to operate on 11 different cleavage sites of the polyproteins PP1ab (replicase 1ab, ~790 kDa) and PP1a (replicase 1a).[10−12] It has been reported that substrate specificity of the Mpro is determined by amino acids at positions -P2-P1↓P1′- of PP1ab[13] (with a preference for residues Leu-Gln↓(Ser,Ala,Gly),[4,14] where ↓ marks the cleavage site). The nonstructural proteins generated from proteolytic cleavage play a vital role in viral transcription and replication,[15,16] making the Mpro essential for the viral lifecycle.[7,17−19]

There are currently many efforts globally aimed at identifying drug targets for COVID-19.[18,19] Molecular simulation and modeling methods have the potential to contribute to the discovery and development of SARS-CoV-2 Mpro inhibitors, e.g., in structure-based ligand discovery and drug repurposing efforts for this and other COVID-19 targets.[20,21] Recent crystallographic determination of the structure of the enzyme complexed with inhibitors[22,23] allows structure-based drug discovery methods to be used; at the time of writing, there are currently 196 SARS-CoV-2 Mpro structures in the Protein Data Bank. A standard initial approach in structure-based ligand discovery is computational molecular docking of approved drugs and other small molecules from docking libraries (such as the ZINC database[24]) to identify "hit" molecules that may bind to and inhibit the enzyme. Several docking studies on SARS-CoV-2 targets have already appeared, including for Mpro.[4,20,23,25,26] Such automated docking methods are typically highly approximate to allow rapid, high throughput screening of large numbers of compounds. For example, automated docking methods usually, by necessity, have a very limited treatment of protein dynamics and structural changes and use simplified descriptions of molecular interactions. Poses obtained with these high throughput methods typically require testing and further refinement. Recognized limitations of automated docking methods include the approximate nature of the energy functions used, limited treatment of solvation, and lack of protein and/or ligand conformational variability, particularly for large, flexible ligands. These simple methods allow automated docking to test millions of compounds in a relatively short time, but also mean that the results are prone to error and may identify many false positives.

More sophisticated biomolecular simulation methods, such as molecular dynamics (MD), can be used to filter out false positives from docking.[27] MD simulations can be used to account for protein flexibility, e.g., by generating ensembles of structures to dock ligands.[28] Human intuition and expertise play an essential role in drug discovery, e.g., in refining structures in crystallographic studies or for predicting binding modes of ligands. Emerging tools based on virtual reality (VR) can provide a useful addition to the armory of computational methods for drug discovery and development. Furthermore, VR potentially allows for sharing structural information in an intuitive and accessible form, as well as enabling distributed, virtual collaboration, e.g., when supported by cloud-based resources.[29] VR frameworks for protein visualization, such as Nanome[30] and ProteinVR,[31] are being used for the SARS-CoV-2 Mpro.[32,33] While VR is undoubtedly a useful tool for visualizing complex structures in 3D, many such representations are static and do not include protein dynamics. Narupa, an open-source software framework, allows users to manipulate rigorous, physics-based atomistic MD simulations within a VR environment, a method which we call 'interactive molecular dynamics in virtual reality' (iMD-VR).[34]

Recently, we have shown that iMD-VR using Narupa provides an effective and accurate approach for predicting the structures of protein−ligand complexes.[34] It allows fully flexible docking through the inclusion of molecular motion in a physically rigorous MD simulation. This is a different approach to the high-throughput, rigid docking methods described previously. iMD-VR allows interactive, dynamic, and flexible manipulation of structures, enabling the user to discover favorable binding modes. We applied this framework to interactively recreate crystallographic binding modes in several systems: pertinently, two of the three proteins studied were viral enzymes (HIV protease and influenza neuraminidase) with clinically approved drugs. We showed that nonexpert iMD-VR users can recreate experimentally observed structures of protein−ligand complexes, with a comparable level of accuracy to standard docking methods (within 2.15 Å RMSD of the crystallographic binding pose).[35]

Here, we apply iMD-VR to dock both a small ligand and an oligopeptide substrate to variants of Mpro (both structures are shown in Figure 1), testing different protocols. As in previous work on other enzymes, we compare results from iMD-VR with the crystal structures. We use iMD-VR to dock a small drug-like inhibitor into an apo SARS-CoV-2 Mpro structure. Taking advantage of the fully flexible docking possible with iMD-VR, we also dock an 11-mer oligopeptide into the original SARS-CoV Mpro with the H41A mutant, for which an oligopeptide complex crystal structure exists. The models are similar to the crystal structures. Finally, we use this knowledge of docking to the original SARS Mpro to create substrate complexes for the SARS-CoV-2 Mpro for which, at the time of writing, no such crystal structure exists (and, of course, capturing active enzyme−substrate structures crystallographically is generally not possible). MD simulation was performed on each iMD-VR docked structure to test the generated binding poses; overall, we find them to be stable and structurally consistent with the experimental crystal structures. We make these models available for the wider community to use.

This iMD-VR framework is open source, uses commodity hardware, and so is easily applicable. These tools are freely available to the global research community, and we believe that they should find wide application. This will be useful in structure-based drug discovery and development efforts targeting the SARS-CoV-2 Mpro and may help in developing inhibitors, complementing other computational and experimental methods, including more traditional high-throughput screening docking approaches. It will also be useful for education on viral enzyme structure and function (as we have shown in other contexts)[36] and testing hypotheses, e.g., on substrate binding and Mpro mechanism, complementing other types of simulations.[37]

2. METHODS

2.1. Simulation Parameters and Setup. The AMBER ff14sb force field[38] was used for protein and peptide structures. X77 was parametrized with GAFF.[39] All simulations used the OBC2 implicit water model[40] and used OpenMM as the force engine. Prior to iMD-VR, crystallographic water molecules were removed, and the structures were minimized and equilibrated: details are given in the S1 Text. As with previous work,[34] all iMD-VR simulations used an implicit solvent

https://dx.doi.org/10.1021/acs.jcim.0c01030
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

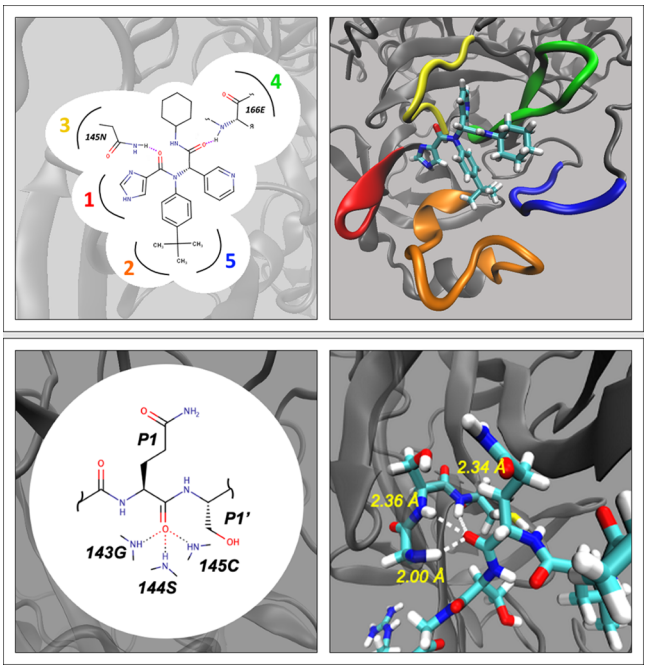


Figure 1. Important ligand and oligopeptide substrate interactions with Mpro. Top panels: details of the binding of the X77 inhibitor revealed by crystallography. (top left) The X77 inhibitor forms two hydrogen bonds with the SARS-CoV-2 Mpro; five active site loops around the ligand are also indicated schematically. (top right) X77 bound covalently in the active site of SARS-CoV-2 Mpro (PDB code: 6W63). The five active site loops are colored as follows: Loop (1), residues 22 to 27, is leftmost in red, Loop (2), residues 41 to 54, is at the bottom in orange, Loop (3), residues 140 to 145, is at the top left in yellow, Loop (4), residues 163 to 173, is at the top right in green, and Loop (5), residues 186 to 191, is rightmost in blue. Bottom panels: substrate complex models created in this work. (bottom left) Interactions between the 11-mer and Mpro that prime the carbonyl of the scissile bond for attack by the nucleophilic cysteine (these interactions are known as the oxyanion hole and are observed both for the SARS-CoV and SARS-CoV-2 Mpros). (bottom right) Oxyanion hole interactions in a structure from iMD-VR docking to an apo SARS-CoV Mpro structure (PDB code: 6M03).

model; although this is generally less accurate than explicitly modeling water molecules, implicit solvent was used to improve the latency and responsiveness of iMD-VR simulations. Furthermore, the workflow is more straightforward, insofar as the positions of water molecules do not need to be considered (allowing greater focus on the ligand). To prevent unrealistic perturbation of the protein structure during iMD-VR (e.g., protein unfolding), except where otherwise indicated, a harmonic positional restraint of 50 kJ/mol was applied to the protein backbone atoms, as in previous work.[34] Each iMD-VR docking experiment was repeated a total of five times, generating a total of five final structures per run.

To guide iMD-VR docking, a trace representation was created using the ligand atom positions from a known crystal structure, which was then superimposed into the VR simulation space; this visual guide is referred to as "trace atoms" and is described in our previous work.[34] These trace atoms are not part of the MD simulation itself but provide a target for docking. Further details of trace atom implementation can be found in the S1 Text and in our previous work on other targets.

To evaluate how closely the iMD-VR complexes replicate experimental structures (from which the trace atom positions were derived), further MD simulation and analysis were performed. The five representative structures were subjected to energy minimization and re-equilibrated using the protocol described in the S1 Text, allowing them to relax prior to any further analysis. After energy minimization, a 10 ns MD simulation was run, and the resulting trajectories were analyzed to evaluate similarity to the experimental structure for each complex. Details of these MD runs, and analysis, are given in the S1 Text.

2.2. iMD-VR Docking of a Small Ligand to Mpro. *2.2.1. Docking X77 into Ligand-Complexed Sars-CoV-2 Mpro.* We first tested the iMD-VR framework using the protocol from our previous work[34] on a known SARS-CoV-2 Mpro protein−ligand complex (PDB ID: 6W63). The ligand (X77) started in the bound position, as observed in the crystal structure. The operator interactively applied force on the ligand to move it out of the pocket to a position where it was clearly not bound to the protein, and then moved the ligand back to its initial pose, using trace atoms as a visual guide (i.e., a transluscent representation of X77 in the starting, complexed position which does not take part in the equation of motion). As noted in Section 2.1, to simplify and speed docking, we used an implicit solvent model (as in our previous work). We also applied positional restraints on the protein backbone to avoid large deformations (see S1 Text).

The small inhibitor, X77, was selected because it does not covalently bind to the protein and is similar in size and number of rotatable bonds to the compounds that we previously interactively undocked and redocked using iMD-VR.[34] Figure 1 shows the structure and binding mode of X77 in SARS-CoV-2 Mpro. There is uncertainty about the protonation state of the SARS-CoV-2 Mpro catalytic His41-Cys145 dyad, and it has been suggested that ligand binding to the active site promotes formation of the zwitterionic state.[41] Furthermore, previous work has shown different behavior between protein tautomers in iMD-VR.[34] To test this, we interactively unbound and docked X77 with both the neutral and zwitterionic forms of the dyad (Figure S1). After iMD-VR docking, a representative structure was selected based on having the lowest ligand heavy atom RMSD compared to the starting structure (i.e., the minimized and equilibrated experimental structure). This docking protocol is referred to as the 'cognate RMSD protocol (X77)'.

2.2.2. Docking X77 into Apo SARS-CoV-2 Mpro. We also tested iMD-VR docking of X77 into the apo form of SARS-CoV-2 Mpro (PDB code: 6M03). In this case, only the neutral tautomer was modeled because the redocked complexes were generally more stable than with the zwitterionic form (Figure S1). The same trace atoms from the 'cognate RMSD protocol (X77)' (described in Section 2.2.1) were superimposed onto the apo protein structure and used as a visual guide. We tested two iMD-VR protocols. In one protocol, the user tries to superimpose X77 over the trace atoms as closely as possible. A representative frame from this iMD-VR docking run was chosen based on having the lowest X77 heavy atom RMSD compared to the crystal structure of the complex. In the other protocol, the user still has trace atoms to guide ligand orientation but instead primarily focuses on reforming two hydrogen bonds (shown in Figure 1). Representative structures were chosen based on the two interactions being

https://dx.doi.org/10.1021/acs.jcim.0c01030
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

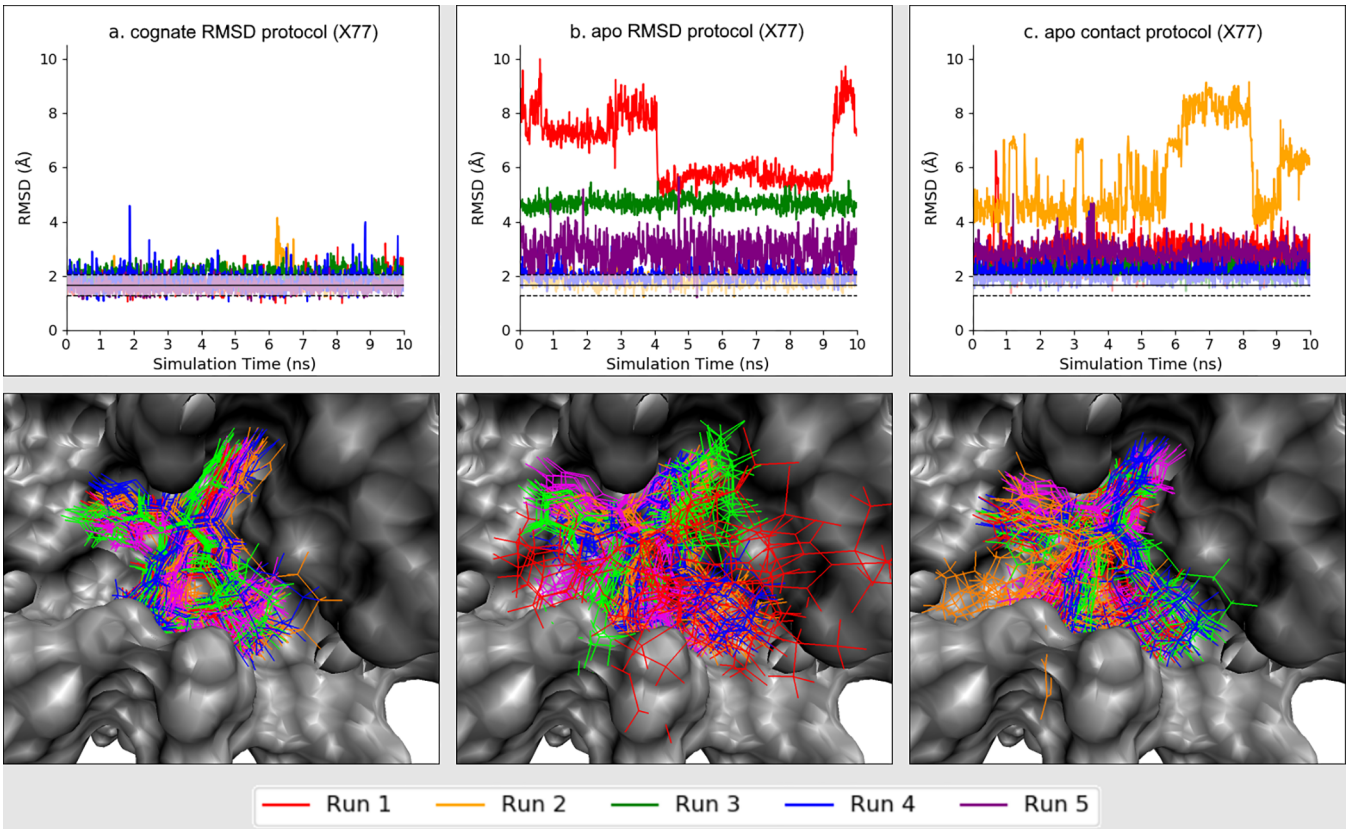


Figure 2. Validation MD simulations of iMD-VR docked X77, showing RMSD during 10 ns MD simulation. top: Results from MD simulations for structures from iMD-VR docking of the X77 ligand into (a) the cognate structure using the 'cognate X77 RMSD protocol (X77)', (b) the apo structure using the 'apo RMSD protocol (X77)', and (c) the apo structure using the 'apo contacts protocol (X77)'. For each of the 5 runs in each protocol, the heavy-atom RMSD of X77 is shown over 10 ns of MD simulation. Note, the RMSDs do not start at (or close to) zero because the complexes generated by iMD-VR were first relaxed by minimizing and equilibrating the structures. For reference, the black line represents the average X77 RMSD from 10 ns of MD on the original crystal structure (PDB code: 6W63). The translucent bars represent the standard deviation around this mean. bottom: For each of the five runs for each of the three docking protocols, snapshots from the 10 ns of MD simulation (taken every nanosecond) are superimposed on the X77 complex crystal structure (PDB code: 6W63). Each image corresponds to the graph directly above it.

present. These two protocols are referred to as the 'apo RMSD protocol (X77)' and the 'apo contact protocol (X77)'.

2.3. iMD-VR Docking of an Oligopeptide to Mpro. *2.3.1. Docking a Substrate into the H41A Mutant SARS-CoV Mpro.* The 11-mer oligopeptide (derived from one of the 11 cleavage sites of PP1ab, sequence TSAVLQSGFRK)[8] was interactively removed from and then docked into the active site of the SARS-CoV Mpro (inactive H41A mutant, PDB code: 2Q6G).[17] The aim was to test iMD-VR docking of long, flexible molecules and how closely iMD-VR docked structures resemble the experimental structure. We chose this substrate because it is processed by both the SARS-CoV and SARS-CoV-2 Mpro.[42] A SARS-CoV Mpro crystal structure is available for comparison. As with the small ligand docking described above, a trace representation of the substrate heavy atoms in the bound position from the crystal structure was superimposed onto the protein as a visual guide (trace atoms). Based on our results with the X77 ligand, we only modeled the neutral dyad for all substrate docking.

Before iMD-VR docking, 10 ns of MD simulation in implicit solvent was performed on the crystal structure. This MD simulation removes any initial strain in the structure and relaxes it in the solvent environment. It also provides a benchmark of how a substrate complex behaves in the same implicit solvent as in iMD-VR. The setup of this MD simulation followed the same protocol as all the other MD simulations and is described in the S1 Text. This simulation is referred to as the 'substrate reference simulation'.

We tested two protocols for iMD-VR docking. In both protocols, the docked oligopeptide trace atoms were rendered as a visual guide, and the focus of docking was on reforming 14 important hydrogen bonds observed in the crystal structure. A representative docked structure was extracted in which interactions that prime the carbonyl of the scissile bond for attack by the nucleophilic cysteine are present. These interactions are important for stabilizing high-energy oxyanion intermediates[43] and are known as the "oxyanion hole" (Figure 1). In the first protocol, we applied positional restraints to the backbone of the main protease, like those used for X77 (see above), and in our previous work.[34] In the second protocol, the main protease is fully flexible. These protocols are called the 'Backbone Restrained Protocol (Oligopeptide)' and the 'Fully Flexible Protocol (Oligopeptide)', respectively. Another protocol, the 'Lowest RMSD Protocol' was also tested. This protocol is similar to the 'Backbone Restrained Protocol (Oligopeptide)', but focuses on superimposing the substrate atoms directly on the trace atoms. For this protocol, we extracted the structure with the lowest substrate RMSD compared to the crystal structure.

https://dx.doi.org/10.1021/acs.jcim.0c01030
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

The resulting structures lacked important interactions observed in the crystal structure, and so the 'Lowest RMSD Protocol' was not used for further oligopeptide docking. These findings emphasize the limitations of RMSD alone as a metric for measuring quality of docking solutions. Formation of binding interactions, such as hydrogen bonds, is crucial. iMD-VR allows the user to find such interactions intuitively and correctly (within the limitations of the force field). Results of the docking experiment with 'Lowest RMSD Protocol' are detailed in the S1 Text.

2.3.2. Docking Substrate into Apo and Inhibitor Structures of SARS-CoV-2 Mpro. To the best of our knowledge, there is no crystal structure available of substrate bound to the SARS-CoV-2 Mpro. iMD-VR was employed here to dock the oligopeptide into two different structures of the SARS-CoV-2 Mpro: an apo structure (PDB code: 6M03) and the structure of the X77 inhibitor complex (PDB code: 6W63), described as the 'apo' structure and the 'inhibitor-complexed' structure, respectively. The 'Backbone Restrained Protocol (Oligopeptide)' and 'Fully Flexible Protocol (Oligopeptide)' were tested for iMD-VR docking to both SARS-CoV-2 Mpro structures, using trace atoms taken from the crystal structure of the substrate bound to the H41A mutant SARS-CoV. These trace atoms were used as a visual guide to indicate the position and orientation of the binding pocket.

3. RESULTS AND DISCUSSION

3.1. iMD-VR Docking of a Small Molecule Inhibitor (X77). *3.1.1. Docking X77 into Complexed SARS-CoV-2 Mpro.* We used iMD-VR to dock a known inhibitor (X77) into both the cognate protein complex (PDB code: 6W63) and the apo (PDB code: 6M03) forms of the SARS-CoV-2 Mpro. For the ligand-complexed structure, we applied our tested iMD-VR ligand docking protocol: the user interactively pulled X77 out of Mpro and then manually reintroduced it into the binding site using iMD-VR. The user was able to recreate the X77 bound structure, within a similar RMSD range to the systems studied using the previous protocol (i.e., less than 2.5 Å).[34] Five structures from iMD-VR docking were then subject to 10 ns of unrestrained MD simulation. The iMD-VR redocked complexes did not change significantly: the heavy atom RMSD of X77 remained within 2.5 Å of the reference structure (i.e., the equilibrated and minimized complex used to initiate iMD-VR). Furthermore, the RMSD for the five 10 ns runs was similar, indicating that the user consistently found a stable state (Figure 2a). The trajectory with the lowest RMSD had an average value of 1.6 Å. Combining the data from all five runs, the structures also showed little variation in position, with a total average RMSD of 1.8 Å and standard deviation of 0.3 Å. These results show that iMD-VR docking is an effective tool for (re)creating stable binding poses of ligands of the SARS-CoV-2 Mpro. With further development of the protocols described here, iMD-VR shows promise as a tool for predicting binding poses of novel ligands where no known experimental complex exists.

3.1.2. Docking X77 into apo SARS-CoV-2 Mpro. Crystal structures of apo enzymes (i.e., structures with no ligand bound) are sometimes not suitable for predicting structures of protein−ligand complexes (e.g., if the protein is in a different conformation, or the binding site is occluded).[44] We tested the use of an apo structure for predicting X77 binding using iMD-VR docking. iMD-VR has the advantage (compared to many other docking approaches) that both the ligand and protein are flexible and can therefore adapt their structures to each other, more so if further noninteractive dynamics is run to relax the structures (as here). For apo SARS-CoV-2 Mpro, two iMD-VR protocols for docking X77 were tested: one aimed on finding a low RMSD, the other on forming two important hydrogen bonds (shown in Figure 1). When the apo structure was viewed, a challenge became apparent: the active site loop containing residues 41 to 54 (highlighted in orange in Figure 1) is in a different position to the cognate structure, such that Tyr54 occupies part of the X77 binding pocket. Molecular dynamics simulations of the apo structure showed that this loop exhibits a high degree of flexibility, especially compared to X77-complexed Mpro (Figure S4); due to this higher flexibility, we captured a state where Tyr54 is blocking the pocket, hence X77 could not be fully replaced (Figure 3). We

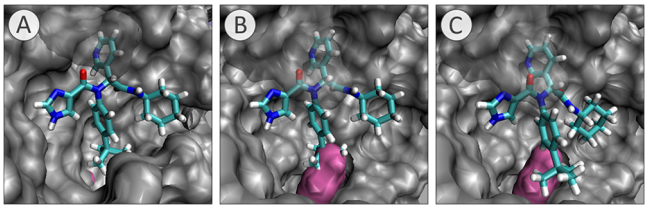


Figure 3. X77 binding poses overlaid on SARS-CoV-2 Mpro structures. (A) Crystal structure of X77 bound to the SARS-CoV-2 Mpro (PDB code: 6W63). Tyr54 (highlighted in pink) is buried in the protein structure. (B) X77 from the cognate crystal complex superimposed on the apo structure, i.e., the protein structure used in the two apo X77 docking protocols (PDB code: 6M03). Tyr54 (highlighted in pink) has shifted so that it is no longer buried and is occupying the binding pocket, causing a steric clash with the *tert*-butylbenzene group of X77. (C) X77 from a representative iMD-VR docked complex taken from 'apo contact protocol (X77)', showing that X77 could not be docked correctly due to the presence of Tyr54 in the binding pocket (highlighted in pink).

tested whether the iMD-VR apo-docked structures relaxed towards the crystal structure of the complex with additional MD. With re-equilibration of the system after iMD-VR docking, Tyr54 spontaneously moved back after X77 docking. Furthermore, the 41 to 54 loop exhibited lower flexibility than the complexed form of Mpro (Figure S4). Despite this initial steric block, the ligand also shifted into the binding pocket and formed the crystallographic binding pose (Figure 2b and 2c). Encouragingly, this suggests that the apo structure can be used for docking small ligands in iMD-VR, as long as the structure is allowed to relax.

RMSD Analysis. Not surprisingly, docking to the apo structure (using either protocol) results in higher RMSDs than docking to the cognate ligand complex structure; even with the structures which relaxed closer to the reference pose, the ligand remained less consistently positioned during subsequent MD (Figure 2). However, for both protocols, at least one structure was found with an average X77 RMSD lower than 2.5 Å. For reference, 10 ns of MD simulation of the unperturbed liganded structure (prior to iMD-VR) yielded an average RMSD of 1.7 Å. Similarly, the lowest average X77 RMSD for iMD-VR docking to the complexed structure for the cognate RMSD protocol (X77) was 1.6 Å. In comparison, the lowest average X77 RMSD for the apo RMSD protocol (X77) was 1.8 Å, and the lowest for the apo contact protocol (X77) was 2.1 Å. Although the first protocol produced a pose with lower average RMSD during subsequent MD, overall, there was more overall

https://dx.doi.org/10.1021/acs.jcim.0c01030
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

variation in the X77 position (Figure 2): The total average RMSD for the apo RMSD docking protocol (X77) was 3.6 Å (σ 1.9 Å), while the total average RMSD for the apo contact docking protocol (X77) was 3.2 Å (σ 1.5 Å). Interactions may be captured by RMSD when docking to cognate protein structures but this may not be the case for apo binding pockets due to structural differences (Figure 3). Therefore, hydrogen bonding analysis and additional evaluation metrics were employed to test the quality of these structures, discussed below.

Hydrogen Bond Analysis. Figure 4 shows the evolution of the two important hydrogen bonds in 10 ns validation MD simulations for both apo X77 protocols. For the apo RMSD protocol (X77), runs 1 and 3 show the poorest hydrogen bonding and highest X77 RMSD of the five runs (Figure 4b); it would appear docking according to minimizing RMSD did not always preserve these hydrogen bonds. During iMD-VR docking, a hydrogen bond did not form in run 2 of the apo contact protocol (X77) and rarely in subsequent MD simulation (Figure 4 c). This resulted in an X77 RMSD which is both higher and prone to large fluctuations (Figure 2c). However, these two contacts were well preserved in the other four (more stable) runs. RMSD alone may not yield structures with important binding interactions preserved (especially postminimization). Therefore, users of iMD-VR should consider how well formed contacts are, especially when the protein structure is apo or otherwise noncognate.

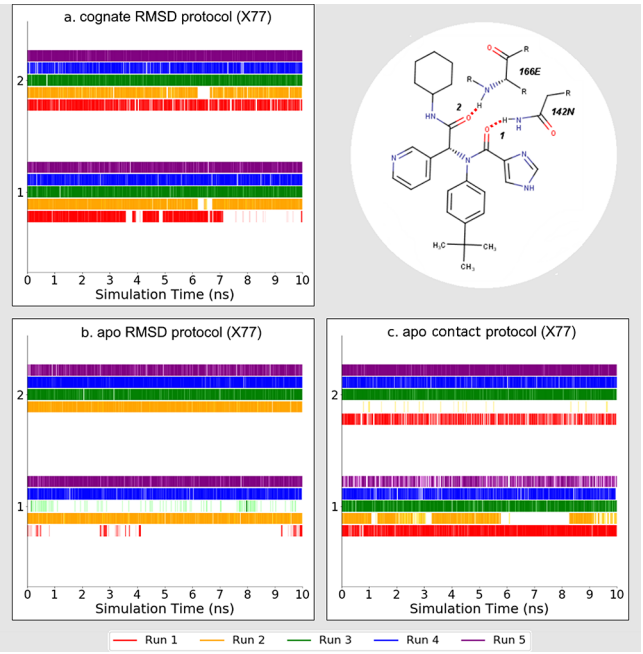


Figure 4. iMD-VR docked X77 hydrogen bonding during 10 ns of validation MD. Results from iMD-VR X77 docking from (a) the cognate Mpro structure using the 'cognate RMSD protocol (X77)', (b) the apo Mpro structure using the 'apo RMSD protocol (X77)', and (c) the apo Mpro structure using the 'apo contact protocol (X77)'. Each graph shows the presence of two hydrogen bonds over 10 ns of MD simulations performed on five representative structures taken from each protocol. Bold lines with a solid color indicate that a hydrogen bond is present in that snapshot. Fainter lines indicate that a hydrogen bond is not present, but the interatomic distance between the donor and acceptor atom is still less than 4 Å, with an applied gradient to become lighter as the atoms move further away. Therefore, faint lines which are harder to distinguish denote only slight displacement.

Additional Analysis. When a reference (e.g., crystal) structure is known, RMSD provides an objective value to rank iMD-VR generated poses; however, it has known limitations.[45] Therefore, users should pay careful attention to interactions in generating structures by iMD-VR docking. We tested other analyses of the iMD-VR generated poses. Figure S2 shows the fraction of contacts recreated for each of the five runs for each of the three docking experiments. Similarly, Figures S3−S7 show the root mean square fluctuation (RMSF) of the five active site protein loops during MD simulations with no backbone restraints. Generally, good performance in these metrics is consistent with the RMSD and hydrogen bonding data. However, while this lends confidence to the iMD-VR generated structures, the protocols suggested here are not an exhaustive list; users should pay attention to both protein−ligand interactions and dynamics when evaluating iMD-VR docked structures. A scoring function or free energy method could be used to evaluate iMD-VR generated poses.[46,47]

3.1.3. Discussion. At least one run from each docking experiment was deemed successful based on good performance across the evaluation metrics (including those discussed in the S1 Text). All of the poses generated by iMD-VR docking for each of the three protocols are given in the S2 and S3 Files. Overall, iMD-VR docking successfully recreated SARS-CoV-2 Mpro complexes. Docking to an apo structure gave complexes with higher RMSDs but that were still usefully close to the crystallographically observed X77 complex. Based on these results, we make general recommendations for using iMD-VR docking for protein−ligand complexes. RMSD has some limitations,[45] chiefly, it only measures how well one structure superimposes on another, and not whether binding interactions are present. Users should focus on forming favorable interactions (such as hydrogen bonds) when docking with iMD-VR. Docking to apo proteins is more challenging, which is not surprising.[44] Apo proteins exhibit different dynamical behavior from complexes (Figures S3−7), so it may be harder to "see" a solution if the binding pocket has changed shape (for both humans and docking algorithms alike). With flexible docking in iMD-VR, even though the apo structure did not allow complete binding of X77 (Figure 3), an improved docking pose was found by relaxing the system with further MD. iMD-VR can be used to explore protein−ligand binding interactions in a fully flexible system, combining human chemical knowledge and spatial intuition work with rigorous MD simulation to find stable complexes.

3.2. iMD-VR Docking of an Oligopeptide Substrate.

3.2.1. Docking Oligopeptide into Complexed SARS-CoV Mpro. A particular advantage of iMD-VR is in the docking of large, flexible molecules because the user can manipulate both their structures and that of the protein. Conventional docking methods often perform well when the ligand has 10 or fewer rotatable bonds but fail when the complexity of the ligand structure increases.[35] iMD-VR docking, on the other hand, allows fully flexible manipulation of large molecules with many rotatable bonds. We tested this approach for an oligopeptide bound to the original SARS Mpro H41A mutant, for which a crystal structure is available (PDB code: 2Q6G).[17] In this crystal structure, the catalytic histidine is replaced by alanine, rendering the enzyme inactive and allowing crystallization of a substrate complex. A 10 ns MD simulation of this crystal structure (the 'substrate reference simulation') was performed

https://dx.doi.org/10.1021/acs.jcim.0c01030
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

to compare with structures generated by iMD-VR. The RMSD of the oligopeptide in this simulation was 3.7−5.6 Å.

Two docking protocols were tested for the oligopeptide substrate with the SARS-CoV Mpro H41A mutant: in the first, restraints were applied to the protease backbone (Restrained Backbone Protocol (Oligopeptide)); in the second, the protease was fully flexible (Fully Flexible Protocol (Oligopeptide)). In both protocols, the aim was to reform 14 hydrogen bonds between the protease and substrate that are observed in the crystal structure. Docking was performed 5 times following each protocol, making a total of 10 iMD-VR oligopeptide docking experiments. From each, a structure in which hydrogen bonds 8 and 10 are formed (the oxyanion hole, excluding hydrogen bond 9 because it is not often observed in the substrate reference simulation) was extracted, and a 10 ns MD simulation was run for each of these docked structures. These interactions (shown in Figure 1) were chosen because they play an integral part in the mechanism of proteolytic cleavage, stabilizing the tetrahedral intermediate by interaction with the carbonyl oxygen of the substrate's scissile bond.[48] Since it is not guaranteed that a structure with a low RMSD (i.e. a structure close to the crystal structure) contains these interactions, we chose structures based on formation of these important hydrogen bonds, to ensure iMD-VR generated structures were similar to the crystal structure.

As a test, substrate docking was also performed following the Lowest RMSD Protocol, similar to the Cognate RMSD Protocol (X77) (Section 2.2.1), where the focus was placing the oligopeptide atoms directly on the trace atoms. However, the docked structures with this protocol did not resemble the crystal structure. While this protocol was successful for docking small molecules (including for the small ligand docking to the complexed SARS-CoV-2 Mpro), we do not recommend this protocol for docking large, flexible molecules. Results from this docking protocol are detailed in the S1 Text.

RMSD Analysis. RMSD analysis of the substrate in MD simulations of all the docked structures showed that the oligopeptide remains within 3.5−6.5 Å of the crystal structure (Figure S8). This value is similar to the substrate RMSD values from the substrate reference simulation (3.7−5.6 Å, Figure S8), indicating that the iMD-VR docked structures are similar to the crystal structure and are also stable in MD simulations. The substrste residues with RMSDs above 5.6 Å are the terminal residues which are solvent exposed, less tightly bound, and more mobile (see Additional Analysis).

Hydrogen Bond Analysis. Structural determination of the SARS-CoV Mpro H41A mutant complex with the 11-mer oligopeptide substrate by Xue et al. revealed 13 hydrogen bonds important for substrate binding.[17] These 13 hydrogen bonds, plus an additional hydrogen bond between Ser144 and P1-Gln that is part of the oxyanion hole, were analyzed in each simulation of the SARS-CoV docked structures. Figures S9 and S10 show that the 14 hydrogen bonds are present throughout the 10 ns MD simulations. The hydrogen bonds reformed in all these simulations are very similar to those observed in the substrate reference simulation. It is encouraging that in all the simulations of iMD-VR docked structures, the oxyanion hole interactions are present.

Additional Analysis. RMSF analysis of the substrate iMD-VR docked structures showed that the P1-Gln residue (that is involved in proteolytic cleavage, and closest to the catalytic dyad) is tightly bound and fluctuates by less than 0.5 Å (Figure S16b,c). In contrast, the substrate's terminal residues (P6-Thr and P5′-Lys) are much more mobile, with RMSF values ranging between 0.8−1.3 Å and 0.8−2.5 Å, respectively (oligopeptide sequence shown in Section 2.3.1). This is not surprising because they are not involved with the 14-hydrogen bond network and not as deeply buried as those in the active site of the protease. Similar behavior is observed in the substrate reference simulation (RMSF of the P1-Gln residue below 0.5 Å and RMSFs for P6-Thr and P5′-Lys of 1 and 2.5 Å respectively, Figure S16a).

Altogether, this shows iMD-VR docking following both the Restrained Backbone Protocol (Oligopeptide) and the Fully Flexible Protocol (Oligopeptide) produces stable structures, which are structurally and dynamically similar to MD simulations based directly on the crystal structure. This demonstrates the power of iMD-VR docking for flexible ligands, including peptide substrates.

3.2.2. Docking Oligopeptide into Two Structures of SARS-CoV-2 Mpro. We then tested the Restrained Backbone Protocol (Oligopeptide) and the Fully Flexible Protocol (Oligopeptide) for iMD-VR docking of the substrate 5 times each into the active site of the 'apo' and 'inhibitor complexed' SARS-CoV-2 Mpro, a total of 20 VR docking simulations. While the SARS-CoV-2 Mpro has a very similar sequence to the SARS-CoV Mpro (96% sequence identity),[5] the structures are slightly different. Figure 5 shows the SARS-CoV Mpro and both SARS-CoV-2 Mpro starting structures for iMD-VR simulations overlaid based on secondary structure. It shows that there are some small structural differences. As in the SARS-CoV Mpro VR docking, a structure in which the oxyanion hole interactions had been reformed was extracted from each iMD-VR simulation. Ten ns of implicit solvent MD

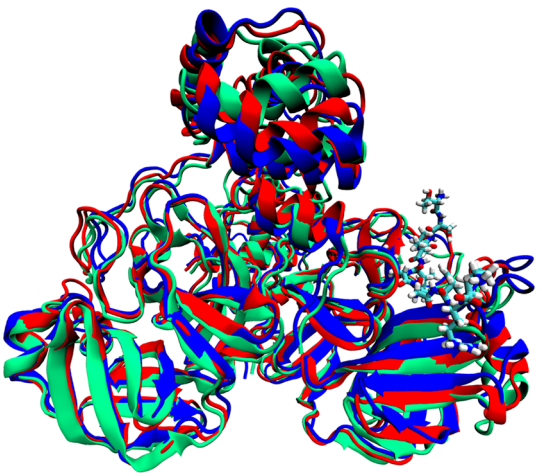


Figure 5. The three Mpro structures used as input for iMD-VR substrate docking, overlaid based on secondary structure. The structures are the SARS-CoV Mpro (red), the apo SARS-CoV-2 Mpro (blue), and the inhibitor-complexed SARS-CoV-2 Mpro (green). The substrate is also shown on the right, bound to the SARS-CoV structure. The two SARS-CoV-2 structures are aligned against the secondary structure of the SARS-CoV structure, yielding an RMSD of 3.17 and 3.29 Å for the apo and inhibitor-complexed structures, respectively. For consistency, the structures were also aligned on the secondary structure of the apo Mpro and the inhibitor-complexed Mpro. The RMSD of SARS-CoV Mpro and inhibitor-complexed Mpro aligned on the apo structure are 3.17 and 3.19 Å, respectively, and the RMSD of the SARS-CoV Mpro and the apo Mpro aligned on the ligand-complexed structure are 3.29 and 3.19 Å, respectively.

https://dx.doi.org/10.1021/acs.jcim.0c01030
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

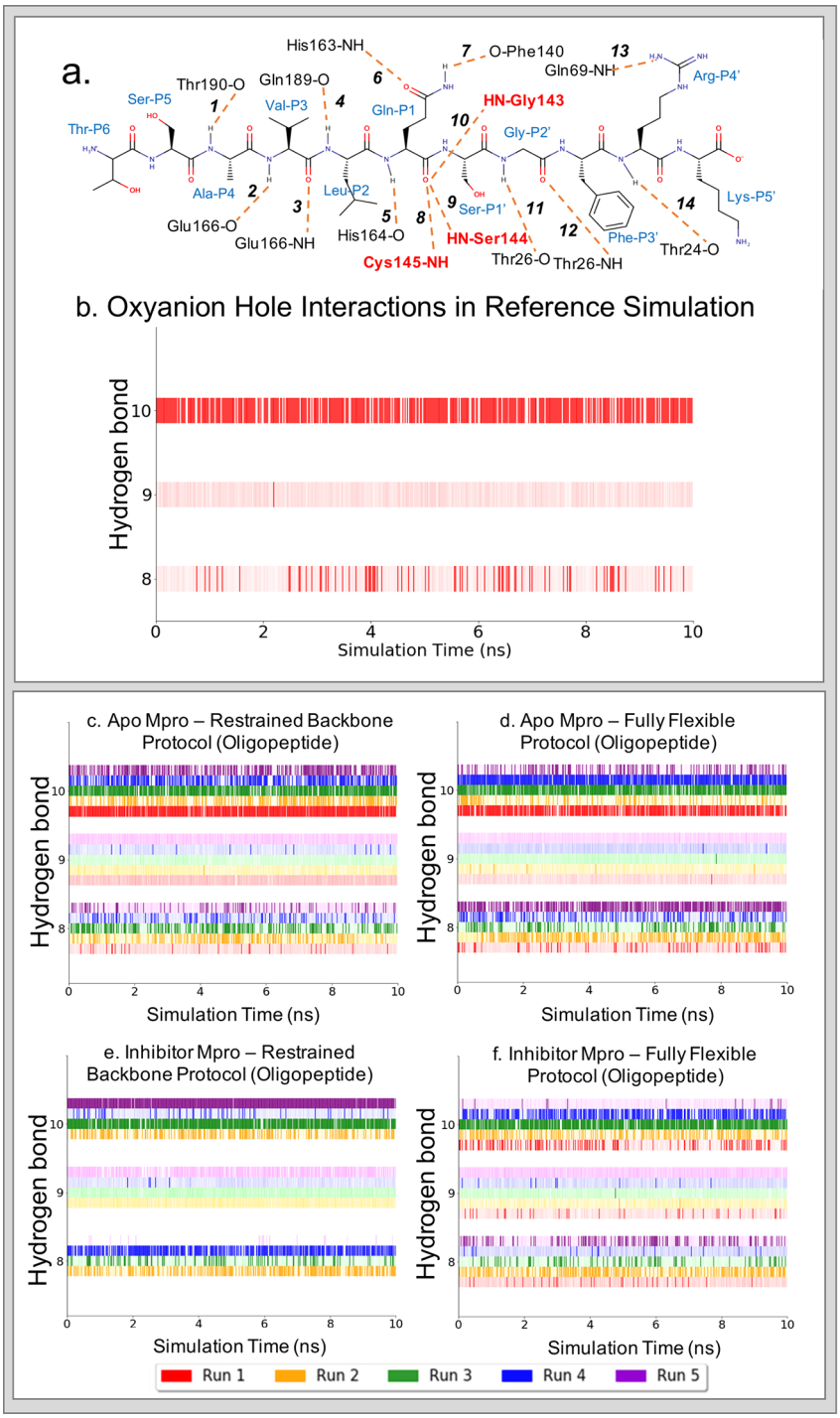


Figure 6. Important hydrogen bonds present throughout MD simulation of oligopeptide substrate-Mpro complexes. (a) The sequence of the 11-mer oligopeptide substrate and the 14 hydrogen bonds which were considered in analysis of VR generated structures. The oxyanion hole interactions are highlighted in bold red. (b) Oxyanion hole interactions (shown on the *y*-axis as hydrogen bonds 8, 9, and 10) during the 'substrate reference simulation'. The solid color represents the hydrogen bond being present, and the lighter/translucent color represents the donor and acceptor atoms being within 4 A of each other, with an applied gradient to become lighter as the atoms move further away. The oxyanion hole interactions are also presented for the simulations of the iMD-VR docked structures of the substrate to (c) the apo SARS-CoV-2 Mpro following the Restrained Backbone Protocol (Oligopeptide) and (d) following the Fully Flexible Protocol (Oligopeptide) and of the substrate to (e) the inhibitor SARS-CoV-2 Mpro following the Restrained Backbone Protocol (Oligopeptide) and (f) the Fully Flexible Protocol (Oligopeptide).

were performed for each of the docked SARS-CoV-2 Mpro structures, giving a total of 20 MD simulations.

RMSD Analysis. RMSD analysis shows that the structures built by iMD-VR docking are close to the crystal structure. The substrate heavy atoms remain within 3−6 Å of the crystal structure coordinates (Figure S8) in all the MD simulations of iMD-VR docked complexes. Visual inspection of the trajectories showed that the substrate remains bound to the protease in all simulations and that the cases where the RMSD is higher than ∼5 Å are due to the mobility of the terminal

https://dx.doi.org/10.1021/acs.jcim.0c01030
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

residues of the oligopeptide (see Additional Analysis). This is similar to the RMSD analysis of the reference simulation (Figure S8a) and also to simulations of the SARS-CoV docked structures (Figure S8b−g).

Hydrogen Bond Analysis. As noted above, the SARS-CoV-2 Mpro has a very high sequence similarity to the SARS-CoV Mpro (96% and contains the same binding residues involved in the 14 hydrogen bonds). The same hydrogen bonds are likely to be involved in substrate binding and therefore are considered here. Figure 6c−f shows the oxyanion hole interactions present during the 10 ns MD simulations of the VR docked SARS-CoV-2 structures.

Docking to the apo SARS-CoV-2 Mpro generally formed the oxyanion hole interactions using either oligopeptide protocol (Figure 6c,d). Similarly, hydrogen bonds 2, 3, 6, 12, and 14 were well maintained throughout simulations of all the iMD-VR docked poses, showing that a common pose is obtained by iMD-VR docking (Figures S12 and S13). The simulations of the inhibitor SARS-CoV-2 Mpro docked structures showed that, while all hydrogen bonds are not always present (Figures S14 and S15), they still show most of the 14 hydrogen bonds have formed. Among the most well maintained in all simulations built from inhibitor complexes are hydrogen bonds 2, 3, 12, and 14. This is similar to the apo structures. Furthermore, the oxyanion hole interactions are present throughout the simulations (Figure 6e,f, except inhibitor Mpro - Restrained Backbone Protocol (Oligopeptide), Run 1).

Finally, hydrogen bonds 4 and 13 are the only two in this set that were not reformed in any iMD-VR docking attempts with either the SARS-CoV or SARS-CoV-2 Mpro. This could be because these two hydrogen bonds do not involve the enzyme backbone. Note that hydrogen bond 13 was also not observed in the substrate reference simulation (Figure S9).

Additional Analysis. RMSF analysis of all 20 MD simulations of the SARS-CoV-2 iMD-VR docked structures showed that the α-carbon of the P1-Gln residue fluctuates by no more than 0.5 Å, whereas the terminal residues P6-Thr and P5′-Lys are more mobile (RMSF between 0.5−3 Å, Figure S16d−g). This is expected, because the ends of the substrate are bound more weakly than the central residues, as also observed for SARS-CoV Mpro (above) and the substrate reference simulation. Altogether, this shows that substrate complexes built by iMD-VR docking are stable and, e.g., will be suitable for further simulation of Mpro dynamics, binding, and mechanism.

3.2.3. Discussion. Docking with iMD-VR was successful in both (i) reproducing the crystal structure of an oligopeptide substrate bound to SARS-CoV Mpro and also (ii) in creating structures of SARS-CoV-2 Mpro with this substrate. This docking was assisted by experience gained from docking the inhibitor and by chemical and spatial intuition. We tested different docking protocols, e.g., testing the effects of protein backbone restraints. Overall, the docked SARS-CoV Mpro structures reformed more of the hydrogen bonds in the 14 hydrogen bond network than in the SARS-CoV-2 structures. This is probably due to the fact that docking to the original SARS protease used the cognate peptide complex structure, which is in a configuration to accommodate oligopeptide binding and so will form these hydrogen bonds more easily than apo enzyme structures or structures with small molecule inhibitors. Nonetheless, the substrate RMSD in simulations of the apo and inhibitor SARS-CoV-2 Mpro docked structure stayed in the same range as the substrate reference simulation. The iMD-VR docked structures also consistently reproduce the majority of the hydrogen bonds with the substrate, particularly the important oxyanion hole interactions and hydrogen bonds 2, 3, 12, and 14. It is interesting that the iMD-VR generated SARS-CoV Mpro oligopeptide complexes were structurally very similar to both the apo and inihibitor complexed SARS-CoV-2 Mpro. High-throughput screening approaches are known to be less successful at docking to apo proteins than to holo protein targets,[44] and it is promising to see that using iMD-VR produces similar docked structures to both apo and inhibitor Mpros. The two oligopeptide docking protocols yielded reproducible docked structures, which indicates that SARS-CoV-2 Mpro substrate binding is similar to SARS-CoV-1. Future work will further analyze these models.

There is no significant discernible difference in the docked structures generated from either of the oligopeptide docking protocols. We therefore recommend that expert iMD-VR users should follow the Fully Flexible Protocol (Oligopeptide), and novice iMD-VR users should follow the Restrained Backbone Protocol (Oligopeptide), so as to not accidentally disrupt the protein secondary structure during docking.

3.3. Software Availability. We used NarupaXR to perform the iMD-VR simulations described in this work. The software is available under a GPL open-source license at https://irl.itch.io/narupaxr (note: trace atoms are not enabled in this version). An updated version of Narupa (Narupa iMD), which allows cloud-hosted single- or multiuser iMD-VR sessions, can be accessed at https://app.narupa.xyz. This new version is available, under the same open-source license as its predecessor, at https://narupa.readthedocs.io. It utilizes a Python framework, making customization straightforward, and also offers improved visuals. The input files for the iMD-VR simulations presented in this work are available in the S4 File for use with the previous version of Narupa. Additionally, the S5 File contains equivalent input files for the new version, Narupa iMD. The reader can access a Narupa iMD cloud-hosted sole- or multiuser session at https://app.narupa.xyz/. We believe that this will be widely useful for structure-based design for SARS-CoV-2 and for education on the structure, mechanisms, and inhibition of this viral protein. Further, iMD-VR can be straightforwardly applied and will be useful for other COVID-19 targets and in other areas.

4. CONCLUSIONS

In this work, we have demonstrated that iMD-VR is an effective tool for generating structures of complexes of the SARS-CoV-2 Mpro, using an inhibitor and an oligopeptide substrate. In previous work,[34] we have shown that even novice users can quickly generate structures in good agreement with crystal structures of protein-ligand complexes, with minimal guidance. With a little practice, researchers can dock ligands to proteins in a few minutes. Here, we show that iMD-VR docking is effective for building and modeling substrate and inhibitor complexes of SARS-CoV-2 Mpro and present our best practices for iMD-VR docking in both cases. Structures constructed by iMD-VR docking reproduce the key structural and binding motifs found in reference crystallographic structures. As expected,[44] iMD-VR docking performs better on cognate protein models, i.e., redocking to an observed complex, than for docking to apo forms of the protein. However, the results with apo and other protein structures are also in good agreement with experiment and stable in MD. Reactive enzyme complexes are difficult to study experimen-

https://dx.doi.org/10.1021/acs.jcim.0c01030
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

tally, so computational models are often necessary to make predictions. Predicting these structures with standard docking methods is challenging, however, particularly in the case of oligopeptide substrates that are larger and more conformationally complex than small molecule drug-like compounds.[46,47] Our models were built first by recreating related experimental structures and showed good agreement with them. iMD-VR shows promise as a tool to study substrate binding because it allows for full flexibility of the enzyme and the substrate. Their conformational behavior and interactions are modeled by a detailed atomistic force field, allowing the system to respond in a realistic manner. Structures created by iMD-VR may then be refined by relaxation with standard MD simulations.

iMD-VR users should pay attention to forming specific binding interactions. Further MD simulations of docked complexes provide a test of stability, as here. Proteins are known to undergo conformational changes upon binding and, although some docking methods attempt to account for this,[49] high-throughput screening methods often treat the protein or ligand as rigid out of practical necessity; simplified interaction potentials are also typically used in automated docking. This, and imperfections in scoring functions,[50] limit accuracy. Our iMD-VR framework will complement high-throughput docking, e.g., to test structures, taking advantage of the protein and ligand flexibility in iMD-VR, and human intuition, to refine these structures. With further development, iMD-VR can also be used as a predictive tool to model binding of potential inhibitors or substrates, for example, by integrating a scoring function or free energy calculation into these protocols. Any ligand can be modeled, given appropriate molecular mechanics parameters. The accuracy of the simulation is limited by the accuracy of the underlying MD simulation. In this work, we used an implicit solvent model, which speeds the simulations and simplifies the binding problem. However, implicit solvent models have limitations for modeling protein structure and binding. Explicit solvent models may improve accuracy, particularly in modeling binding pocket water interactions but would increase the computational cost and the complexity of navigating and manipulating the system.

IMD-VR allows the user to "step inside" the SARS-CoV-2 Mpro and manipulate its structure and interactions—and dynamics—in atomic detail. This allows the user to perceive ligand binding in three dimensions and to interact with molecules in an intuitive way, similar to how people interact with macroscopic objects. It is a better interface for molecular modeling tasks than, e.g., traditional screen/mouse-based methods.[29] We believe that such emerging interactive VR technologies will find widespread application in structure-based drug design,[34] in other molecular modeling, design and simulation applications, and in education.[36] IMD-VR should be directly useful in ongoing efforts to develop Mpro inhibitors as potential antivirals against COVID-19. They could also find application in efforts to crowd-source the problem.[51]

Narupa, the software framework that we have developed, is open source and uses commodity VR hardware, which is widely accessible.[29,34,36,52,53] It can therefore readily be used. We make our Mpro structures and simulations here freely available (Files S2, S3, S4, and S5); these can be used for ligand discovery and structure−activity studies for SARS-CoV-2. They run straightforwardly on a suitably configured laptop computer. Alternatively, these simulations are available via the cloud, such that anyone with a compatible VR headset can easily access them. We note that the cloud also offers the possibility of virtual collaboration and distributed working via iMD-VR. We encourage others to use these potentially transformative tools and believe that they will be widely useful.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.0c01030.

S1 text, "interactive molecular dynamics in virtual reality (iMD-VR) is an effective tool for flexible substrate and inhibitor docking to the SARS-CoV-2 main protease" (PDF)

S2 file, supplementary structure files (inhibitor), structures from iMD-VR simulations of inhibitor docking (ZIP)

S3 file, supplementary structures files (oligopeptide) (ZIP)

S4 file, supplementary simulation files (NarupaXR), simulation files for iMD-VR setup of inhibitor and oligopeptide docking, using NarupaXR (ZIP)

S5 file, supplementary simulation files (Narupa iMD), simulation files for iMD-VR setup of inhibitor and oligopeptide docking, using Narupa iMD (ZIP)

■ AUTHOR INFORMATION

Corresponding Authors

David R. Glowacki − *Intangible Realities Laboratory, School of Chemistry, Centre for Computational Chemistry, School of Chemistry, and Department of Computer Science, Merchant Venturers Building,, University of Bristol, Bristol BS8 1TS, United Kingdom;* orcid.org/0000-0002-9608-3845; Email: glowacki@bristol.ac.uk

Adrian J. Mulholland − *Centre for Computational Chemistry, School of Chemistry, University of Bristol, Bristol BS8 1TS, United Kingdom;* orcid.org/0000-0003-1015-4567; Email: adrian.mulholland@bristol.ac.uk

Authors

Helen M. Deeks − *Intangible Realities Laboratory, School of Chemistry, Centre for Computational Chemistry, School of Chemistry, and Department of Computer Science, Merchant Venturers Building,, University of Bristol, Bristol BS8 1TS, United Kingdom*

Rebecca K. Walters − *Intangible Realities Laboratory, School of Chemistry, Centre for Computational Chemistry, School of Chemistry, and Department of Computer Science, Merchant Venturers Building,, University of Bristol, Bristol BS8 1TS, United Kingdom;* orcid.org/0000-0001-7266-2175

Jonathan Barnoud − *Intangible Realities Laboratory, School of Chemistry and Centre for Computational Chemistry, School of Chemistry, University of Bristol, Bristol BS8 1TS, United Kingdom;* orcid.org/0000-0003-0343-7796

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.0c01030

Author Contributions

¶H.M.D. and R.K.W. contributed equally to this work.

Notes

The authors declare no competing financial interest.

https://dx.doi.org/10.1021/acs.jcim.0c01030
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

■ ACKNOWLEDGMENTS

H.M.D. and R.K.W. thank the Engineering and Physical Sciences Research Council (EPSRC) for PhD studentships. A.J.M. and H.M.D. thank the British Society for Antimicrobial Chemotherapy for support (grant number BSAC-COVID-30). J.B. acknowledges funding from the EPSRC (programme grant EP/P021123/1). A.J.M. thanks EPSRC (grant number EP/M022609/1, CCP-BioSim), BrisSynBio, a BBSRC/EPSRC Synthetic Biology Research Centre (Grant Number: BB/L01386X/1), and MRC (MR/T016035/1) for support. A.J.M. and D.R.G. thank Oracle Research (University Partnership Cloud award) for support. D.R.G. acknowledges funding the Royal Society (URF\R\180033), EPSRC (impact acceleration award, institutional sponsorship award, and EP/P021123/1), and the Leverhulme Trust (Philip Leverhulme Prize).

■ REFERENCES

(1) Fehr, A. R.; Perlman, S. Coronaviruses: An Overview of Their Replication and Pathogenesis. Maier, H. J., Bickerton, E., Britton, P., Eds.; In *Coronaviruses: Methods and Protocols [Internet]*; Springer New York: New York, NY, 2015; pp 1−23, DOI: 10.1007/978-1-4939-2438-7_1.

(2) Wang, L.-F.; Zhengli, S.; Shuyi, Z.; Hume, E. F.; Peter, D.; Bryan, T. E. Review of Bats and SARS. *Emerging Infect. Dis.* **2006**, *12* (12), 1834.

(3) Chen, Y.; Liu, Q.; Guo, D. Emerging coronaviruses: Genome structure, replication, and pathogenesis. *J. Med. Virol.* **2020**, *92* (4), 418−23.

(4) 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.

(5) Morse, J. S.; Lalonde, T.; Xu, S.; Liu, W. R. Learning from the Past: Possible Urgent Prevention and Treatment Options for Severe Acute Respiratory Infections Caused by 2019-nCoV. *ChemBioChem* **2020**, *21* (5), 730−8.

(6) Kiemer, L.; Lund, O.; Brunak, S.; Blom, N. Coronavirus 3CLproproteinase cleavage sites: Possible relevance to SARS virus pathology. *BMC Bioinf.* **2004**, *5* (1), 72.

(7) Chuck, C.-P.; Chong, L.-T.; Chen, C.; Chow, H.-F.; Wan, DC-C; Wong, K.-B. Profiling of Substrate Specificity of SARS-CoV 3CLpro. *PLoS One* **2010**, *5* (10), e13197.

(8) 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* (5626), 1763.

(9) Grottesi, A.; Beš ker, N.; Emerson, A.; Manelfi, C.; Beccari, A. R.; Frigerio, F.; Lindahl, E.; Cerchia, C.; Talarico, C. Computational Studies of SARS-CoV-2 3CLpro: Insights from MD Simulations. *Int. J. Mol. Sci.* **2020**, *21* (15), 5346.

(10) Chen, S.; Jonas, F.; Shen, C.; Higenfeld, R. Liberation of SARS-CoV main protease from the viral polyprotein: N-terminal autocleavage does not depend on the mature dimerization mode. *Protein Cell* **2010**, *1* (1), 59−74.

(11) Marra, M. A.; Jones, S. J. M.; Astell, C. R.; Holt, R. A.; Brooks-Wilson, A.; Butterfield, Y. S. N.; Khattra, J.; Asano, J. K.; Barber, S. A.; Chan, S. Y.; Cloutier, A.; Coughlin, S. M.; Freeman, D.; Girn, N.; Griffith, O. L.; Leach, S. R.; Mayo, M.; McDonald, H.; Montgomery, S. B.; Pandoh, P. K.; Petrescu, A. S.; Robertson, A. G.; Schein, J. E.; Siddiqui, A.; Smailus, D. E.; Stott, J. M.; Yang, G. S.; Plummer, F.; Andonov, A.; Artsob, H.; Bastien, N.; Bernard, K.; Booth, T. F.; Bowness, D.; Czub, M.; Drebot, M.; Fernando, L.; Flick, R.; Garbutt, M.; Gray, M.; Grolla, A.; Jones, S.; Feldmann, H.; Meyers, A.; Kabani, A.; Li, Y.; Normand, S.; Stroher, U.; Tipples, G. A.; Tyler, S.; Vogrig, R.; Ward, D.; Watson, B.; Brunham, R. C.; Krajden, M.; Petric, M.; Skowronski, D. M.; Upton, C.; Roper, R. L. The Genome Sequence of the SARS-Associated Coronavirus. *Science* **2003**, *300* (5624), 1399.

(12) Krichel, B.; Falke, S.; Hilgenfeld, R.; Redecke, L.; Uetrecht, C. Processing of the SARS-CoV pp1a/ab nsp7−10 region. *Biochem. J.* **2020**, *477* (5), 1009−19.

(13) Hegyi, A.; Ziebuhr, J. Conservation of substrate specificities among coronavirus main proteases. *J. Gen. Virol.* **2002**, *83* (3), 595−9.

(14) Fan, K.; Wei, P.; Feng, Q.; Chen, S.; Huang, C.; Ma, L.; Lai, B.; Pei, J.; Liu, Y.; Chen, J. Biosynthesis, purification, and substrate specificity of severe acute respiratory syndrome coronavirus 3C-like proteinase. *J. Biol. Chem.* **2004**, *279* (3), 1637−1642.

(15) Liang, P.-H. Characterization and Inhibition of SARS-Coronavirus Main Protease. *Curr. Top. Med. Chem.* **2006**, *6* (4), 361−376.

(16) Graziano, V.; McGrath, W. J.; Yang, L.; Mangel, W. F. SARS CoV Main Proteinase: The Monomer−Dimer Equilibrium Dissociation Constant. *Biochemistry.* **2006**, *45* (49), 14632−41.

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

(18) Barnard, D. L.; Kumaki, Y. Recent developments in anti-severe acute respiratory syndrome coronavirus chemotherapy. *Future Virol.* **2011**, *6* (5), 615−31.

(19) Haagmans, B. L.; Osterhaus, A. D. M. E. Coronaviruses and their therapy. *Antiviral Res.* **2006**, *71* (2), 397−403.

(20) Chen, Y.; CPB, Y.; Wong, K. Y. Prediction of the SARS-CoV-2 (2019-nCoV) 3C-like protease (3CLpro) structure: virtual screening reveals velpatasvir, ledipasvir, and other drug repurposing candidates [version 2; peer review: 3 approved]. *F1000Research* **2020**, *9*, 129.

(21) Dai, W.; Zhang, B.; Jiang, X.-M.; Su, H.; Li, J.; Zhao, Y.; Xie, X.; Jin, Z.; Peng, J.; Liu, F.; Li, C.; Li, Y.; Bai, F.; Wang, H.; Cheng, X.; Cen, X.; Hu, S.; Yang, X.; Wang, J.; Liu, X.; Xiao, G.; Jiang, H.; Rao, Z.; Zhang, L.-K.; Xu, Y.; Yang, H.; Liu, H. Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease. *Science* **2020**, *368* (6497), 1331−5.

(22) Ma, C.; Sacco, M. D.; Hurst, B.; Townsend, J. A.; Hu, Y.; Szeto, T.; Zhang, X.; Tarbet, B.; Marty, M. T.; Chen, Y.; Wang, J. Boceprevir, GC-376, and calpain inhibitors II, XII inhibit SARS-CoV-2 viral replication by targeting the viral main protease. *Cell Research.* **2020**, *30* (8), 678−92.

(23) 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 Mpro from COVID-19 virus and discovery of its inhibitors. *Nature* **2020**, *582* (7811), 289−93.

(24) Irwin, J. J.; Shoichet, B. K. ZINC–a free database of commercially available compounds for virtual screening. *J. Chem. Inf Model.* **2005**, *45* (1), 177−82.

(25) Liu, X.; Wang, X.-J. Potential inhibitors for 2019-nCoV coronavirus M protease from clinically approved medicines. *bioRxiv*; 2020;2020.01.29.924100, DOI: 10.1101/2020.01.29.924100.https://www.biorxiv.org/content/10.1101/2020.01.29.924100v1 (accessed 2020-11-02).

(26) 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* (8), 2000028.

(27) Ge, Y.; van der Kamp, M.; Malaisree, M.; Liu, D.; Liu, Y.; Mulholland, A. J. Identification of the quinolinedione inhibitor binding site in Cdc25 phosphatase B through docking and molecular dynamics simulations. *J. Comput.-Aided Mol. Des.* **2017**, *31* (11), 995−1007.

(28) Amaro, R. E.; Baudry, J.; Chodera, J.; Demir, Ö; McCammon, J. A.; Miao, Y.; Smith, J. C. Ensemble Docking in Drug Discovery. *Biophys. J.* **2018**, *114* (10), 2271−8.

(29) O'Connor, M.; Deeks, H. M.; Dawn, E.; Metatla, O.; Roudaut, A.; Sutton, M.; Thomas, L. M.; Glowacki, B. R.; Sage, R.; Tew, P. et al.; et al. Sampling molecular conformations and dynamics in a

multiuser virtual reality framework. *Science Advances.* **2018**, *4* (6), eaat2731.
(30) Kingsley, L. J.; Brunet, V.; Lelais, G.; McCloskey, S.; Milliken, K.; Leija, E.; Fuhs, S. R.; Wang, K.; Zhou, E.; Spraggon, G. Development of a virtual reality platform for effective communication of structural data in drug discovery. *J. Mol. Graphics Model.* **2019**, *89*, 234−41.
(31) Cassidy, K. C.; Šefčík, J.; Raghav, Y.; Chang, A.; Durrant, J. D. ProteinVR: Web-based molecular visualization in virtual reality. *PLoS Comput. Biol.* **2020**, *16* (3), e1007747.
(32) Calvelo, M.; Piñeiro, Á; Garcia-Fandino, R. An immersive journey to the molecular structure of SARS-CoV-2: Virtual reality in COVID-19. *Comput. Struct. Biotechnol. J.* **2020**, *18*, 2621−8.
(33) Zhavoronkov, A.; Zagribelnyy, B.; Zhebrak, A.; Aladinskiy, V.; Terentiev, V.; Vanhaelen, Q.; Bezrukov, D.; Polykovskiy, D.; Shayakhmetov, R.; Filimonov, A.; Bishop, M.; McCloskey, S.; Leija, E.; Bright, D.; Funakawa, K.; Lin, Y.-C.; Huang, S.-H.; Liao, H.-J.; Aliper, A.; Ivanenkov, Y. Potential Non-Covalent SARS-CoV-2 3C-like Protease Inhibitors Designed Using Generative Deep Learning Approaches and Reviewed by Human Medicinal Chemist in Virtual Reality. *ChemRxiv [Internet]*; 2020. https://doi.org/10.26434/chemrxiv.12301457.v1 (accessed 2020-11-01).
(34) Deeks, H. M.; Walters, R. K.; Hare, S. R.; O'Connor, M. B.; Mulholland, A. J.; Glowacki, D. R. Interactive molecular dynamics in virtual reality for accurate flexible protein-ligand docking. *PLoS One* **2020**, *15* (3), e0228461.
(35) Pagadala, N. S.; Syed, K.; Tuszynski, J. Software for molecular docking: a review. *Biophys. Rev.* **2017**, *9* (2), 91−102.
(36) Bennie, S. J.; Ranaghan, K. E.; Deeks, H.; Goldsmith, H. E.; O'Connor, M. B.; Mulholland, A. J.; Glowacki, D. R. Teaching Enzyme Catalysis Using Interactive Molecular Dynamics in Virtual Reality. *J. Chem. Educ.* **2019**, *96* (11), 2488−96.
(37) Świderek, K.; Moliner, V. Revealing the molecular mechanisms of proteolysis of SARS-CoV-2 Mpro by QM/MM computational methods. *Chemical Science.* **2020**, *11* (39), 10626−30.
(38) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11* (8), 3696−3713.
(39) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general Amber force field. *J. Comput. Chem.* **2004**, *25* (9), 1157−74.
(40) Onufriev, A.; Bashford, D.; Case, D. A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins: Struct., Funct., Genet.* **2004**, *55* (2), 383−94.
(41) Paasche, A.; Zipper, A.; Schäfer, S.; Ziebuhr, J.; Schirmeister, T.; Engels, B. Evidence for Substrate Binding-Induced Zwitterion Formation in the Catalytic Cys-His Dyad of the SARS-CoV Main Protease. *Biochemistry* **2014**, *53* (37), 5930−46.
(42) 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* (7798), 265−9.
(43) Yang, H.; Wong, M. W. Oxyanion Hole Stabilization by C−H···O Interaction in a Transition State—A Three-Point Interaction Model for Cinchona Alkaloid-Catalyzed Asymmetric Methanolysis of meso-Cyclic Anhydrides. *J. Am. Chem. Soc.* **2013**, *135* (15), 5808−18.
(44) Erickson, J. A.; Jalaie, M.; Robertson, D. H.; Lewis, R. A.; Vieth, M. Lessons in Molecular Recognition: The Effects of Ligand and Protein Flexibility on Molecular Docking Accuracy. *J. Med. Chem.* **2004**, *47* (1), 45−55.
(45) Kufareva, I, Abagyan, R. Methods of Protein Structure Comparison. Orry, A. J. W., Abagyan, R, Eds.; In *Homology Modeling: Methods and Protocols [Internet]*; Humana Press: Totowa, NJ, 2012; pp 231−57 DOI: 10.1007/978-1-61779-588-6_10.
(46) Ciemny, M.; Kurcinski, M.; Kamel, K.; Kolinski, A.; Alam, N.; Schueler-Furman, O.; Kmiecik, S. Protein−peptide docking: opportunities and challenges. *Drug Discovery Today.* **2018**, *23* (8), 1530−7.
(47) Weng, G.; Gao, J.; Wang, Z.; Wang, E.; Hu, X.; Yao, X.; Cao, D.; Hou, T. Comprehensive Evaluation of Fourteen Docking Programs on Protein−Peptide Complexes. *J. Chem. Theory Comput.* **2020**, *16* (6), 3959−69.
(48) Simón, L.; Goodman, J. M. Enzyme Catalysis by Hydrogen Bonds: The Balance between Transition State Binding and Substrate Binding in Oxyanion Holes. The. *J. Org. Chem.* **2010**, *75* (6), 1831−40.
(49) Hart, K. M.; Ho, C. M. W.; Dutta, S.; Gross, M. L.; Bowman, G. R. Modelling proteins' hidden conformations to predict antibiotic resistance. *Nat. Commun.* **2016**, *7* (1), 12965.
(50) Li, J.; Fu, A.; Zhang, L. An Overview of Scoring Functions Used for Protein−Ligand Interactions in Molecular Docking. Interdisciplinary Sciences: Computational. *Interdiscip. Sci.: Comput. Life Sci.* **2019**, *11* (2), 320−8.
(51) Chodera, J.; Lee, A. A.; London, N.; von Delft, F. Crowdsourcing drug discovery for pandemics. *Nat. Chem.* **2020**, *12* (7), 581−581.
(52) Amabilino, S.; Bratholm, L. A.; Bennie, S. J.; Vaucher, A. C.; Reiher, M.; Glowacki, D. R. Training neural nets to learn reactive potential energy surfaces using interactive quantum chemistry in virtual reality. *J. Phys. Chem. A* **2019**, *123* (20), 4486−4499.
(53) Jamieson-Binnie, A. D.; O'Connor, M. B.; Barnoud, J.; Wonnacott, M. D.; Bennie, S. J.; Glowacki, D. R. Narupa IMD: A VR-Enabled Multiplayer Framework for Streaming Interactive Molecular Simulations. In *SIGGRAPH '20: ACM SIGGRAPH 2020 Immersive Pavilion [Internet]*; Association for Computing Machinery: New York, NY, USA, 2020; DOI: 10.1145/3388536.3407891.

https://dx.doi.org/10.1021/acs.jcim.0c01030
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX