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

Helen M. Deeks[a,b,c†], Rebecca K. Walters[a,b,c†], Jonathan Barnoud[a,b], David R. Glowacki[a,b,c*], Adrian J. Mulholland[b*]

*corresponding authors

† These authors contributed equally to this work

(a) Intangible Realities Laboratory, School of Chemistry, Cantock's Close, Bristol, BS8 1TS (b) Centre for Computational Chemistry, School of Chemistry, Cantock's Close, Bristol, BS8 1TS, (c) Department of Computer Science, Merchant Venturers Building, Woodland Road Bristol, BS8 1UB

The main protease (Mpro) of the SARS-CoV-2 virus is one focus of drug development efforts for antivirals to combat COVID-19. Here, we show that interactive molecular dynamics in virtual reality (iMD-VR) is a useful and effective tool for predicting structures of Mpro-ligand complexes. We have recently demonstrated that iMD-VR is an effective method for interactive, flexible docking of small molecule drugs into their protein targets. iMD-VR provides an immersive environment in which users can interact with fully dynamic molecular simulations, and interactively manipulate them to form stable complexes, an important challenge in predicting binding of potential drug leads. Here, we apply this approach to an Mpro inhibitor, and also to create a model of Mpro with an oligopeptide substrate. For the latter, we test against a crystallographic structure of the original SARS Mpro. Docking with iMD-VR gives structures close to experimentally observed (crystal) structures. These docked structures are also tested in MD simulations and are generally found to be stable. We test different protocols for iMD-VR docking, e.g. with and without restraints on protein backbone, and 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 only on use simple metrics (such as RMSD) in creating realistic, stable complexes. We also test the use of apo (uncomplexed) crystal structures for docking and find that they can give good results because of the flexibility and dynamic response allowed by the physically rigorous, atomically detailed simulation

approach of iMD-VR. We make these 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 wide research community working on Mpro and other COVID-19 targets. These should be widely useful in drug development programmes, 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. Coronaviruses of this type can be pernicious as they can spread between numerous 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). As a result of 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 '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 non-structural proteins, using a catalytic Cys145/His41 dyad present in each subunit of the dimer. [8] Both the SARS-CoV and SARS-CoV-2 Mpro are believed to operate on 11 different cleavage sites of the large overlapping polyprotein 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 non-structural proteins generated from proteolytic cleavage play a vital role in viral transcription and replication [15, 16], making the Mpro an essential component for the viral lifecycle. [7, 17-19]

There are currently many efforts globally aimed at identifying promising drug targets for COVID-19 [18, 19]. Molecular simulation and modelling 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 179 SARS-CoV-2 Mpro structures on 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] can identify 'hit' molecules that may bind to and inhibit the enzyme. Several docking studies on SARS-CoV-2 targets have already appeared, including for the 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, can 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 often require further refinement. Recognised limitations of automated docking methods include the approximate nature of energy functions used, limited treatment of solvation, and limited treatment of protein and ligand conformational variability, particularly for large, flexible ligands. These simple methods mean that automated docking can test millions of compounds in a relatively short time but also that the results are prone to error and may identify many false positives.

More sophisticated biomolecular simulation methods, such as molecular dynamics (MD) simulations, 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 armoury of computational methods for drug discovery and development. Furthermore, VR potentially allows for sharing structural information in an intuitive and accessible form, and for distributed, virtual collaboration, e.g. when supported by cloud-based resources. [29] VR frameworks for protein visualisation, such as ProteinVR, are being used for the SARS-CoV-2 Mpro. [30] While VR is undoubtedly a useful tool for visualising 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). [31]

Recently, we have shown that iMD-VR using Narupa provides an effective and accurate approach for recreating known protein-ligand complexes. [31] It allows fully flexible docking through the inclusion of molecular motion in a physically rigorous MD simulation. iMD-VR allows interactive, dynamic, and flexible manipulation of structures, enabling the user to discover favourable binding modes. We applied this framework to several systems: pertinently, two of the three proteins studied were viral enzymes (HIV

protease and influenza neuraminidase) with clinically approved drugs. We showed that non-expert iMD-VR users could accurately 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). [32] In VR, users share a 3D virtual space with proteins and drug molecules (and can also share the space with each other, enabling multiple users to interact with the same system).

Here, we apply iMD-VR to dock both a small ligand and an oligopeptide substrate to variants of Mpro (both structures are shown in Fig 1), testing different protocols. As in previous work on other enzymes, we compare results from iMD-VR with the crystal structure. We use iMD-VR to dock the inhibitor into an apo SARS-CoV-2 Mpro structure. We also dock an 11-mer oligopeptide into the original SARS-CoV Mpro with H41A mutant, giving structures in good agreement with the crystal structure. Finally, we created substrate complexes for SARS-CoV-2 Mpro for which, at the time of writing, no crystal structure exists (and of course capturing active substrate structures crystallographically is generally not possible). MD simulation was performed on each docked structure from iMD-VR to test the stability of the predicted binding poses. 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. It will also be useful for education on viral enzyme structure and function (as we have shown in other contexts) [33] and testing hypotheses e.g. on substrate binding and Mpro mechanism, complementing other types of simulations. [34]

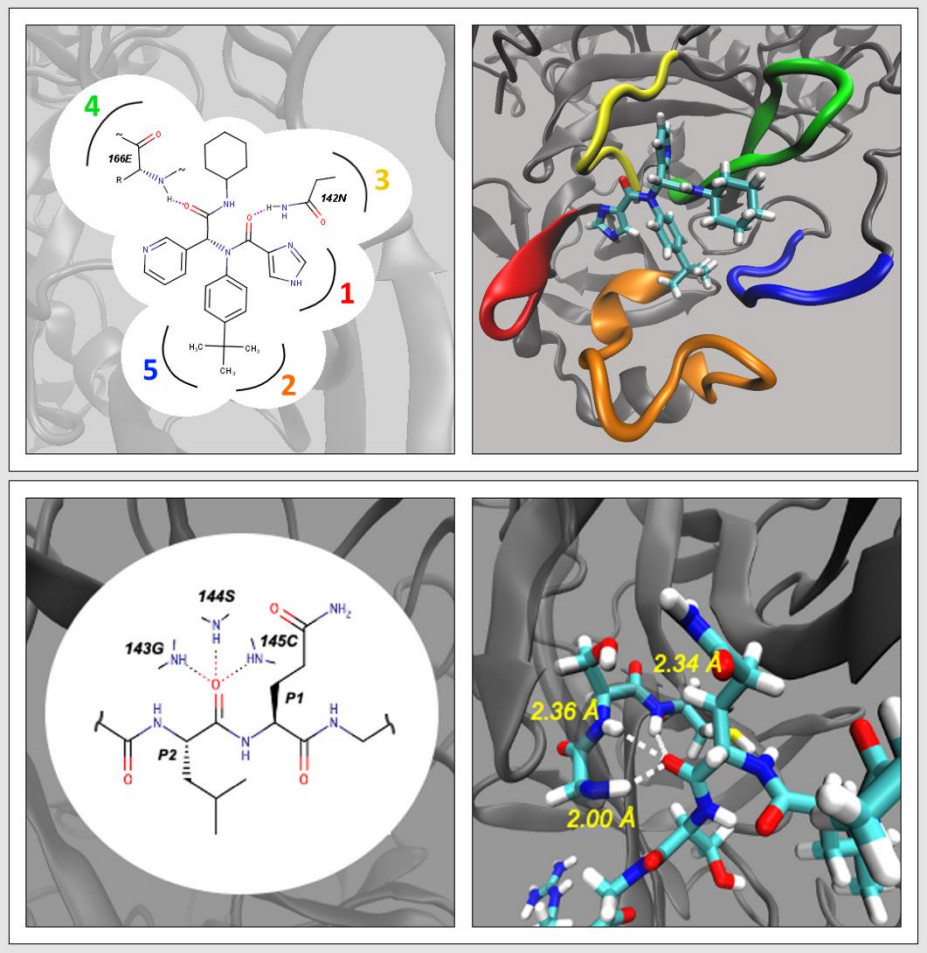

Figure 1 Top: Details of the binding of the X77 inhibitor revealed by crystallography (TOP LEFT) The X77 inhibitor forms two hydrogen bonds with the SARS-CoV-2 Mpro; five active site loops around the ligand are also indicated schematically. (TOP RIGHT) X77 bound covalently in the active site of SARS-CoV-2 Mpro (PDB ID: 6W63). The five active site loops are coloured as follows: Loop (1), residues 22 to 27, is leftmost in red, (2) residues 41 to 54 is at the bottom in orange, (3) residues 140 to 145 is at the top left in yellow, (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) Oxyanion hole interactions between the 11-mer oligopeptide substrate (observed both for the 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).

2. Methods

2.1 Simulation parameters and set up

The AMBER ff14sb forcefield [35] was used for protein and peptide structures. X77 was parameterised with GAFF. [36] All simulations used the OBC2 implicit water model [37] and used OpenMM as the force engine. Prior to iMD-VR, crystallographic water molecules were removed and the structures were minimised and equilibrated: details given in the ESI. To prevent unrealistic perturbation of the protein

structure during iMD-VR (e.g. protein unfolding), a harmonic positional restraint of 50 kJ/mol was applied to the protein backbone atoms in most of the protocols (except protocol 3 in which the protease is fully flexible), as in previous work. [27] Each iMD-VR docking experiment was repeated a total of five times, generating a total of five final structures per run.

To evaluate the stability of enzyme complexes generated by iMD-VR, all five final structures were subjected to energy minimization and re-equilibrated using the protocol described in the ESI. After energy minimization, an additional 10ns of MD simulation was run and the resulting trajectories were analysed to evaluate each structure. Details of these MD runs, and analysis, are given in the ESI.

To guide iMD-VR docking, a trace representation of the docked ligand or oligopeptide in the crystallographic position in the binding pocket was superimposed onto the simulation space and used as a 3D visual guide to where X77 or the oligopeptide should be placed; these are referred to here as 'trace atoms'. Further details can be found in the ESI.

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 [31] based on a known SARS-CoV-2 Mpro protein-ligand complex (PDB: 6W63). The ligand 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 ot a position where it was clearly not bound to the protein. Then, they moved the ligand back to its initial pose. Trace atoms were placed at the initial position of the ligand to guide the user, taking the form of a fixed and translucent representation of the molecule. Trace atoms are only visual clues and do not take part in the equation of motion. As noted in the Methods, we used an implicit solvent model (as in our previous work) to simplify and speed docking, and also applied positional restraints on the protein backbone to avoid large deformations (see ESI).

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 that compounds that we docked using iMD-VR in our previous work. Fig 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. [38] Previous work has shown differential docking behaviour between different protein tautomers [31], so we tested SARS-CoV-2 Mpro with in neutral and zwitterionic forms of the dyad. We interactively unbound and docked X77 from both using iMD-VR. After iMD-VR docking, a representative structure was selected based on having the lowest

ligand heavy atom RMSD compared to the ligand coordinates in the minimized and equilibriated structure (i.e. compared to the starting bound pose).

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 ID: 6M03). In this case, only the neutral tautomer was modelled, because the complexes were generally more stable than the zwitterionic form (Fig S1). The same trace atoms from the iMD-VR docking experiments on the complexed protein above 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 uses trace atoms to orientate the ligand, but instead primarily focuses on reforming the two hydrogen bonds shown in Fig 1. A representative structure was chosen based on the two interactions being present. Henceforth, these two protocols are referred to as "apo X77 docking protocol 1" and "apo X77 docking protocol 2", respectively.

2.3 iMD-VR docking of an oligopeptide to Mpro

2.3.1 Docking a substrate into H41A mutant SARS-CoV Mpro

The 11-mer oligopeptide (derived from one of the 11 cleavage sites of PP1ab [8], sequence TSAVLQSGFRK) was interactively removed from, and then docked into the active site of the SARS-CoV Mpro with an inactivating H41A mutant (PDB: 2Q6G). [17] The aim was to test iMD-VR docking of long, flexible molecules. We chose this substrate because it is processed by both the SARS-CoV and SARS-CoV-2 Mpro. [39] 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 modelled the neutral dyad for all substrate docking tasks with the substrate here.

Before beginning iMD-VR docking, 10ns of MD 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 set-up of this MD simulation followed the same protocol as all the other MD simulations and is described in the ESI. This simulation will henceforth be 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 key hydrogen bonds observed in the crystal structure (Fig. 5a). A representative docked structure was extracted in which the oxyanion hole interactions are present (Fig. 1). In the first protocol, we applied positional restraints to the backbone of the protease, like those used for X77 (see above), and in our previous work. [31] In the second protocol, the protease is fully flexible. These protocols will henceforth be called "protocol 2" and "protocol 3", respectively. Another protocol, "protocol 1", was also tested. This protocol is similar to protocol 2, but the focus lies on superimposing the substrate atoms directly on top of the trace atoms. For this protocol, we extracted the structure with the lowest substrate RMSD compared to the crystal structure. However, the resulting structures lacked important interactions observed in the crystal structure, and so protocol 1 was not used for further oligopeptide docking. These findings emphasise 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 forcefield). Results of the docking experiment with protocol 1 are detailed in the ESI.

2.3.2 Docking substrate into apo and complex structures of SARS-CoV-2 Mpro

There is as yet no crystal structure available of substrate bound to the SARS-CoV-2 Mpro. iMD-VR was employed here to dock a substrate 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), which will henceforth be described as the 'inhibitor' structure. These two structures were chosen to test the effect of starting structure on the docking solution.

Protocols 2 and 3 were tested for iMD-VR docking to both SARS-CoV-2 Mpro structures, using trace atoms taken from 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 its cognate protein complex (PDB: 6W63) and the apo (PDB: 6M03) form of the SARS-CoV-2 Mpro. For the ligand-complexed structure, we applied our tested iMD-VR ligand docking protocol [31] to recreate the ligand-protein complex: 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 as for previous

protein-ligand systems, i.e. less than 2.5 Å. [31] Five structures from iMD-VR docking attempts were then subject to 10 ns of unrestrained MD simulation, to test their stability. The complexes were stable: the heavy atom RMSD of X77 during remained within 2.5 Å of the equilibrated and minimized structure used to initiate iMD-VR (Fig 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 and and should be useful in predicting binding poses of novel ligands.

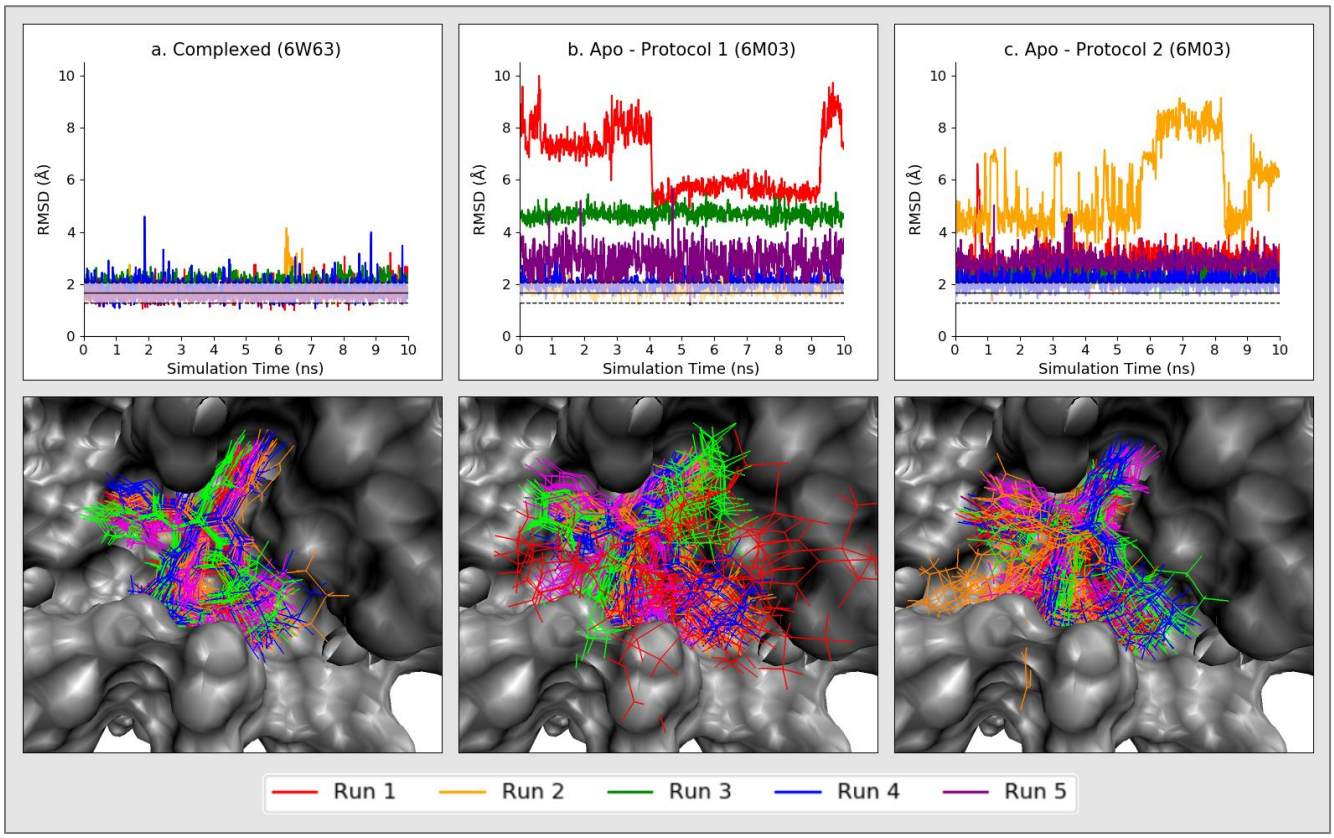

Figure 2 (TOP) Results from MD simiulations for structures from iMD-VR docking of the X77 ligand into (a) its cognate crystal structure (ligand complexed X77 docking) (b) the apo structure (using X77 docking protocol 1), and (c) the apo structure (using X77 docking protocol 2. For eachthe RMSD of X77 is shown for 10 ns of MD simulation starting from five independent iMD-VR docking experiments (5 runs, one for each iMD-VR docked structure). For reference, the black line represents the average X77 RMSD from 10 ns of MD for the original, unperturbed bound structure (6W63). The translucent bars represent the standard deviation around this mean. (BOTTOM) For each of the five runs for each of the three docking experiments, snapshots from the 10 ns of MD simulation (taken every nanosecond), superimposed on the X77 complex crystal structure. Each image corresponds to the graph directly above it.

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). We tested the use of an apo structure for predicting X77 binding using iMD-VR docking. IMD-VR has the advantage (compared to many docking approaches) that both the ligand and protein are flexible and so can adapt their structures to each other. Two iMD-VR protocols for docking X77 to the apo form of SARS-CoV-2 Mpro were tested: one with the aim of finding a low RMSD, the other instead focuses on forming two important hydrogen bonds (shown in Fig. 1). A challenge became immediately apparent in the apo crystal structure (Fig. 3). In the apo structure, the active site loop containing residues 41 to 54 (highlighted in orange in Fig. 1) is in a different position from the ligand complex structure, such that Tyr54 occupies part of the X77 binding pocket. Molecular dynamics performed on the apo structure showed this loop exhibits a high degree of flexibility, especially compared to X77-complexed Mpro (Fig. S4); due to this higher flexibility, we captured a state where the Tyr54 is blocking the pocket, hence X77 could not be fully replaced (Fig. 3). We tested whether apo-docked structures generated by iMD-VR, with the ligand correctly oriented, and forming important contacts with the protein, relaxed into a structure closer to the crystal structue of the complex with additional MD. During minimization and requilibration on the iMD-VR generated complexes, Tyr54 spontaneously moved back after X77 docking and, furthermore, the 41 to 54 loop exhibited a lessened degree of flexibility, comparable the complexed form of Mpro (Fig. S4). Encouragingly, this suggests that the apo structure can be used for docking small ligands, so long as the structure is given opportunity to relax.

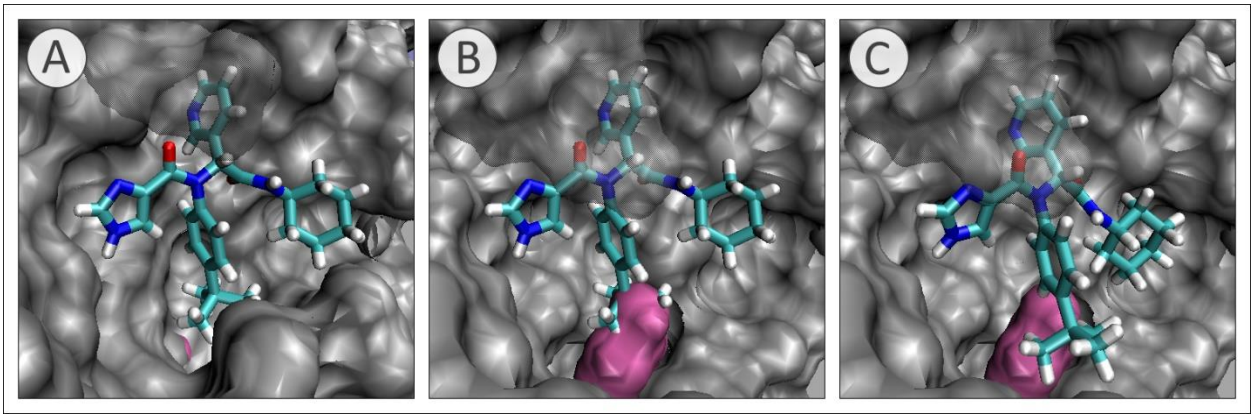

Figure 3 (A) Crystal structure of X77 bound to the SARS-CoV-2 Mpro (PDB ID: 6W63). Tyr54 (highlighted in pink) is buried in the protein structure. (B) The X77 structure from 6W63 superimposed on the apo structure starting coordinates, used in the two apo X77 docking protocols. Tyr54 (highlighted in pink) has shifted so that it is no longer buried and is occupying the X77 binding pocket, causing a steric clash with the tert-butylbenzene group of X77. (C) A representative iMD-VR generated X77 bound atom positions (taken from apo X77 docking protocol 2), showing how X77 could no longer be docked correctly due to the presence of Tyr54 (highlighted in pink).

RMSD analysis

Not surprisingly, docking to the apo structure (using either docking protocol 1 or 2) results in slightly higher RSMDs and less stable structures than docking to the cognate ligand complex structure, with less overall consistency between runs (Fig. 2). However, for each protocol, 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 (prior to iMD-VR) yielded an average RMSD of 1.7 Å, and the lowest average X77 RMSD for iMD-VR docking to the complexed structure was 1.6 Å. In comparison, the lowest average X77 RMSD for the apo X77 docking protocol 1 was 1.8 Å, and the lowest for apo X77 docking protocol 2 was 2.1 Å. Although the first protocol produced a pose with lower average RMSD during subsequent MD, overall, it performed worse: The total average RMSD for apo X77 docking protocol 1 was 3.6 Å (σ 1.9 Å), whilst the total average RMSD for apo X77 docking protocol 2 was 3.2 Å (σ 1.5 Å). This indicates that a protocol in which the user aims to form important bindinigbinding contacts gives better results than one in which selection is based on RMSD alone.

Hydrogen bond analysis

Fig. 4 shows analsysis of hydrogen bonds in the 10 ns validation MD simulations of structures from the two apo X77 iMD-VR docking protocols. For apo X77 docking protocol 1, runs 1 and 3 show poorest hydrogen bonding and highest X77 RMSD of the five runs (Fig. 4 b). In iMD-VR docking, hydrogen bond 2 did not form in run 2 of the apo X77 docking protocol 2, and rarely in subsequent MD simulation (Fig. 4

c), which resulted in the highest X77 RMSD (Fig. 2 c). RMSD is clearly an important metric, but consideration of RMSD may not give structures with important binding interactions (as noted above). Therefore, as with other manual docking methods, consideration of key hydrogen bonds is important in creating ligand complexes with iMD-VR.

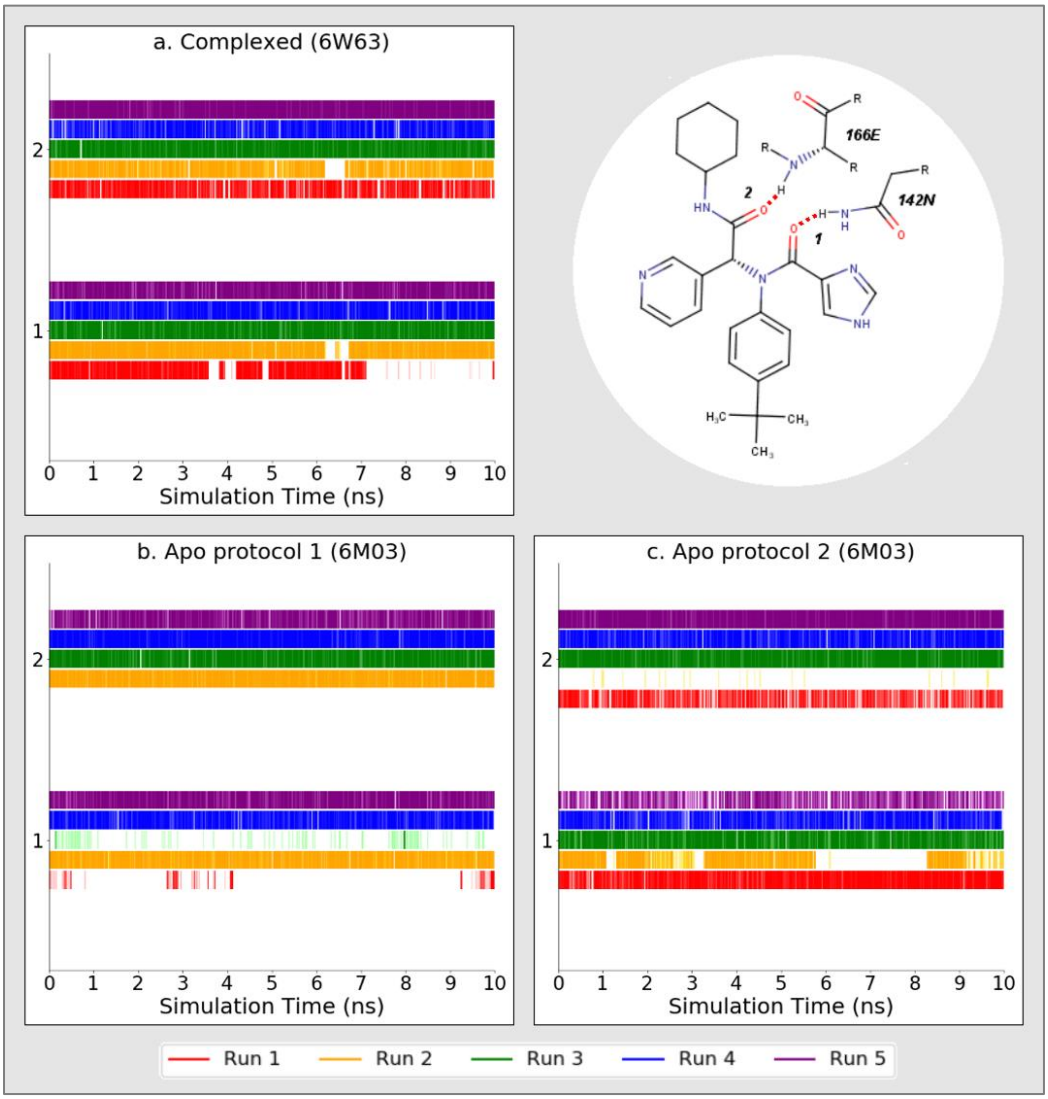

Figure 4 Results from (a) ligand complexed X77 iMD-VR docking, (b) apo X77 docking protocol 1, and (c) apo X77 docking protocol 2, showing presence of two hydrogen bonds over 10 ns of MD simulations of the resulting structures. Bold lines with a solid color indicate that hydrogen bond is present in that snapshot; faint lines indicate that a hydrogen bond Is not present, but the interatomic distance between the donor and acceptor atom being less than 4 Å.

Additional analysis

When a reference (e.g. crystal) structure is known, RMSD provides an objective value to rank iMD-VR generated poses but, as noted above, users should pay careful attention to interactions in generating structures by iMD-VR docking. [40] We tested other analyses of the iMD-VR generated poses. Fig. S2 shows the fraction of contacts recreated for each of the five runs for each of the three docking experiments. Similarly, Figs. S3-8 show the RMSF of the five active site protein loops during MD simulatiosn with no backbone restraints. These analyses are more extensively discussed in the ESI, however, while this lends further confidence to the iMD-VR generated structures, the methods suggested here should not be treated as an exhaustive list. As has been previously recommended, users should pay attention to both protein-ligand interactions and dynamics when evaluating iMD-VR docked structures; with this in mind, more data analysis will always be preferable to less.

3.1.3 Discussion

At least one run from each docking experiment was deemed sucessful based on all the evaluation metrics (including those discussed in the ESI). All the poses generated by iMD-VR docking, for each of the three protocols, are given in the ESI. Overall, iMD-VR docking successfully recreated X77 and SARS-CoV-2 Mpro complexes. Docking to an apo structure gave complexes with hgher 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 is useful for quickly evaluating the similarity between a known structure and a pose generated by iMD-VR docking, but has significant limitations. [40] RMSD of a ligand clearly only measures how well one ligand structure superimposes on another; it does not indicate whether binding interactions are present. Docking to apo structures is a challenge, but as we show here, iMD-VR docking of small molecule ligands to the apo structure of SARS-CoV-2 Mpro is successful, which is encouraging. This success reflects the flexibility of the protein and ligand in the iMD-VR approach. Users should focus on forming favourable interactions such as hydrogen bonds when docking in iMD-VR.

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. We tested this approach initially for an oligopeptide binding 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 10ns MD simulation of this crystal structure was performed to compare with structures generated by iMD-VR. The RMSD of the oligopeptide in this simulation was 3.7–5.6 Å. The hydrogen bonds formed are shown in Fig. 6 b.

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 (protocol 2); in the second, the protease was fully flexible (protocol 3). In both protocols, the aim was to reform 14 hydrogen bonds between the protease and substrate that are observed in the crystal structure (Fig. 5 a). 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 (the oxyanion hole, excluding hydrogen bond 9 as it is not strongly observed in the reference simulation) are formed was extracted, and 10 ns of MD simulation was run for each of these docked structures. These interactions (which are shown in Fig. 1) were chosen because they play an integral part in the mechanism of proteolytic cleavage, stabilising the tetrahedral intermediate by interaction with the carbonyl oxygen of the substrate's scissile bond. [41]

As a test, substrate docking was also performed following the previous protocol (protocol 1), as for small ligand docking (Section 3.1.1), where the focus was placing the oligopeptide atoms directly on top of the trace atoms. However, the docked structures generated from following 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), it was unsuitable for docking long, flexible molecules. Results from this docking protocol are detailed in the ESI. We therefore recommend that novice iMD-VR users should follow protocol 2 (I.e. apply backbone restraints) when docking large, flexible molecules to not accidentally perturb the protease structure, and expert iMD-VR users should follow protocol 3 where the protease is fully flexible.

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 (Fig. S8 b-c). This value is similar to the substrate RMSD values from the reference simulation (3.7–5.6 Å, Fig. S8 a), which indicates that the iMD-VR docked structures are similar to the crystal structure and are also stable in MD simulations. The 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

Xue *et al's* structural determination of the SARS-CoV Mpro H41A mutant complex with the 11-mer oligopeptide substrate revealed 13 hydrogen bonds important for substrate binding [17] (Fig. 6 a). The stability of these 13 hydrogen bonds, plus an additional hydrogen bond between Ser144 and P1-GLN that is part of the oxyanion hole, were analysed in each simulation of the SARS-CoV docked structures. The hydrogen bonds in all these simulations are very similar to those observed in the reference simulation (Fig. 6 b), in which the most stable interactions are hydrogen bonds 1, 2, 3, 6, 8, 10, 12, and 14 (Fig. S9 and S10). It is especially 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 (which is involved in proteolytic cleavage, and closest to the catalytic dyad) is tightly bound and fluctuates by less than 0.5 Å (Fig. S15 b-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 behaviour is observed in the reference simulation (RMSF of the P1-Gln residue below 0.5 Å and RMSFs for P6-Thr and P5'-Lys of 1 Å and 2.5 Å respectively, Fig. S15 a).

Altogether, this shows iMD-VR docking with protocols 2 or 3 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 iMD-VR protocols 2 and 3 for docking the substrate interactively 5 times each into the active site of two SARS-CoV-2 Mpro structures (an apo and an 'inhibitor' structure, PDB codes: 6M03 and 6W63, respectively), 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. Fig. 5 shows the SARS-CoV Mpro and both SARS-CoV-2 Mpro starting structures for iMD-VR simulations overlaid on the 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 VR simulation. 10 ns of implicit solvent MD was performed on each of the docked SARS-CoV-2 Mpro structures, giving a total of 20 MD simulations.

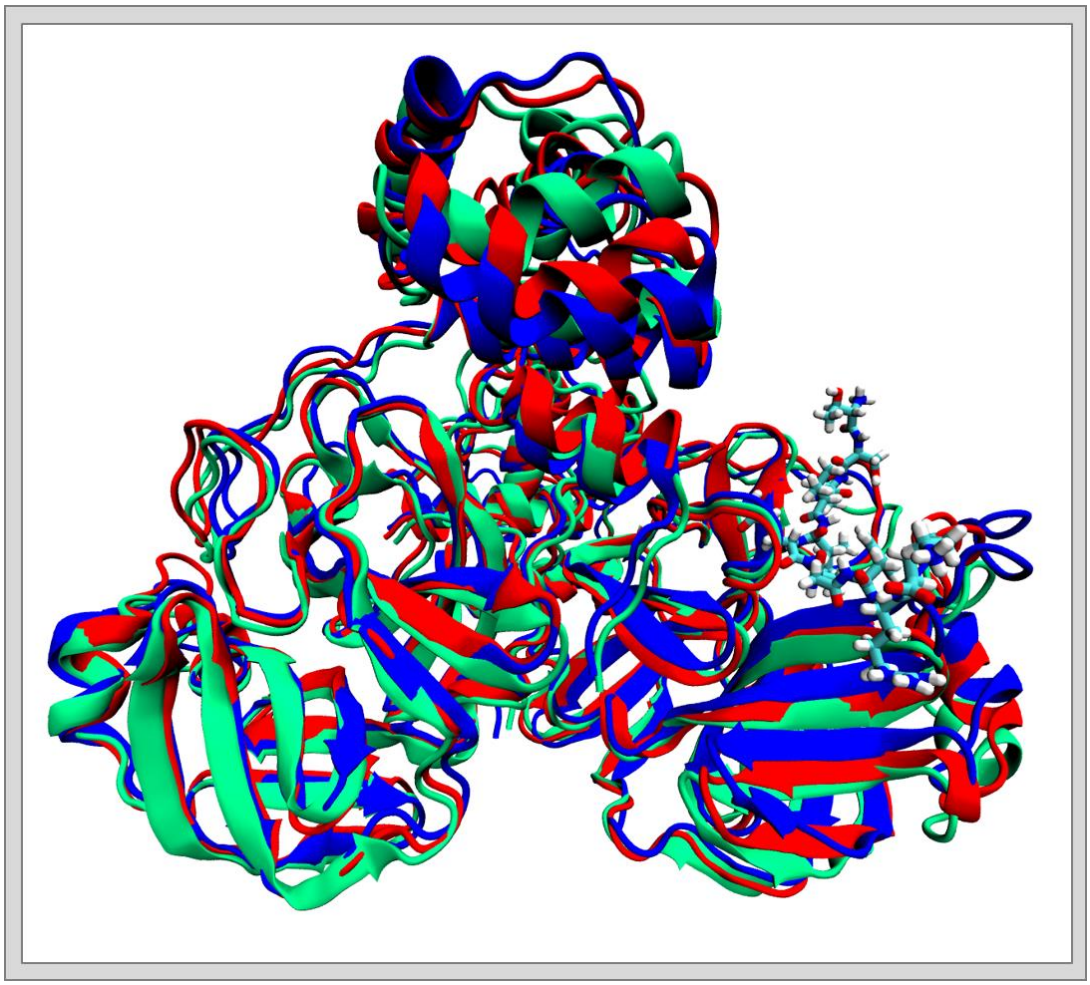

Figure 5 The three Mpro structures used as input for iMD-VR substrate docking overlaid. The structures are the SARS-CoV Mpro (red), the apo SARS-CoV-2 Mpro (blue) and the inhibitor-complexed SARS-CoV-2 Mpro (green. The substrate is also shown on the right, bound to the SARS-CoV structure. The two SARS-CoV-2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

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 (Fig. S8 d-g) 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 residues of the oligopeptide (see additional analysis). This is similar

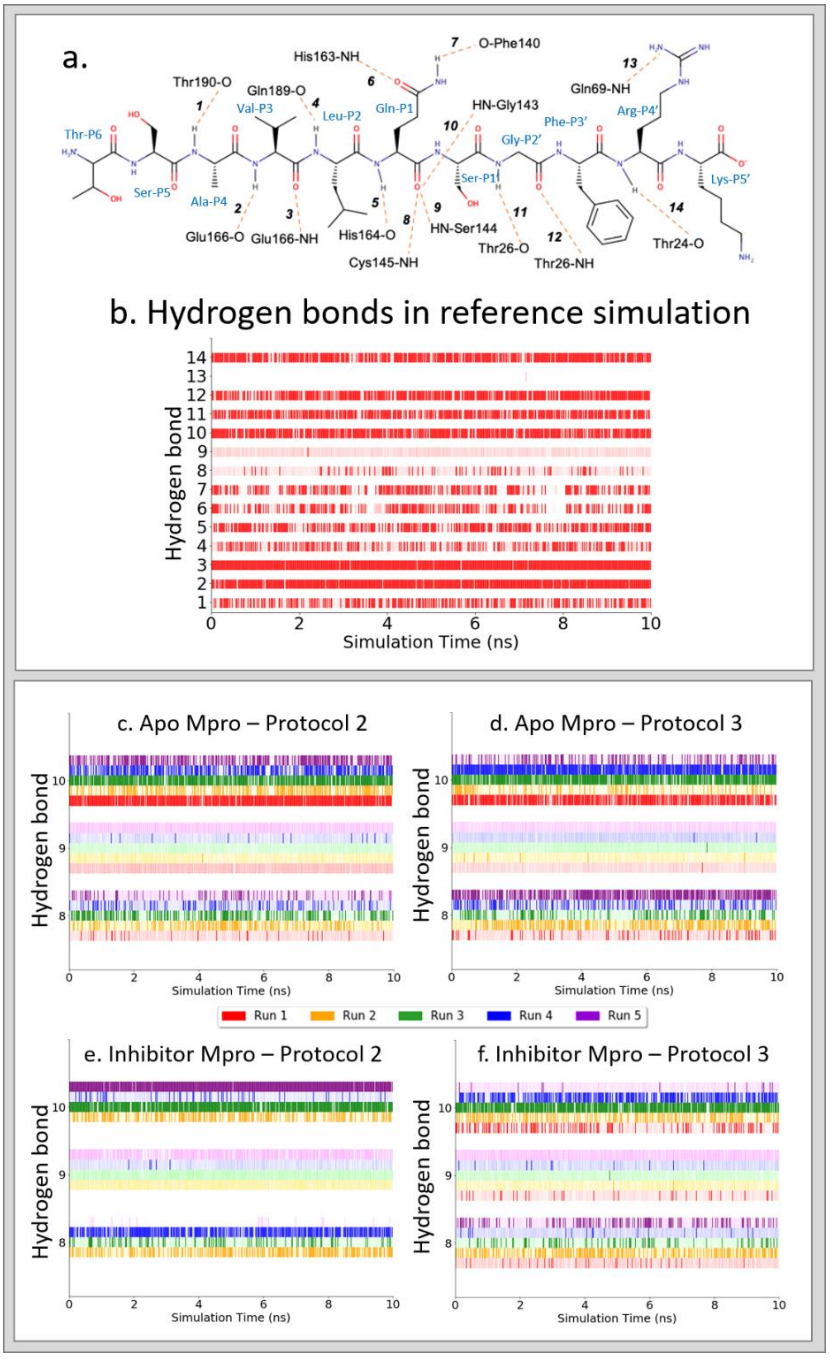

Figure 6 (a) The sequence of the 11-mer oligopeptide substrate and the 14 hydrogen bonds which were considered in analysis of VR generated structures. (b) The presence of the 14 hydrogen bonds throughout the MD simulation of the docked SARS-CoV Mpro crystal structure. The solid colour represents the hydrogen bond being present, and the lighter/translucent colour represents the donor and acceptor atoms being within 4 A of eachother. Hydrogen bond 13 is not included as it was never fully formed. The hydrogen bonds involved in the oxyanion hole (shown on the y-axis as 8, 9 and 10) are presented for throughout the simulations of the VR generated docked structures of the substrate to (c) the apo SARS-CoV-2 Mpro following VR docking protocol 2 and (d) following docking protocol 3, and of the substrate to (e) the inhibitor SARS-CoV-2 Mpro following docking protocol 2 and (f) docking protocol 3..

to the RMSD analysis of the reference simulation (Fig. S8 a) and also to simulations of the SARS-CoV docked structures (Fig. S8 b-c).

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. Docking to the apo SARS-CoV-2 Mpro generally formed the oxyanion hole interactions (hydrogen bond 8 and 10) using either protocol (Fig. 6 c-d). Similarly, hydrogen bonds 2, 3, 6, 12, and 13 were maintained throughout simulations of all the docked poses, showing that a common, stable pose is obtained by iMD-VR docking (Fig. S11 and S12). Hydrogen bond 9 is also present, although the interaction is weaker.

The simulations of the inhibitor SARS-CoV-2 Mpro docked structures, showed hydrogen bonds are not always present (similar to results from docking to the apo structure, Fig. S13 and S14), they still how most of these hydrogen bonds. Among the most stable in all simulations built from inhibtor complexes are hydrogen bonds 2, 3, 12, and 14. Furthermore, the oxyanion hole interactions are present throughout the simulations (Fig. 6e-f, except inhibitor - protocol 2 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 reference simulations of the substrate complexes.

Additional analysis

RMSF analysis of all 20 MD simulations of the SARS-CoV-2 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, Fig. S15 d-g). This is expected, as the ends of the substrate are bound more weakly than the central residues, as also observed for SARS-CoV Mpro (above) and the reference simulation. Altogether, this shows that substrate complexes build 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 stable structures of SARS-CoV-2 Mpro with this substrate. This docking was assisted by experience gained from docking the inhibitor, and chemical and spatial intuition. We tested different docking protocols, e.g. testing the effects of protein backbone

restraints. The docked SARS-CoV Mpro structures reformed more of the hydrogen bonds in 14 hydrogen bond network than in the SARS-CoV-2 structures. This is likely due to the fact 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, which may have to undergo conformational changes. Nonetheless, the substrate RMSD in simulations of the apo and inhibitor SARS-CoV-2 Mpro docked structure stayed in the same range as the 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. This is very encouraging: iMD-VR docking to all the different crystal structures gave stable structures that were also in agreement with experiment.

There is no significant discernible difference in the docked structures generated from protocols 2 or 3. We therefore recommend that expert iMD-VR users should follow protocol 3 (I.e. to treat the protease as fully flexible), and novice iMD-VR users should follow protocol 2 (applying backbone restraints, 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 the GPL open-source licence at https://irl.itch.io/narupaxr. A new, more flexible, version of Narupa is currently under development. This new version is available, under the same open-source licence 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 article are available in ESI for use with the previous version of Narupa. 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 is 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 with an inhibitor- and an oligopeptide substrate. In previous work [31] we have shown that even novice users can quickly generate structures in good agreement with crystal structures, 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 modelling substrate and inhibtors complexes of SARS-CoV-2 Mpro.Structures constructed by iMD-VR docking reproduce the key structural and binding motifs found in reference crystallographic structures. As expected, 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 in good agreement with experiment also, and stable in MD. This shows that iMD-VR docking can be used predictively to generate structures of complexes of Mpro even when specific experimental structural information is lacking. This is because the iMD-VR protocols described here allow for full flexibility of the enzyme, and of the ligand, and their conformational behaviour and interactions are modelled by a detailed atomistic forcefield, allowing the system to respond in a realistic manner.

In iMD-VR docking, users should pay attention to forming specific binding interactions. RMSD alone can be a poor indicator of the model quality. Other evaluation metrics should be used in addition. Furhter 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 can account for this [42], 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. Our iMD-VR framework will complement high-throughput docking e.g. to test structures predicted by computational screening, taking advantage of the protein and ligand flexibility in iMD-VR, and human intuition, to refine these structures. iMD-VR can also be used as a predictive tool, to model binding of potential inhibitors or substrates. It should be a useful tool for testing binding hypotheses. Any ligand can be modelled, 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 a little. Implicit solvent models have limitations for modelling protein structure and binding. Explicit solvation can straightforwardly be included in iMD-VR; this simply increases computational cost (which would reduce responsiveness for large systems). In cases where specific solvent molecules are invovled in binding [35, 36], use of an explicit solvent representation may be preferable.

IMD-VR allows the user to ‘step inside’ the SARS-CoV-2 Mpro, and manipulate its streucture and interactions – and dynamics- in atomic-level detail. This allows the user to perceive liigand 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 modelling tasks than e.g. traditional screen/mouse-based methods. [29] We believe that such technologies will find widespread application in structure-based drug design [31], in other molecular modelling, design and simulation applications, and in education. [33] IMD-VR should be directly useful in ongoing efforts to develop Mpro inhibitors as potential antivirals against COVID-19, especially as others look to crowd-source the problem. [43]

Narupa, the software framework that we have developed, is open source, and uses commodity VR hardware, which is widely accessible. [29, 31, 33, 44] It can therefore readily be used. We make our Mpro simulations here freely available; these can be used for ligand discovery and structure-activity studies for SARS-CoV-2. They run straightforwardly on a suitably configured laptop computer. In future work, we will aim to have these simulations 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 worked via iMD-VR. We encourage others to use these potentially transformative tools and believe that they will be widely useful.

5. Acknowledgements

HMD and RKW are funded by PhD studentships from the Engineering and Physical Sciences Research Council (EPSRC). JB acknowledges funding from the the EPSRC (programme grant EP/P021123/1). AJM thanks EPSRC (grant number EP/M022609/1, CCP-BioSim), BrisSynBio BrisSynBio, a BBSRC/EPSRC Synthetic Biology Research Centre (Grant Number: BB/L01386X/1) and MRC (MR/T016035/1) for support. AJM and DRG thank Oracle Research (University Partnership Cloud award) for support. DRG 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).

6. References

1. Fehr, A.R. and S. Perlman, *Coronaviruses: An Overview of Their Replication and Pathogenesis*, in *Coronaviruses: Methods and Protocols*, H.J. Maier, E. Bickerton, and P. Britton, Editors. 2015, Springer New York: New York, NY. p. 1-23.
2. Lin-Fa, W., et al., *Review of Bats and SARS.* Emerging Infectious Disease journal, 2006. **12**(12): p. 1834.
3. Chen, Y., Q. Liu, and D. Guo, *Emerging coronaviruses: genome structure, replication, and pathogenesis.* Journal of medical virology, 2020. **92**(4): p. 418-423.
4. Zhang, L., et al., *Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors.* Science, 2020: p. eabb3405.
5. Morse, J.S., et al., *Learning from the Past: Possible Urgent Prevention and Treatment Options for Severe Acute Respiratory Infections Caused by 2019-nCoV.* ChemBioChem, 2020. **21**(5): p. 730-738.
6. Kiemer, L., et al., *Coronavirus 3CLproproteinase cleavage sites: Possible relevance to SARS virus pathology.* BMC Bioinformatics, 2004. **5**(1): p. 72.
7. Chuck, C.-P., et al., *Profiling of substrate specificity of SARS-CoV 3CLpro.* PloS one, 2010. **5**(10): p. e13197.
8. Anand, K., et al., *Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs.* Science, 2003. **300**(5626): p. 1763-1767.
9. Grottesi, A., et al., *Computational Studies of SARS-CoV-2 3CLpro: Insights from MD Simulations.* International Journal of Molecular Sciences, 2020. **21**(15): p. 5346.

10. Chen, S., et al., *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): p. 59-74.
11. Marra, M.A., et al., *The Genome Sequence of the SARS-Associated Coronavirus.* Science, 2003. **300**(5624): p. 1399.
12. Krichel, B., et al., *Processing of the SARS-CoV pp1a/ab nsp7-10 region.* The Biochemical journal, 2020. **477**(5): p. 1009-1019.
13. Hegyi, A. and J. Ziebuhr, *Conservation of substrate specificities among coronavirus main proteases.* Journal of general virology, 2002. **83**(3): p. 595-599.
14. Fan, K., et al., *Biosynthesis, purification, and substrate specificity of severe acute respiratory syndrome coronavirus 3C-like proteinase.* Journal of Biological Chemistry, 2004. **279**(3): p. 1637-1642.
15. Po-Huang, L., *Characterization and Inhibition of SARS-Coronavirus Main Protease.* Current Topics in Medicinal Chemistry, 2006. **6**(4): p. 361-376.
16. Graziano, V., et al., *SARS CoV Main Proteinase: The Monomer–Dimer Equilibrium Dissociation Constant.* Biochemistry, 2006. **45**(49): p. 14632-14641.
17. Xue, X., et al., *Structures of Two Coronavirus Main Proteases: Implications for Substrate Binding and Antiviral Drug Design.* Journal of Virology, 2008. **82**(5): p. 2515.
18. Barnard, D.L. and Y. Kumaki, *Recent developments in anti-severe acute respiratory syndrome coronavirus chemotherapy.* Future Virology, 2011. **6**(5): p. 615-631.
19. Haagmans, B.L. and A.D.M.E. Osterhaus, *Coronaviruses and their therapy.* Antiviral Research, 2006. **71**(2): p. 397-403.
20. Chen, Y.a.Y., CPB and Wong, KY, *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., et al., *Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease.* Science, 2020. **368**(6497): p. 1331-1335.
22. Ma, C., et al., *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): p. 678-692.
23. Jin, Z., et al., *Structure of Mpro from COVID-19 virus and discovery of its inhibitors.* Nature, 2020.
24. Irwin, J.J. and B.K. Shoichet, *ZINC--a free database of commercially available compounds for virtual screening.* Journal of chemical information and modeling, 2005. **45**(1): p. 177-182.
25. Liu, X. and X.-J. Wang, *Potential inhibitors for 2019-nCoV coronavirus M protease from clinically approved medicines.* bioRxiv, 2020: p. 2020.01.29.924100.
26. Ton, A.-T., et al., *Rapid Identification of Potential Inhibitors of SARS-CoV-2 Main Protease by Deep Docking of 1.3 Billion Compounds.* Molecular Informatics, 2020. **n/a**(n/a).
27. Ge, Y., et al., *Identification of the quinolinedione inhibitor binding site in Cdc25 phosphatase B through docking and molecular dynamics simulations.* Journal of Computer-Aided Molecular Design, 2017. **31**(11): p. 995-1007.
28. Amaro, R.E., et al., *Ensemble Docking in Drug Discovery.* Biophysical Journal, 2018. **114**(10): p. 2271-2278.
29. O'Connor, M., et al., *Sampling molecular conformations and dynamics in a multiuser virtual reality framework.* Science Advances, 2018. **4**(6): p. eaat2731.
30. Cassidy, K.C., et al., *ProteinVR: Web-based molecular visualization in virtual reality.* PLOS Computational Biology, 2020. **16**(3): p. e1007747.
31. Deeks, H.M., et al., *Interactive molecular dynamics in virtual reality for accurate flexible protein-ligand docking.* PLOS ONE, 2020. **15**(3): p. e0228461.

32. Pagadala, N.S., K. Syed, and J. Tuszynski, *Software for molecular docking: a review.* Biophysical Reviews, 2017. **9**(2): p. 91-102.
33. Bennie, S.J., et al., *Teaching Enzyme Catalysis Using Interactive Molecular Dynamics in Virtual Reality.* Journal of Chemical Education, 2019. **96**(11): p. 2488-2496.
34. Świderek, K. and V. Moliner, *Revealing the molecular mechanisms of proteolysis of SARS-CoV-2 Mpro by QM/MM computational methods.* Chemical Science, 2020.
35. Maier, J.A., et al., *ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB.* J Chem Theory Comput, 2015. **11**(8): p. 3696-713.
36. Wang, J., et al., *Development and testing of a general amber force field.* Journal of computational chemistry, 2004. **25**(9): p. 1157-1174.
37. Onufriev, A., D. Bashford, and D.A. Case, *Exploring protein native states and large-scale conformational changes with a modified generalized born model.* Proteins: Structure, Function, and Bioinformatics, 2004. **55**(2): p. 383-394.
38. Paasche, A., et al., *Evidence for Substrate Binding-Induced Zwitterion Formation in the Catalytic Cys-His Dyad of the SARS-CoV Main Protease.* Biochemistry, 2014. **53**(37): p. 5930-5946.
39. Wu, F., et al., *A new coronavirus associated with human respiratory disease in China.* Nature, 2020. **579**(7798): p. 265-269.
40. Kufareva, I. and R. Abagyan, *Methods of Protein Structure Comparison*, in *Homology Modeling: Methods and Protocols*, A.J.W. Orry and R. Abagyan, Editors. 2012, Humana Press: Totowa, NJ. p. 231-257.
41. Simón, L. and J.M. Goodman, *Enzyme Catalysis by Hydrogen Bonds: The Balance between Transition State Binding and Substrate Binding in Oxyanion Holes.* The Journal of Organic Chemistry, 2010. **75**(6): p. 1831-1840.
42. Hart, K.M., et al., *Modelling proteins' hidden conformations to predict antibiotic resistance.* Nature Communications, 2016. **7**(1): p. 12965.
43. Chodera, J., et al., *Crowdsourcing drug discovery for pandemics.* Nature Chemistry, 2020. **12**(7): p. 581-581.
44. Amabilino, S., et al., *Training Neural Nets To Learn Reactive Potential Energy Surfaces Using Interactive Quantum Chemistry in Virtual Reality.* The Journal of Physical Chemistry A, 2019. **123**(20): p. 4486-4499.