



SARS-CoV-2 M^{Pro} conformational changes induced by covalently bound ligands

Glaucio Monteiro Ferreira^{a,b,*}, Thales Kronenberger^{b,c*}, Arun Kumar Tonduru^c, Rosario Dominguez Crespo Hirata^a, Mario Hiroyuki Hirata^a and Antti Poso^{b,c}

^aDepartment of Clinical and Toxicological Analyses, School of Pharmaceutical Sciences, University of São Paulo, São Paulo, Brazil; ^bDepartment of Oncology and Pneumology, Internal Medicine VIII, University Hospital Tübingen, Tübingen, Germany; ^cSchool of Pharmacy, Faculty of Health Sciences, University of Eastern Finland, Kuopio, Finland

Communicated by Ramaswamy H. Sarma

ABSTRACT

SARS-CoV-2's main protease (M^{Pro}) interaction with ligands has been explored with a myriad of crystal structures, most of the monomers. Nonetheless, M^{Pro} is known to be active as a dimer but the relevance of the dimerization in the ligand-induced conformational changes has not been fully elucidated. We systematically simulated different M^{Pro}-ligand complexes aiming to study their conformational changes and interactions, through molecular dynamics (MD). We focused on covalently bound ligands (N1 and N3, ~9 μs per system both monomers and dimers) and compared these trajectories against the apostructure. Our results suggest that the monomeric simulations led to an unrealistically flexible active site. In contrast, the M^{Pro} dimer displayed a stable oxyanion-loop conformation along the trajectory. Also, ligand interactions with residues His41, Gly143, His163, Glu166 and Gln189 are postulated to impact the ligands' inhibitory activity significantly. In dimeric simulations, especially Gly143 and His163 have increased interaction frequencies. In conclusion, long-timescale MD is a more suitable tool for exploring *in silico* the activity of bioactive compounds that potentially inhibit the dimeric form of SARS-CoV-2 M^{Pro}.

ARTICLE HISTORY

Received 20 May 2021
Accepted 16 August 2021

KEYWORDS

COVID-19; main protease; M^{Pro}; Molecular dynamics simulation

Introduction

A virus-caused illness later called Coronavirus disease 2019 (COVID-19) by World Health Organization (WHO) - has been a worldwide concern since its first report in December 2019 (Wuhan, China), named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (Wu et al., 2020). The disease caused by this new coronavirus was classified by the World Health Organization (WHO), in February 2020, as Coronavirus Disease 2019 (COVID-19). The outbreak was declared a pandemic in March 2020. By October 2020, ~39 million cumulative cases were recorded globally, with over a million deaths (*Coronavirus Disease (COVID-19) Situation Reports*, 2021).

Currently, patients with COVID-19 are treated with repurposed drugs, which effects are often controversial due to the adverse events or the lack of fully proven clinical verification of their therapeutic effects. Therefore, there is still a need for novel treatments and the investigation of potential drug targets remains the cornerstone when designing novel, safe and effective antiviral drugs (Penman et al., 2020).

The main protease of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2 M^{Pro}, herein referred as M^{Pro} for short) is a cysteine protease that plays a crucial role in the virus' life cycle since it releases replicases pp1a and pp1ab; these functional peptides are essential for replication and

transcription of the virus (Pillaiyar et al., 2016). M^{Pro} is conserved in all coronaviruses and lacks a homolog with human proteins, increasing its attractiveness as a druggable target. Recent studies reported noncovalent M^{Pro} inhibitors with high antiviral activity (IC₅₀=1 μM) and no cytotoxicity (Mendoza et al., 2020; Zhang et al., 2021). However, to the best of our knowledge, there are no SARS-CoV-2 M^{Pro} inhibitors available for clinical use.

M^{Pro} consists of a polypeptide chain with 306 amino acids structured in three domains (S-I, S-II and S-III) connected by a flexible loop (Figure 1A). The S-I and S-II domains have a complementary antiparallel β-barrel fold relevant to the protease mechanism. The S-III domain contains five α-helices arranged in a broadly antiparallel globular cluster linked to the S-II domain through the flexible loop (Jin et al., 2020). SARS-CoV-2 M^{Pro} has a conserved catalytic dyad composed of Cys145 and His41 (Figure S1). Further, the substrate-binding site contained on the surface between the S-I and S-II domains (in green and orange, respectively, in Figure 1A) includes oxyanion hole residues, relevant for the substrate binding (Suárez & Díaz, 2020), and is covered by a loop (in yellow, Figure 1A). M^{Pro} has been co-crystallized with two lead compounds (N1 and N3 (Dai et al., 2020; Zhang et al., 2020)). The main sub-pockets (color-coded in Figure 1C and D) are P1 (containing Phe140 and Glu166; in green), P1'

CONTACT Antti Poso  antti.poso@uef.fi  School of Pharmacy, University of Eastern Finland, P.O.BOX 1627, 70211 Kuopio, Finland; Mario Hiroyuki Hirata  mhirata@usp.br  Av Prof. Lineu Prestes, 580 - Bloco15 - São Paulo, Brazil, CEP 05508-900

*The authors equally contributed to this work.

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/07391102.2021.1970626>.

© 2021 Informa UK Limited, trading as Taylor & Francis Group

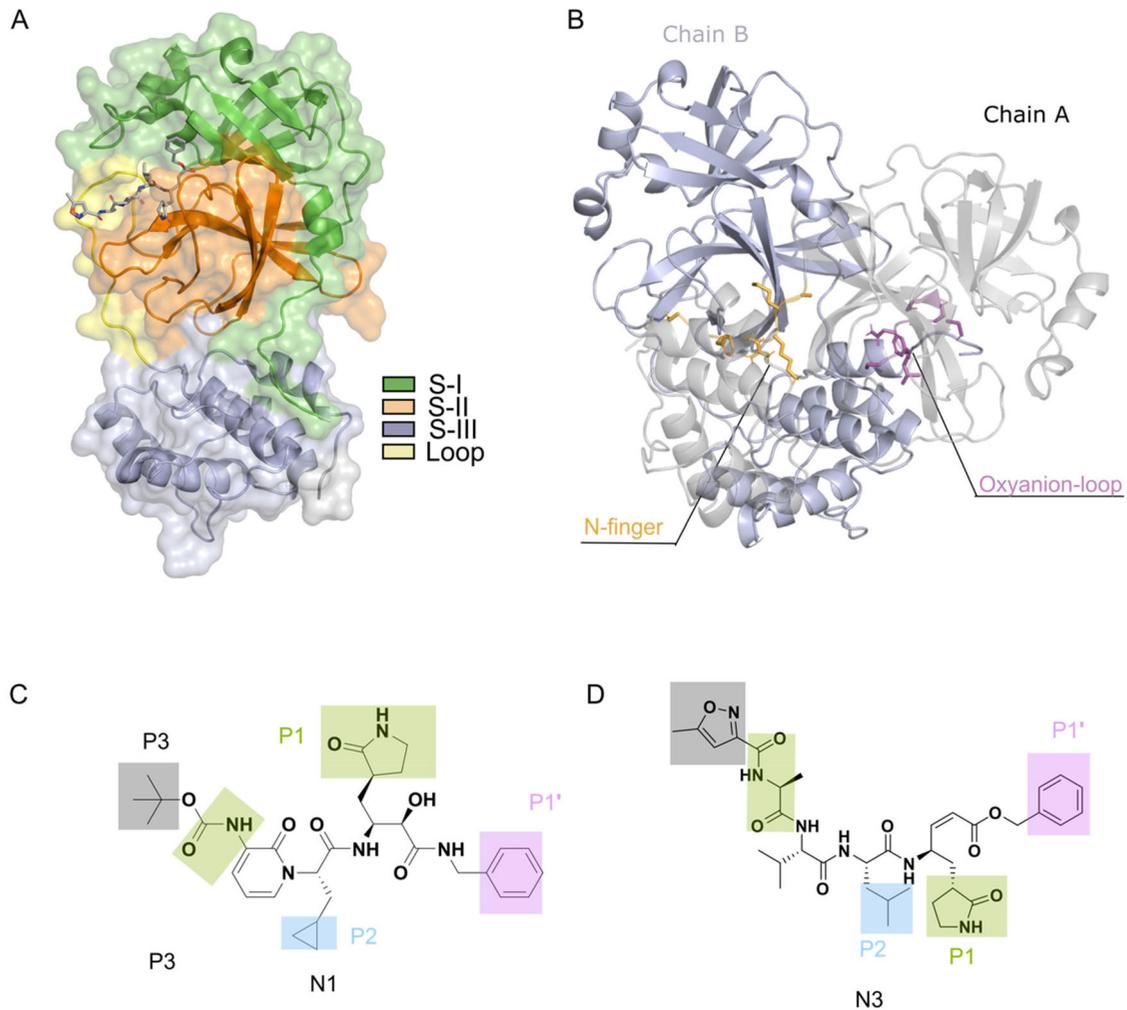


Figure 1. SARS-CoV-2 M^{pro} structure overview, ligands from PDB and evolutionary conservation. (A) Tertiary structure of the SARS-CoV-2 M^{pro} with domains S-I (in green, residues 8-101), S-II (in orange, residues 102-184), S-III domain (in purple, residues 201-303), long loop (in yellow, residues 185-200) and covalent ligands (sticks), (B) Dimer structure of the SARS-CoV-2 M^{pro} with highlights on the N-finger (in orange, residues 1-7) and oxyanion-loop (in magenta, residues 138-146), (C) Two-dimensional representation of the covalent ligands N1 (from PDB: 6y2g) and (D) N3 (from PDB: 6lu7).

(His41, Gly143, Ser144, Cys145 and His163; in pink), **P2** (His164 and Met49; in cyan) and **P3** (Pro168 and Gln189; in black).

The oxyanion-loop (residues 138 – 146) and catalytic residues are localized between subpockets **P1** and **P1'**, both of which are critical for substrate binding (Suárez & Díaz, 2020). The oxyanion-loop stabilizes the partial negative charge in the **P1** carbonyl group of the peptide substrate during the hydrolysis of the **P1–P1'** bond. The catalytically active conformation is stabilized by the interaction between the substrate and main-chain atoms of Gly143 and Cys145. In active conformations of M^{pro} the Cys145 is in direct interaction with the His41.

SARS-CoV-2 M^{pro} is active as a dimer. The main dimer interface includes S-III domains (in purple, Figure 1A), with the participation of the N-finger domain (Figure 1B). The N-finger is composed of the first seven residues of N-terminal (from the S-I domain, highlighted in pink in Figure 1B) and deletion of these residues in the 3CLpro homologue reduces the dimerization and, consequently, abolishes enzymatic

activity (<1%) (Chen et al., 2005). Further, also in 3CLpro, it has been shown that Arg4Ala mutation reduces enzymatic activity (Chen et al., 2005).

Despite the relevance of the dimerization to the active site's conformation (Suárez & Díaz, 2020), most of the current SARS-CoV-2 M^{pro} crystal structures in the Protein Data Bank (PDB) are presented as monomers. Additionally, the few simulation-based studies available for M^{pro} rely on monomeric structures with insufficient sampling due to their short timescale (50 ns – 2 μ s) (Komatsu et al., 2020; Peterson, 2020). It is well known that long-timescale simulations are needed to ensure that the observed conformational changes are statistically relevant (Henzler-Wildman & Kern, 2007). For this reason, we have simulated covalently bound ligands (using N1 and N3 as model ligands, $\sim 9 \mu$ s per system, monomers and dimers) and compared these trajectories against the apostructure. We have analyzed the major protein movements and observed that the dimeric state is more stable than the monomeric state, especially if the interaction between N-finger with oxyanion-loop is concerned. Our

investigation aims to clarify the relevance of the dimerization for the active conformation and ligand binding for M^{Pro} studies.

Results and discussion

SARS-CoV-2 M^{Pro} is more stable as a dimer than as a monomer

Principal Component Analyses (PCA) of all the simulations pooled together indicated that most of the M^{Pro} 's large moments were captured in the two components, with the first component accounting for 82.8%, while the second component was responsible for 4.3% of the total motion. The first PC separates the ligand-bound monomers into two conformations (Figure 2A and B) and accounting for 24% and 31% of the analyzed trajectory for N3 and N1, respectively. However, similar behavior is not seen with apostructure simulations. In a more detailed analysis monomer simulations (ligand bound), PC's motion is characterized by the coordinated movement between the S-II and S-III domains (Figure 2C), where S-III turns from its original conformation, potentially interfering with the dimerization interface (Suárez & Díaz, 2020). Dimeric M^{Pro} simulations do not show this movement, as the dominant feature is a small variation in loop conformation (between S-II and S-III). This indicates that the dimeric protein is more stable than monomer, at least when it comes to the S-III movement. It is interesting to note that the SARS-CoV-2 M^{Pro} is biologically active as a homodimer (Jin et al., 2020; Silvestrini et al., 2021; Suárez & Díaz, 2020).

In addition, this is supported by our finding that monomeric simulations have higher protein backbone's flexibility (as represented by RMSF values, Figure 3A) when compared to their dimeric counterparts (Figure 3B). This is especially manifested within the S-III domain region.

SARS-CoV-2 M^{Pro} dimer is stabilized by interactions in the S-III interface and N-finger

We also investigated whether the higher flexibility of S-III residues would interfere with the dimerization interface. M^{Pro} dimer X-ray structure is held together by interactions between several residues: (Ser1A-Phe140B, Ser1A-Glu166B, Ser1A-His172B, Arg4A-Glu290B, Arg4A-Lys137B, Ala7A-Val125B, Ser10A-Ser10B, Gly11A-Glu14B and Ser139A-Gln299B) (Suárez & Díaz, 2020). However, during MD simulations most frequent interactions between subunits are detected as following: Arg4A-Glu290B (side-chain), Ala7A-Val125B, Ser10A-Ser10B and Gly11A-Glu14B (Figure 4A and B, and Figure S4B-E). Especially S-II (Arg4A-Val125B) and S-III (Arg4A-Glu290B) interactions seems to contribute for the dimer stability. These findings are supported by Wang et al. (2020) that describes the dimerization as being driven by the interaction between Arg4A-Glu290B (Wang et al., 2020). The sequence conservation and the stability of interaction between Arg4-Glu290, we hypothesize the relevance of Arg4 for the dimerization/activity in SARS-Cov2 M^{Pro} , which could be validated by experimental mutations.

Interestingly, changes in frequency between dimeric ligand bound and apo structures are observed in Arg4-Lys137 and in the Ser1-Phe140 (113°; considered a weak interaction)/Glu166 (92°; a strong hydrogen bond)/His172 (102°; medium) interactions, both of which belong to the N-finger region (Figure S5). The strong hydrogen bond between Ser1A-Glu166A is also observed in the dimeric crystal structure (pdb: 7cb7).

The work of Chou et al. (2004) observed that interference with the Arg4-Glu290 interactions reduced the SARS-COV-1 dimerization and that, specifically Glu290Ala was enzymatically inactive whereas Arg4Ala was not (Chou et al., 2004). This supports the role of dimerization in activity for this enzyme's family.

Further, interactions between Ser1 and the residues Phe140, His172, and Glu166 were less frequent in the overall analyzed trajectory (Figure 4B-E and Figure S4A and B). We postulate that Ser1 helps to shape the substrate pocket in the normal catalysis, but does not contribute to the inhibited state, as Glu166 and Phe140 are involved in the inhibitor stabilization (see below).

The complete deletion of the N-finger in SARS-CoV-1 M^{Pro} , reduces the extent of the dimerization and completely abolishes the enzymatic activity (<1%) (Chen et al., 2005). This was corroborated to also translate in SARS-Cov-2's M^{Pro} by simulations (Suárez & Díaz, 2020), suggesting that the N-finger conformation upon dimerization exerts a direct influence on the oxyanion-loop (namely Ser139) motions.

Our results also indicate a water-mediated interaction between Ser139A-Gln299B with high frequency (>75% of the analyzed trajectory) in all studied systems (Figure 4F and Figure S4F). We can suggest that a water molecule could be structurally integrate this region, contributing to stabilize the intermolecular interactions, however this remains to be confirmed (Raschke, 2006).

Interestingly, Suarez et al. (Suárez & Díaz, 2020) suggested that a direct hydrogen bond between Ser139A-Gln299B (12% of their simulated time), which is something we exclusively observed in our apostructure simulations (30% of the analyzed trajectory). It is noteworthy that Ser139A-Gln299B can also participate in the dimer's oxyanion stabilization (Figure S6A and B), which links the dimerization stability with the conformation of the active site.

Additionally, free-energy calculations with N3 ligand in both monomeric and dimeric states suggested a lower energy interaction in the latter for the SARS-CoV-2 M^{Pro} , but not for the SARS-CoV M^{Pro} (Bello, 2020). The energy decomposition into the most relevant residues suggest that His41, Met49, Ser144, and Cys145 contributed significantly to the binding affinity (Bello, 2020). These results agree with the catalytic mechanism that shows the involvement of the main chain amides of Gly143, Ser144, and Cys145 in substrate cleavage in SARS-COV-1 (Chen et al., 2006). Accordingly, the functionality of the dimer is probably due to the interaction of the N-finger of each of the two monomers with Glu166 of the other monomer, which establishes the S-I domain, the pocket occupied by the substrate (Hsu et al., 2005).

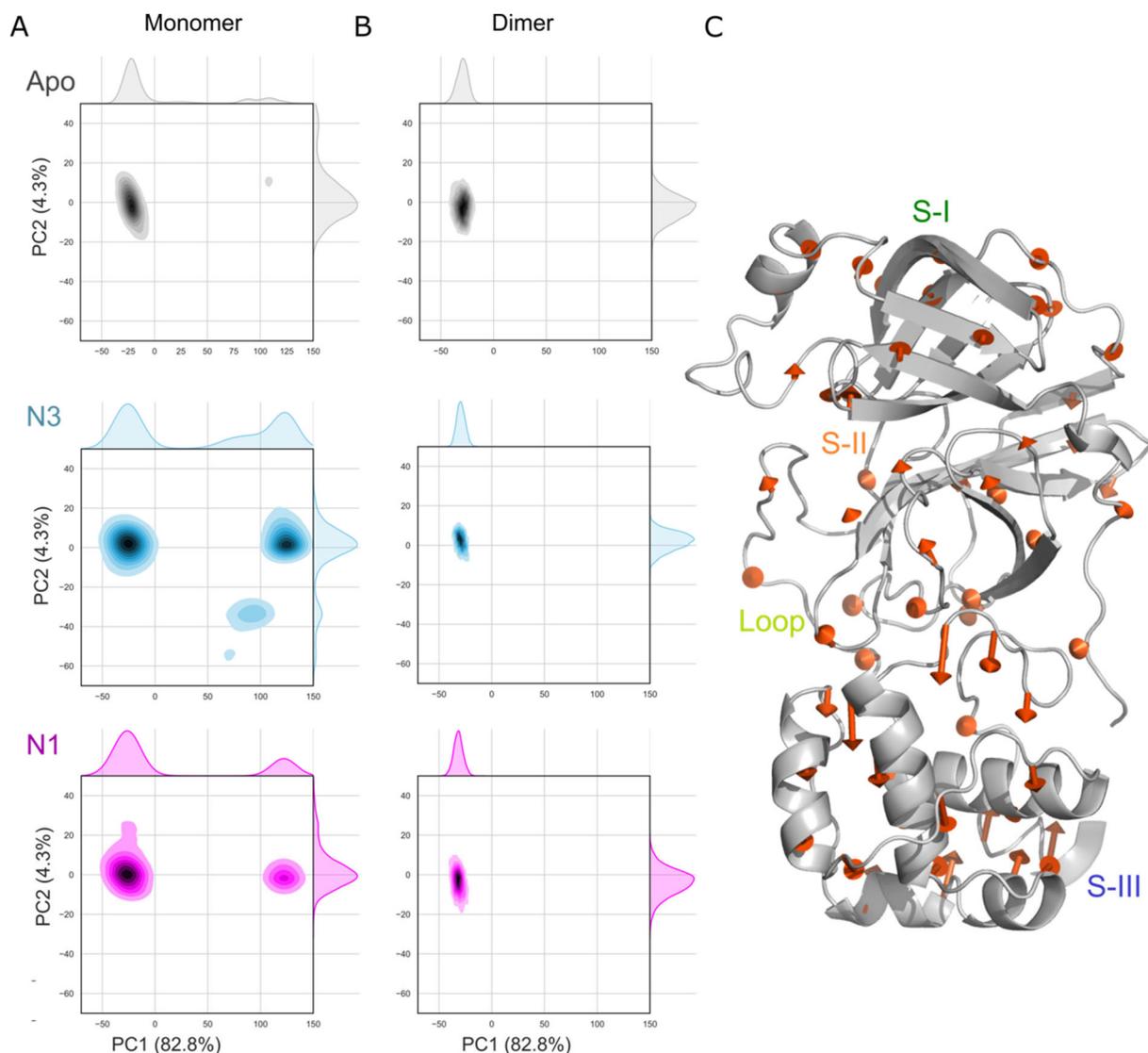


Figure 2. SARS-CoV-2 M^{Pro} PCA data and conformational changes induced by covalent ligands. SARS-CoV-2 M^{Pro} monomer (A) and dimer (B). Data referring to the apostructure (PDB: 6y84) is highlighted in gray; N3 bound structures (N3-6lu7) in blue; and N1 bound structures (N1-6y2g) in magenta. (C) Extreme movements from the PC1 plotted over the monomeric tertiary structure represented by orange arrows. PCA: principal component analysis.

Furthermore, in terms of the catalytic site, we propose that the interaction between Asn28-Cys145 and Gly143-Cys145 backbone atoms would stabilize the reactive conformation of Cys145 (Figure S6G and H). The interaction Asn28-Cys145 had a high frequency (>80%), in all systems, whereas Cys145-Gly143 was 40% more frequent in the dimeric simulations (Figure S6G and H). We also detected a smaller radius of the gyration variation for the Cys145 in dimeric simulations (22-23 Å) than for the monomeric (10-25 Å) forms (Figure S6C and D), and a smaller overall fluctuation (Figure S6E and F). This indicates that dimerization plays a role in stabilizing the active oxyanion-loop conformation.

Hydrogen bond interactions in the P1' region is influenced by dimerization

It is known that the oxyanion-loop is stabilized through a partial negative charge in the P1 carbonyl group of the peptide substrate during the hydrolysis of the P1 – P1' bond (Rut et al., 2021). Given our observation that the dimerization

stabilized the oxyanion-loop region, we further investigated its influence on the inhibitor binding to explain the differences in the ligands' inhibitory effects (Figure 5). We observed that the ligand N3 has different hydrogen bond frequencies with amine group of the Gly143 for the monomer (<30% of the analyzed simulation time) and the dimer (~70%, Figure 5A and D). Further, the mean distances between amine group of the Gly143 and both ligands are smaller for dimeric simulations (N1 = $M < 3.7$ Å and $D < 1.8$ Å; N3 = $M < 3.5$ Å and $D < 2.5$ Å) (Figure 5F).

Another apparently frequent interaction between M^{Pro} and inhibitors was detected for Glu166. The hydrogen bond between its side-chain and the pyrrolidine-2-one moieties from the ligand were stable through all the simulation (~98%, all systems, Figure 5A and D). Meanwhile, Glu166 backbone NH amide had a water-bridge with both N3 (~60% to monomer and ~30% to dimer, Figure 5C and D) and N1 (~50% to monomer and ~15% to dimer, Figure 5C and E). This water-mediated interaction seems to be more relevant for monomeric simulations, whereas Glu166 in

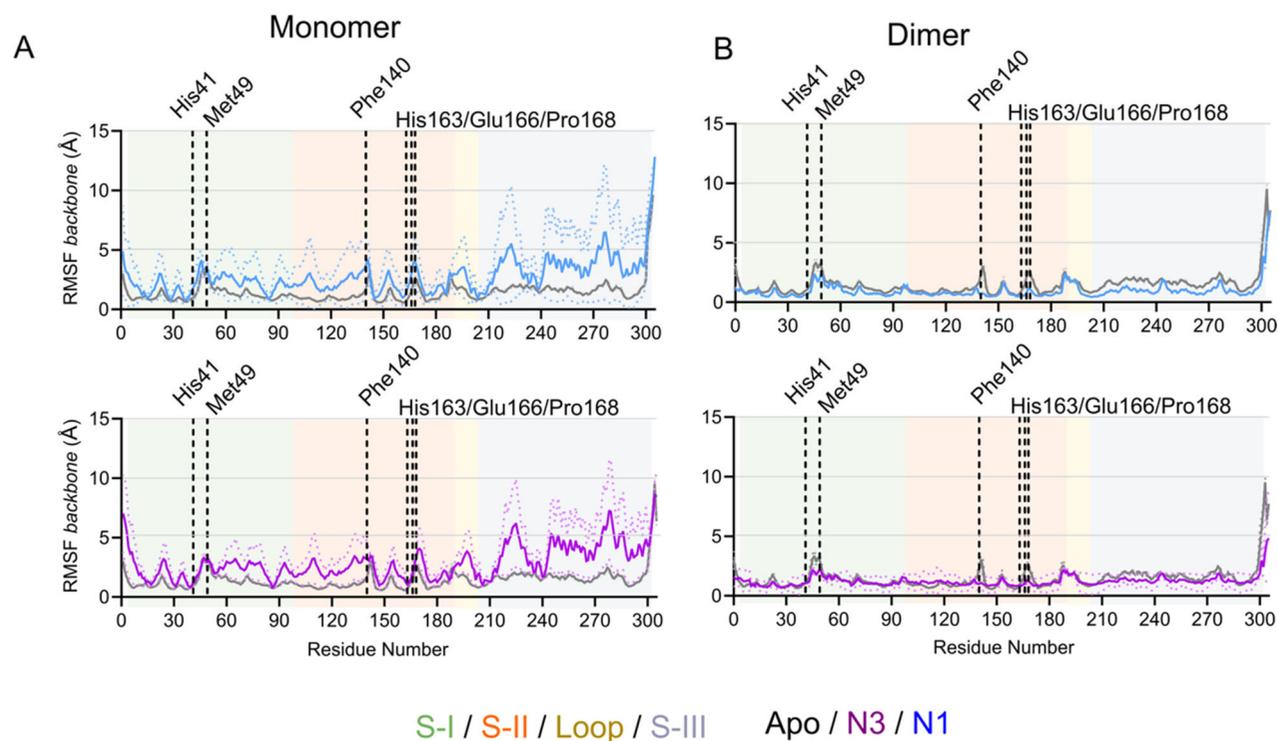


Figure 3. Root-mean-square fluctuation (RMSF) of SARS-CoV-2 M^{Pro} monomeric and dimeric simulations. RMSF graphics with highlighted residues separated by domains for simulations with the monomer (A) and the dimer (B). Mean RMSF are shown in colors for Apo (gray), N3 (sky-blue) and N1 (magenta). Domains are shown in colors: S-I (green), S-II (light orange), Loop (yellow) and S-III (light purple). Apo: apostructure; N1; N3; RMSF: root mean square fluctuation.

dimers are in close interaction with N-finger from the other subunit (Ser1).

Interestingly, the N1 carbonyl group of the pyrrolidine-2-one ring displayed hydrogen bond interactions with His163 (P1) in dimeric simulations (Figure 5A), whereas in N3 the monomeric states had water-mediated interactions. We hypothesize that the Phe140-His163 π - π interaction (with an average distance of ~ 4.1 Å, Figure 5I) would be relevant to lock His163 in a hydrogen bond prone conformation. Moreover, Phe140 had more frequent hydrophobic contacts in the benzyl moiety of the N1 simulations ($>40\%$) than in N3 ($<5\%$) (Figure 5B). This low frequency of hydrophobic contacts in both ligands is in agreement with a previous work, which suggested that the corresponding hydrophobic interactions were not crucial for the inhibition but more relevant in maintaining the His163 hydrogen bonding with the ligand (Ghosh et al., 2020; Zhang et al., 2020).

Biochemical data for N3 ($IC_{50} = 9.0 \mu M \pm 0.8$ on enzymatic inhibition assay) (Wang et al., 2016; Yang et al., 2005) and for N1 ($IC_{50} = 0.6 \mu M \pm 0.1$) (Zhang et al., 2020) indicated similar binding mechanism, with N1 being a more potent inhibitor (Zhang et al., 2020). We suggest that the differences in the His41 (P1') interaction frequency between inhibitors may explain their distinct inhibitory potencies. The inhibitor N1 exclusively showed an interaction between the hydroxyl ethene group and His41 (Figure 5A and E), while N3's ethene group cannot form these interactions (Figure 5D). Further, it was reported that His41 is considered a cold spot, among other homologue sequences, and missense changes led to lack of activity (Krishnamoorthy & Fakhro, 2021).

Importantly, in SARS-CoV2 M^{pro}, the key active site residues His41 (3 mutations), Phe140 (1 mutation), Cys145 (3

mutations), Glu166 (3 mutations), and His172 (1 mutation) showed low mutation frequencies (a total of 11 out of 525 mutations at the active site)

The hydration site in the P3 site can be explored for increasing potency

We also observed that Gln189 (P3 pocket), in both monomer and dimer, established a hydrogen bonding interaction with a the N3 carbonyl ($\sim 40\%$, Figure 5A, D, and E) and a weak water-bridge with N1 pyrrolidine ($\sim 20\%$, Figure 5C–E). As a result of these interactions, the loop Gln189 – Gln192 becomes less flexible in the presence of the inhibitor. This is in line with the previously reported M^{Pro}-inhibitor simulations suggesting that the loop connecting S-II – III would have a decrease in mobility upon inhibitor binding (Suárez & Díaz, 2020).

WaterMap calculations (Abel et al., 2008) were performed to analyze the solvation impact within the M^{Pro}. Specifically, the protein's hydration sites surrounding the residues His41, Cys145, His163, Gln166, and Glu189 (P1, P1', and P3 sites) were calculated. The hydration sites in the P3 site (near the Pro168 and Gln189 residues) had the highest occupancy values ($>80\%$) and free-energy (ΔG) values (3.15 kcal/mol). Specifically, hydration site 2 exhibited the highest occupancy values (0.87-0.89; Figure 6A and B). It has been suggested that P3 site is a conserved hydrophobic pocket and therefore moieties bulkier than 2-pyridone in the P3 region could substantially contribute to increasing the inhibitory potency. This is supported by our results, as displacing those high-energy water molecules would result in stronger binding.

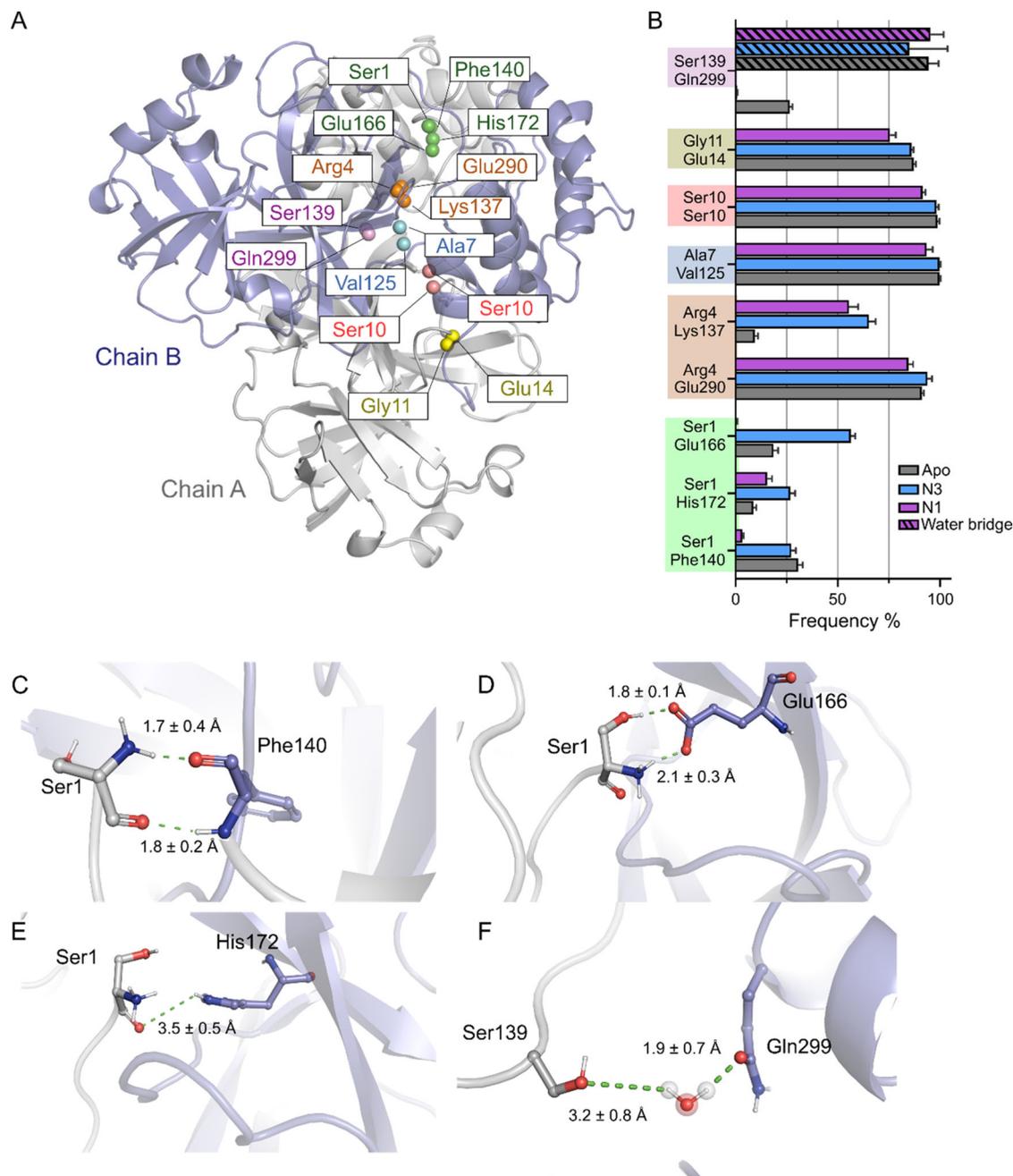


Figure 4. Dimeric interface interactions. (A) Spheres are located in the interface regions of the chains. (B) Frequencies of the dimeric interface hydrogen bonds in the apostructure (gray) or ligands N1 (magenta) and N3 (sky-blue). (C) Hydrogen bond interaction between Ser1A and Phe140B. (D) Hydrogen bond interaction between Ser1A and Glu166B. (E) Hydrogen bond interaction between Ser1A and His172B and (F) Water bridge interaction between Ser139A and Gln299B. Sphere interactions: Ser1-Phe140, Ser1-His172 and Ser1-Glu166 (green); Arg4-Lys137 and Arg4-Glu290 (orange); Ala7-Val125 (cyan); Ser10-Ser10 (salmon); Gly11-Glu14 (yellow) and Ser139-Gln299 (pink).

Previous studies reported that the highly flexible site of M^{pro} (Bzówka et al., 2020) could be addressed by bulkier hydrophobic moieties, however pyrrolidines or amines group were shown to be poor groups to stabilize it (Zhang et al., 2020). In quest for novel drugs to treat SARS-CoV-2 infection, the N1 has already proven functionality and reliable characteristics as a lead compound. Our work will hopefully help a bit this work, a dynamic understanding of the binding mode can be beneficial when developing subsequent strategies, such as scaffold hopping (Böhm et al., 2004) and molecular simplification (Pinacho Crisóstomo et al., 2006). Especially we believe that design larger/bulkier molecules to better occupy

pockets such as **P1** and **P3** (Figure 7) would be beneficial. Finally, we would like to emphasize that although this study discusses the high stability of the dimeric SARS-CoV-2 M^{pro} , by no means can we conclude that the protein would not undergo more extensive conformational changes in simulations with longer timescales (milliseconds).

Conclusion

We report microsecond MD simulations of the SARS-CoV-2 M^{pro} , comparing covalently bound ligand-protein complexes

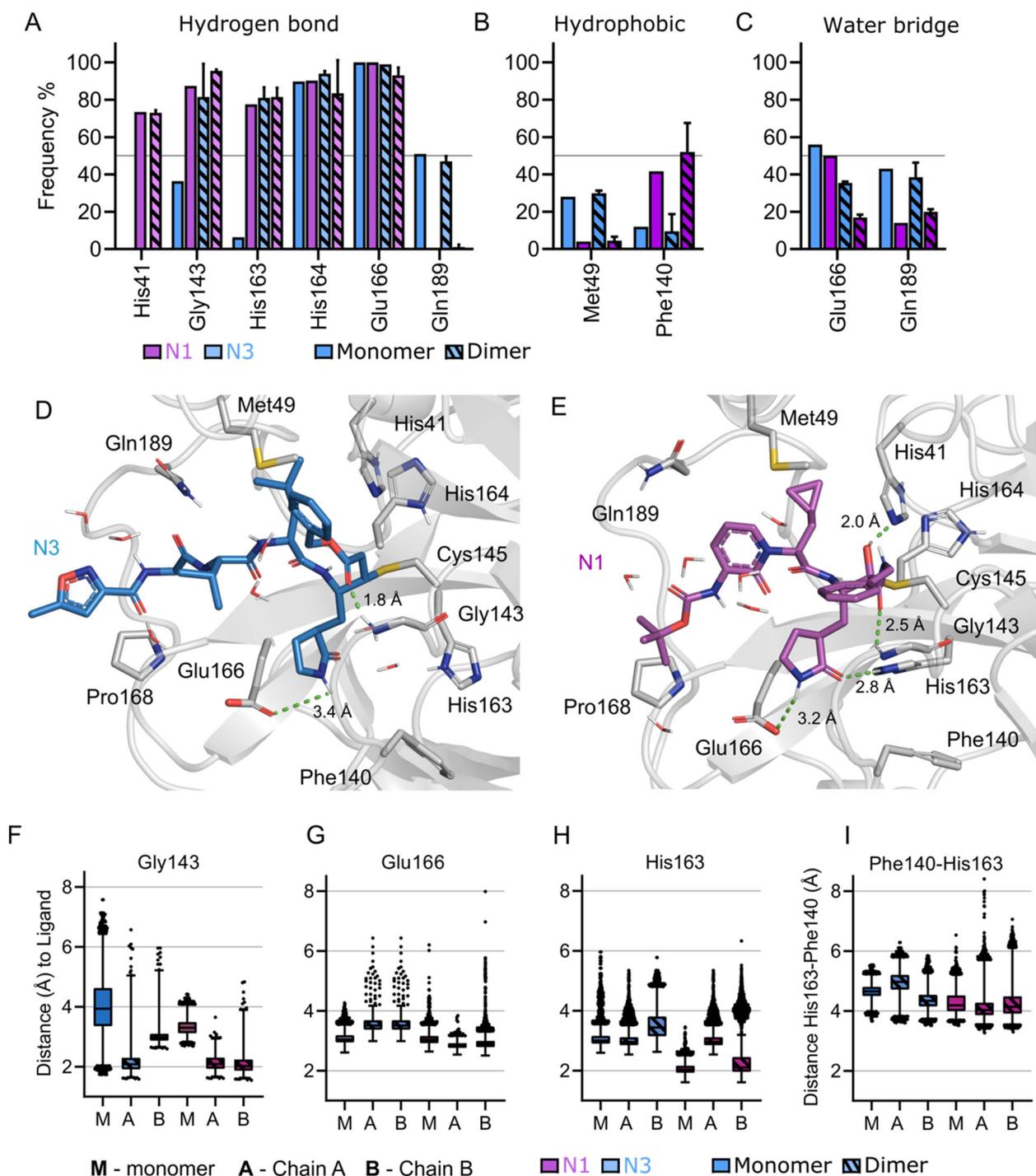


Figure 5. Overall interactions of the SARS-CoV-2 M^{Pro} with inhibitors' sub-pockets (P1, P1', P2 and P3). Frequencies of contacts for hydrogen bonds (A); hydrophobic bonds (B); and water bridge (C) of the M^{Pro} (monomer and dimer) with N1 and N3 ligands. Snapshot frames of interactions between the M^{Pro} dimeric form and the ligands N3(D); and N1(E). Distance between the M^{Pro} amino acid residues (monomer and dimer) and ligands N1 and N3 along the simulations for Gly143 (F); Glu166 (G); His163 (H); Phe140-His163(I). M^{Pro} residues are colored according to the types of atoms in the interacting amino acid residues (protein carbon, light gray; nitrogen, blue; oxygen, red; N3, sky-blue; N1, magenta).

with the apostructure, in both monomeric and dimeric configurations. Simulations with monomeric M^{Pro} have revealed a large conformational change, mainly in the S-III domain. According to the PCA analyses, the dimeric simulations pointed to small conformational changes, being stabilized by a network of hydrogen interactions in the dimerization interface. M^{Pro} is biologically active as a dimer and the results suggested that dimer is more stable

than the monomer, with implications for the oxyanion loop and catalytic sites conformations.

In covalently bound systems, it was observed that the catalytic His41 and Glu166 were keys residues for stabilizing the M^{Pro} 's inhibited state. Additionally, we suggest that substituents bulkier than pyrrolidine could increase activity against SARS-CoV-2 M^{Pro} by occupying the hydrophobic P3 sub-pocket. We envision that this study can set a standard

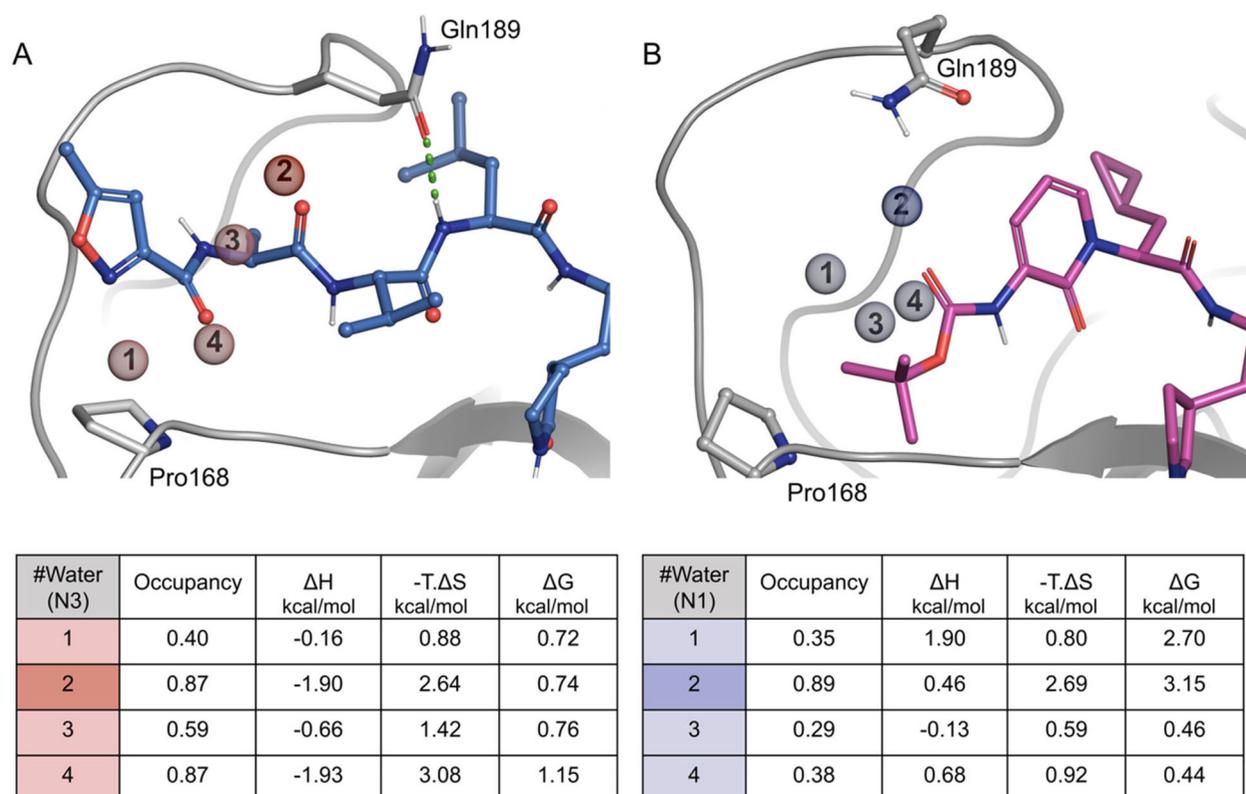


Figure 6. WaterMap in the P3 site of the SARS-CoV-2 M^{pro}. Conserved water of M^{pro} in the presence of ligand N3 (A); and N1 (B). Conserved water molecules are shown in colored spheres, according to their free energy value (ΔG). The adjacent table contains the thermodynamic solvation parameters of four hydration sites in the P3 pocket. N3: Spheres (1,3 and 4) shown in salmon and (2) spheres in red while N1: Spheres (1,3 and 4) shown in violet and (2) spheres in purple. Occupancy: number of water-oxygen atoms that occupy a given hydration site during a short-time simulation (5 ns); ΔG : free-energy value; ΔH : enthalpic energy; $-T \cdot \Delta S$: entropic energy.

for M^{pro} MD simulations and benefit the search for novel bio-active compounds against SARS-CoV-2 M^{pro}.

Materials & methods

SARS-CoV-2 M^{pro} amino acid conservation during evolution

The evolutionary conservation of SARS-CoV-2 M^{pro} amino acid residues was calculated using the ConSurf server (Landau et al., 2005), comparing the M^{pro}'s sequence against known homolog sequences available in PDB. This analysis predicts the conservation score of amino acid residues ranging from 1 to 9, with 1 indicating the least conserved and 9 the highest conserved and provides a structural visualization of it.

Modeling and structure preparation

The SARS-CoV-2 M^{pro} apostructure (PDB: 6y84, Resolution: 1.39 Å) and structures with covalently bound ligands interacting with N1 (PDB: 6y2g, 2.20 Å) and N3 domains (PDB: 6lu7, 2.16 Å) were selected based on the structure's quality and existing ligand.

The selected PDB protein structures were prepared by adding hydrogen atoms and fixing missing side chains using the Protein Preparation Wizard (PrepWiz) (Madhavi Sastry et al., 2013), implemented in the Small Discovery Molecule

Drug Discovery Suite 2019v.3 (Schrödinger LLC, New York, NY, USA). Sulfate ions and other co-crystallization molecules, such as glycerol (GOL) were removed. Within the catalytic site of M^{pro}, His41 can act as a proton shuttle in the catalytic cycle (Pavlova et al., 2021). Accordingly, the His41 (atom NE), His163 (ND), His164 (ND) and His172 (ND) were protonated (apostructures). The His41 ionization and tautomerization states were chosen as previously discussed in (Paasche et al., 2014).

The chosen protein crystals were analyzed according to biological assembly state by using the PISA protein website (Krissinel & Henrick, 2005) (<https://www.ebi.ac.uk/pdbe/pisa/>), to generate their dimeric state. Dimers for the different systems were minimized using Prime (Jacobson et al., 2004), with default options.

Molecular dynamics simulations

Prepared SARS-CoV-2 M^{pro} structures were simulated as apostructures (without ligands) and covalently bound to ligands (N1 and N3). Molecular Dynamics (MD) simulations were carried out by using the Desmond engine (Bowers et al., 2006, 2007) with the OPLS3e force-field (Harder et al., 2016) according to a previously described protocol (Ferreira et al., 2019). OPLS3e accomplishes this by incorporating a broad range of chemical moieties with greater and combining them on-the-fly to generate parameterization, followed by the assignment of partial charges (Roos et al., 2019). In

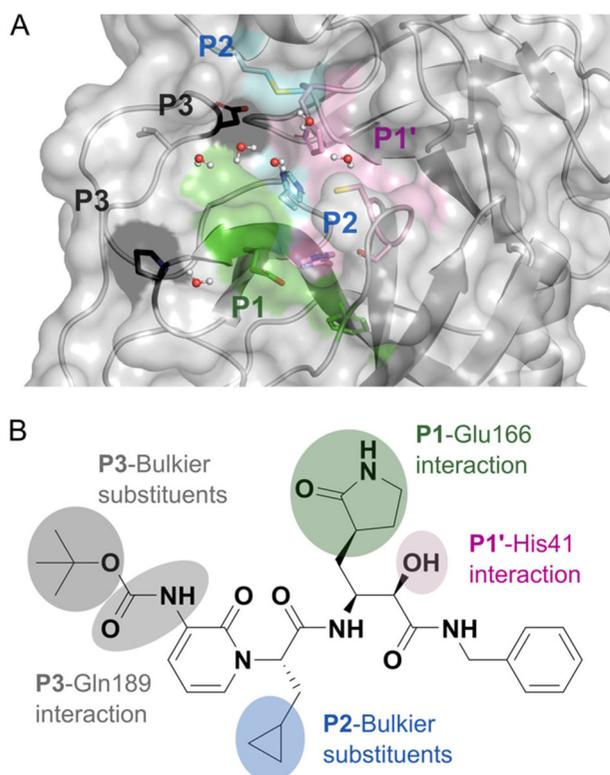


Figure 7. Structure-activity relationship of the inhibition of SARS-CoV-2 M^{pro}. (A) Surface M^{pro} structure representation of all sub-pockets (P1, P1', P2 and P3). (B) Two-dimensional representation of the N1 ligand highlighting interactions with the M^{pro} active site.

short, the system encompassed the protein-ligand/cofactor complex, a predefined water model (TIP3P (Jorgensen et al., 1983)) as a solvent and counterions (Na⁺ or Cl⁻ adjusted to neutralize the overall system charge). The entire system was treated in a cubic box with periodic boundary conditions (PBC), specifying the shape and the size of the box as 13 Å distance from the box edges to any atom of the protein. Short-range coulombic interactions were calculated using 1 fs time steps and 9.0 Å cut-off value, whereas long-range coulombic interactions were estimated using the Smooth Particle Mesh Ewald (PME) method (Darden et al., 1993). Each system was subjected to at least 3 μs simulations (three replicas).

Root mean square deviation (RMSD) values of the protein backbone were used to monitor simulation equilibration and protein folding changes (Figures S2 and S3A, C, and E). The fluctuation (RMSF) by residues was calculated using the initial MD frame as a reference, and compared between ligand-bound and apostructure simulations (Figures S2 and S3B, D, and F). All the trajectory and interaction data are available on the Zenodo repository (code: 10.5281/zenodo.3980660).

Atomic interactions and distances were determined using the Simulation Event Analysis pipeline as implemented in Maestro 2019v.4 (Schrödinger LCC). The criteria for protein-ligand H-bond are 2.5 Å distance between the donor and acceptor atoms (D — H...A); $\geq 120^\circ$ angle between the donor-hydrogen-acceptor atoms (D — H...A); and $\geq 90^\circ$ angle between the hydrogen-acceptor-bonded atoms (H...A — X). Corresponding requirements for protein-water and water-ligand H-bonds are 2.8 Å (D—H...A); $\geq 110^\circ$ (D—H...A); and

$\geq 90^\circ$ (H...A—X). Non-specific hydrophobic interactions are defined by the presence of a hydrophobic side chain within 3.6 Å of the ligand's aromatic or aliphatic carbons. π - π interactions is recorded when two aromatic groups are stacked face-to-face or face-to-edge and within 4.5 Å of distance. MD trajectories were visualized, and figures produced by PyMol v.2.4 (Schrödinger LCC, New York, NY, USA).

Principal Component Analysis (PCA) was used to study the main features of monomeric and dimeric backbone movements. The backbone atoms of chain A and chain B were extracted and aligned using scripts (`trj_selection_dl.py` and `trj_align.py`) from Schrödinger package 2019v.4. Individual simulations from all runs were merged using the `trj_merge.py` script into a final trajectory and CMS file, which was further used to generate the principal components. The actual PCA was done by using the `trj_essential_dynamics.py` script. PCA graphics were generated using the python script, available in the GitHub repository (code: <https://github.com/gmf12/pcaanalysismrpo.git>). All commands were generated using JuPyter (Matplotlib, Seaborn, Numpy, and Pandas).

WaterMap calculations

WaterMap calculations were used to analyze the impact of solvation on the active site region of SARS-CoV-2 M^{pro}. Briefly, short-run MD simulations (5 ns) of the M^{pro} active site (apostructure) were performed using the Desmond molecular dynamic engine with the OPLS3e force field. The binding site was defined to include all protein residues within a 5 Å distance of any atoms in the catalytic dyad (His41 and Cys145) and supporting residues (His163, Gln189 and Glu166). Protein structure was restrained throughout the simulation. Water molecules were clustered into distinct hydration sites. Enthalpy values of the hydration sites were obtained by averaging over the non-bonded interaction for each water molecule in the cluster. Entropy values were calculated using numerical integration of the local expansion of the entropy in terms of spatial and orientational correlation functions (Young et al., 2007).

Acknowledgements

The authors thank the CSC-IT Center for Science, Finland, for the generous computational resources. Authors also would like to thank Prof. Dr. Ewen MacDonald and Prof. Dr. Vinícius G. Maltarollo for the critical reading and language corrections.

Author contributions

G.M.F and T.K. designed the experiments, executed and analyzed the simulations. A.K.T. performed the WaterMap analysis. R.D.C.H., M.H.H and A.P. helped with the discussion and writing. M.H.H and A.P. contributed to resources and overall study supervision. All authors prepared and reviewed the manuscript.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This study was partially supported by CNPq (Grant # 447120/2014-0) and FAPESP (Grant # 2016/12899-6 and 2019/24112-9), Brazil. MHH and RDCH are recipients of fellowships from CNPq, Brazil. GMF is a recipient of a fellowship from FAPESP, Brazil.

References

- Abel, R., Young, T., Farid, R., Berne, B. J., & Friesner, R. A. (2008). Role of the active-site solvent in the thermodynamics of factor Xa ligand binding. *Journal of the American Chemical Society*, *130*(9), 2817–2831. <https://doi.org/10.1021/ja0771033>
- Bello, M. (2020). Prediction of potential inhibitors of the dimeric SARS-CoV2 main proteinase through the MM/GBSA approach. *Journal of Molecular Graphics & Modelling*, *101*, 107762. <https://doi.org/10.1016/j.jmgm.2020.107762>
- Böhm, H. J., Flohr, A., & Stahl, M. (2004). Scaffold hopping. *Drug Discovery Today Technology*, *1*(3), 217–224. <https://doi.org/10.1016/j.ddtec.2004.10.009>
- Bowers, K. J., Chow, D. E., Xu, H., Dror, R. O., Eastwood, M. P., Gregersen, B. A., Klepeis, J. L., Kolossvary, I., Moraes, M. A., Sacerdoti, F. D., Salmon, J. K., Shan, Y., & Shaw, D. E. (2007). Scalable algorithms for molecular dynamics simulations on commodity clusters. Proceedings of the ACM/IEEE Conference on Supercomputing (SC06), 2006, Tampa, FL, 43. <https://doi.org/10.1109/sc.2006.54>
- Bowers, K. J., Sacerdoti, F. D., Salmon, J. K., Shan, Y., Shaw, D. E., Chow, E., Xu, H., Dror, R. O., Eastwood, M. P., Gregersen, B. A., Klepeis, J. L., Kolossvary, I., & Moraes, M. A. (2006). *Molecular dynamics—Scalable algorithms for molecular dynamics simulations on commodity clusters* [Paper presentation]. Proceedings of the 2006 ACM/IEEE Conference on Supercomputing - SC '06, 84. <https://doi.org/10.1145/1188455.1188544>
- Bzówka, M., Mitusińska, K., Raczynska, A., Samol, A., Tuszyński, J. A., & Góra, A. (2020). Structural and evolutionary analysis indicate that the sars-CoV-2 mpro is a challenging target for small-molecule inhibitor design. *International Journal of Molecular Sciences*, *21*(9), 3099–4016. <https://doi.org/10.3390/ijms21093099>
- Chen, H., Wei, P., Huang, C., Tan, L., Liu, Y., & Lai, L. (2006). Only one protomer is active in the dimer of SARS 3C-like proteinase. *The Journal of Biological Chemistry*, *281*(20), 13894–13898. <https://doi.org/10.1074/jbc.M510745200>
- Chen, S., Chen, L., Tan, J., Chen, J., Du, L., Sun, T., Shen, J., Chen, K., Jiang, H., & Shen, X. (2005). Severe acute respiratory syndrome coronavirus 3C-like proteinase N terminus is indispensable for proteolytic activity but not for enzyme dimerization. Biochemical and thermodynamic investigation in conjunction with molecular dynamics simulations. *The Journal of Biological Chemistry*, *280*(1), 164–173. <https://doi.org/10.1074/jbc.M408211200>
- Chou, C. Y., Chang, H. C., Hsu, W. C., Lin, T. Z., Lin, C. H., & Chang, G. G. (2004). Quaternary structure of the severe acute respiratory syndrome (SARS) coronavirus main protease. *Biochemistry*, *43*(47), 14958–14970. <https://doi.org/10.1021/bi0490237>
- Coronavirus Disease (COVID-19) Situation Reports. (2021). Retrieved June 15, from <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/>
- 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, H. (2020). Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease. *Science (New York, NY)*, *368*(6497), 1331–1335. <https://doi.org/10.1126/science.abb4489>
- Darden, T., York, D., & Pedersen, L. (1993). Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *The Journal of Chemical Physics*, *98*(12), 10089–10092. <https://doi.org/10.1063/1.464397>
- Ferreira, G. M., Kronenberger, T., De Almeida, É. C., Sampaio, J., Terra, C. F., Pinto, E., & Trossini, G. H. G. (2019). Inhibition of porcine aminopeptidase M (pAMP) by the pentapeptide microginins. *Molecules (Basel, Switzerland)*, *24*(23), 4369–4383. <https://doi.org/10.3390/molecules24234369>
- Ghosh, A. K., Brindisi, M., Shahabi, D., Chapman, M. E., & Mesecar, A. D. (2020). Drug development and medicinal chemistry efforts toward SARS-Coronavirus and Covid-19 therapeutics. *ChemMedChem*, *15*(11), 907–932. <https://doi.org/10.1002/cmdc.202000223>
- Harder, E., Damm, W., Maple, J., Wu, C., Reboul, M., Xiang, J. Y., Wang, L., Lupyan, D., Dahlgren, M. K., Knight, J. L., Kaus, J. W., Cerutti, D. S., Krilov, G., Jorgensen, W. L., Abel, R., & Friesner, R. A. (2016). OPLS3: A force field providing broad coverage of drug-like small molecules and proteins. *Journal of Chemical Theory and Computation*, *12*(1), 281–296. <https://doi.org/10.1021/acs.jctc.5b00864>
- Henzler-Wildman, K., & Kern, D. (2007). Dynamic personalities of proteins. *Nature*, *450*(7172), 964–972. <https://doi.org/10.1038/nature06522>
- Hsu, W. C., Chang, H. C., Chou, C. Y., Tsai, P. J., Lin, P. I., & Chang, G. G. (2005). Critical assessment of important regions in the subunit association and catalytic action of the severe acute respiratory syndrome coronavirus main protease. *The Journal of Biological Chemistry*, *280*(24), 22741–22748. <https://doi.org/10.1074/jbc.M502556200>
- Jacobson, M. P., Pincus, D. L., Rapp, C. S., Day, T. J. F., Honig, B., Shaw, D. E., & Friesner, R. A. (2004). A hierarchical approach to all-atom protein loop prediction. *Proteins*, *55*(2), 351–367. <https://doi.org/10.1002/prot.10613>
- 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, H. (2020). Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature*, *582*(7811), 289–293. <https://doi.org/10.1038/s41586-020-2223-y>
- Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, *79*(2), 926–935. <https://doi.org/10.1063/1.445869>
- Komatsu, T. S., Okimoto, N., Koyama, Y. M., Hirano, Y., Morimoto, G., Ohno, Y., & Tajji, M. (2020). Drug binding dynamics of the dimeric SARS-CoV-2 main protease, determined by molecular dynamics simulation. *Scientific Reports*, *10*(1), 16986. <https://doi.org/10.1038/s41598-020-74099-5>
- Krishnamoorthy, N., & Fakhro, K. (2021). Identification of mutation resistance coldspots for targeting the SARS-CoV2 main protease. *IUBMB Life*, *73*(4), 670–675. <https://doi.org/10.1002/iub.2465>
- Krissinel, E., & Henrick, K. (2005). Detection of protein assemblies in crystals. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 3695 LNBI, 163–174. https://doi.org/10.1007/11560500_15
- Landau, M., Mayrose, I., Rosenberg, Y., Glaser, F., Martz, E., Pupko, T., & Ben-Tal, N. (2005). ConSurf 2005: The projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res*, *33*(Web Server issue), W299–W302. <https://doi.org/10.1093/nar/gki370>
- Madhavi Sastry, G., Adzhigirey, M., Day, T., Annabhimoju, R., & Sherman, W. (2013). Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments. *Journal of Computer-Aided Molecular Design*, *27*(3), 221–234. <https://doi.org/10.1007/s10822-013-9644-8>
- Mendoza, E. J., Manguiat, K., Wood, H., & Drebot, M. (2020). Two detailed plaque assay protocols for the quantification of infectious SARS-CoV-2. *Current Protocols in Microbiology*, *57*(1), 105. <https://doi.org/10.1002/cpmc.105>
- Paasche, A., Zipper, A., Schäfer, S., Ziebuhr, J., Schirmeister, T., & Engels, B. (2014). Evidence for substrate binding-induced zwitterion formation in the catalytic Cys-His Dyad of the SARS-CoV main protease. *Biochemistry*, *53*(37), 5930–5946. <https://doi.org/10.1021/bi400604t>
- Pavlova, A., Lynch, D. L., Daidone, I., Zanetti-Polzi, L., Smith, M. D., Chipot, C., Kneller, D. W., Kovalevsky, A., Coates, L., Golosov, A. A., Dickson, C. J., Velez-Vega, C., Duca, J. S., Vermaas, J. V., Pang, Y. T., Acharya, A., Parks, J. M., Smith, J. C., & Gumbart, J. C. (2021). Inhibitor binding influences the protonation states of histidines in SARS-CoV-2 main protease. *Chemical Science*, *12*(4), 1513–1527. <https://doi.org/10.1039/D0SC04942E>
- Penman, S. L., Kiy, R. T., Jensen, R. L., Beoku-Betts, C., Alfirevic, A., Back, D., Khoo, S. H., Owen, A., Pirmohamed, M., Park, B. K., Meng, X.,

- Goldring, C. E., & Chadwick, A. E. (2020). Safety perspectives on presently considered drugs for the treatment of COVID-19. In *British Journal of Pharmacology* (Vol. 177, Issue 19, pp. 4353–4374). John Wiley and Sons Inc. <https://doi.org/10.1111/bph.15204>
- Peterson, L. (2020). COVID-19 and Flavonoids: In silico molecular dynamics docking to the active catalytic site of SARS-CoV and SARS-CoV-2 main protease. *SSRN Electronic Journal*, <https://doi.org/10.2139/ssrn.3599426>
- Pillaiyar, T., Manickam, M., Namasivayam, V., Hayashi, Y., & Jung, S. H. (2016). An overview of severe acute respiratory syndrome-coronavirus (SARS-CoV) 3CL protease inhibitors: Peptidomimetics and small molecule chemotherapy. In *Journal of Medicinal Chemistry*, 59(14), 6595–6628. <https://doi.org/10.1021/acs.jmedchem.5b01461>
- Pinacho Crisóstomo, F. R., Carrillo, R., León, L. G., Martín, T., Padrón, J. M., & Martín, V. S. (2006). Molecular simplification in bioactive molecules: Formal synthesis of (+)-muconin. *The Journal of Organic Chemistry*, 71(6), 2339–2345. <https://doi.org/10.1021/jo0524674>
- Raschke, T. M. (2006). Water structure and interactions with protein surfaces. *Current Opinion in Structural Biology*, 16(2), 152–159. <https://doi.org/10.1016/j.sbi.2006.03.002>
- Roos, K., Wu, C., Damm, W., Reboul, M., Stevenson, J. M., Lu, C., Dahlgren, M. K., Mondal, S., Chen, W., Wang, L., Abel, R., Friesner, R. A., & Harder, E. D. (2019). OPLS3e: Extending force field coverage for drug-like small molecules. *Journal of Chemical Theory and Computation*, 15(3), 1863–1874. <https://doi.org/10.1021/acs.jctc.8b01026>
- Rut, W., Groborz, K., Zhang, L., Sun, X., Zmudzinski, M., Pawlik, B., Wang, X., Jochmans, D., Neyts, J., Młynarski, W., Hilgenfeld, R., & Drag, M. (2021). SARS-CoV-2 Mpro inhibitors and activity-based probes for patient-sample imaging. *Nature Chemical Biology*, 17(2), 222–228. <https://doi.org/10.1038/s41589-020-00689-z>
- Silvestrini, L., Belhaj, N., Comez, L., Gerelli, Y., Lauria, A., Libera, V., Mariani, P., Marzullo, P., Ortore, M. G., Palumbo Piccionello, A., Petrillo, C., Savini, L., Paciaroni, A., & Spinozzi, F. (2021). The dimer-monomer equilibrium of SARS-CoV-2 main protease is affected by small molecule inhibitors. *Scientific Reports*, 11(1), 9283. <https://doi.org/10.1038/s41598-021-88630-9>
- Suárez, D., & Díaz, N. (2020). SARS-CoV-2 main protease: A molecular dynamics study. *Journal of Chemical Information and Modeling*, 60(12), 5815–5831. <https://doi.org/10.1021/acs.jcim.0c00575>
- Wang, F., Chen, C., Tan, W., Yang, K., & Yang, H. (2016). Structure of main protease from human coronavirus NL63: Insights for wide spectrum anti-coronavirus drug design. *Scientific Reports*, 6, 22677. <https://doi.org/10.1038/srep22677>
- Wang, Y.-C., Yang, W.-H., Yang, C.-S., Hou, M.-H., Tsai, C.-L., Chou, Y.-Z., Hung, M.-C., & Chen, Y. (2020). Structural basis of SARS-CoV-2 main protease inhibition by a broad-spectrum anti-coronaviral drug. *American Journal of Cancer Research*, 10(8), 2535–2545. www.ajcr.us/
- 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. (2020). A new coronavirus associated with human respiratory disease in China. *Nature*, 579(7798), 265–269. <https://doi.org/10.1038/s41586-020-2008-3>
- Yang, H., Xie, W., Xue, X., Yang, K., Ma, J., Liang, W., Zhao, Q., Zhou, Z., Pei, D., Ziebuhr, J., Hilgenfeld, R., Kwok, Y. Y., Wong, L., Gao, G., Chen, S., Chen, Z., Ma, D., Bartlam, M., & Rao, Z. (2005). Design of wide-spectrum inhibitors targeting coronavirus main proteases. *PLoS Biology*, 3(10), e324. <https://doi.org/10.1371/journal.pbio.0030324>
- Young, T., Abel, R., Kim, B., Berne, B. J., & Friesner, R. A. (2007). Motifs for molecular recognition exploiting hydrophobic enclosure in protein-ligand binding. *Proceedings of the National Academy of Sciences of the United States of America*, 104(3), 808–813. <https://doi.org/10.1073/pnas.0610202104>
- Zhang, C.-H., Stone, E. A., Deshmukh, M., Ippolito, J. A., Ghahremanpour, M. M., Tirado-Rives, J., Spasov, K. A., Zhang, S., Takeo, Y., Kudalkar, S. N., Liang, Z., Isaacs, F., Lindenbach, B., Miller, S. J., Anderson, K. S., & Jorgensen, W. L. (2021). Potent noncovalent inhibitors of the main protease of SARS-CoV-2 from molecular sculpting of the drug perampanel guided by free energy perturbation calculations. *ACS Central Science*, 7(3), 467–475. <https://doi.org/10.1021/acscentsci.1c00039>
- Zhang, L., Lin, D., Sun, X., Curth, U., Drosten, C., Sauerhering, L., Becker, S., Rox, K., & Hilgenfeld, R. (2020). Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α -ketoamide inhibitors. *Science (New York, NY)*, 368(6489), 409–412. <https://doi.org/10.1126/science.abb3405>