SARS-CoV-2 Main Protease:

A Molecular Dynamics Study

Dimas Suárez and Natalia Díaz*

Departamento de Química Física y Analítica. Universidad de Oviedo. Avda. Julián Clavería 8.

33006. Oviedo. Asturias. SPAIN.

Corresponding Author

* diazfnatalia@uniovi.es. Phone-Number +34-985103468.

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval

to the final version of the manuscript. Both authors contributed equally.

ABSTRACT.

Herein, we investigate the structure and flexibility of the hydrated SARS-CoV-2 main protease by means of 2.0 μs Molecular Dynamics (MD) simulations in explicit solvent. After having performed electrostatic pK_a calculations on several X-ray structures, we consider both the native (unbound) configuration of the enzyme and its non-covalent complex with a model peptide, Ace-Ala-Val-Leu-Gln~Ser-Nme, which mimics the polyprotein sequence recognized at the active site. For each configuration, we also study their monomeric and homodimeric forms. The simulations of the unbound systems show that the relative orientation of domain III is not stable in the monomeric form and provide further details about inter-domain motions, protomer-protomer interactions, inter-residue contacts, accessibility at the catalytic site, etc. In the presence of the peptide substrate, the monomeric protease exhibits a stable interdomain arrangement, but the relative orientation between the scissile peptide bond and the catalytic dyad is not favorable for catalysis. By means of comparative analysis, we further assess the catalytic impact of the enzyme dimerization, the actual flexibility of the active site region and other structural effects induced by substrate binding. Overall, our computational results complement previous crystallographic studies on the SARS-CoV-2 enzyme and, together with other simulation studies, should contribute to outline useful structure-activity relationships.

INTRODUCTION

At the end of 2019, a novel coronavirus was identified in the airway epithelial cells of three patients with pneumonia of unknown cause.[1] The new pathogen, named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2),[2] spread rapidly around the globe and the associated COVID-19 disease become a worldwide pandemic. Social distancing and quarantine were imposed in numerous countries to stop dissemination, and the development of effective drugs and/or vaccines against the virus was urged.

Similar to other coronaviruses, two overlapping polyproteins (replicase polyproteins 1a and 1ab) are produced and processed after SARS-CoV-2 infection, to generate multiple functional subunits required for viral replication.[3] The proteolytic processing is accomplished by two internally-encoded proteases that hydrolyze the polyproteins at specific sites. One of these proteolytic enzymes, the so-called main protease or 3C-like protease ($3CL^{pro}$), catalyzes most maturation cleavage events. Thus, $3CL^{pro}$ is an essential enzyme for viral replication and constitutes one of the best characterized drug targets among coronaviruses.

The crystal structure of the $3CL^{pro}$ protein from SARS-CoV-2 is highly similar to that of other coronaviruses.[4, 5]. The protein is formed by three domains: domains I (residues 10-99) and II (residues 100-182) have an antiparallel β-barrel structure, while domain III (residues 198-303) forms a compact α-helical domain connected to domain II by a long linker-loop. The active site is located in a cleft between domains I and II, and it holds a histidine/cysteine catalytic dyad. In the SARS-CoV-2 main protease, Cys_{145} acts as a nucleophile during the first step of the hydrolysis reaction assisted by His_{41} as a base catalyst. The oxyanion hole, which stabilizes the partial negative charge developed at the P_1 carbonyl group of the peptide substrate during the

hydrolysis of the $P_1 \sim P_1$' bond, is formed by the backbone amino groups of Gly_{143} and Cys_{145}, both placed in the so termed oxyanion loop (residues 138-146). The catalytic machinery also included a number of binding sites, with the S_1 site defining the enzyme specificity for a glutamine at the P_1 position of the peptide substrate. Domain III is involved in the dimerization of the protease (see below), and the resulting homodimer is supposed to be the active form of the enzyme.[6]

The high homology shown by the 3CL[pro] enzyme in different coronaviruses suggests that this enzyme family exhibits almost identical biochemical and biophysical properties. Particularly, the main proteases from SARS-CoV-2 and SARS-CoV share a 96% of sequence identity so that much of the experimental results obtained for the later enzyme could be relevant for the present case.[4] Hence, before outlining our goals in this work, we first summarize some biochemical and structural studies on the SARS-CoV main protease.[7]

The critical role of domain III in the 3CL[pro] dimerization has been settled by fragment deletion experiments performed for the SARS-CoV protease, showing that a truncated enzyme lacking domain III remains as a monomer.[8] The catalytic activity of this truncated form, which has been assayed against a 14-mer peptide substrate, is very weak as only ~20 % of the substrate was cleaved after 20 h.[8] Furthermore, serial truncation experiments have confirmed that the last *C*-terminal helix in domain III is essential for dimerization and enzyme activity and, more specifically, deletion of residues Gln_{299} and Arg_{298} significantly reduces both dimerization and enzyme activity (~1-2 %).[9, 10]

Another structural element of 3CL[pro] that seems essential for the activity of the SARS-CoV main protease is the *N*-finger constituted by residues 1-7 in domain I. According to Chen *et*

al.,[11] the complete deletion of the *N*-finger has a minor effect on dimerization, but it abolishes enzymatic activity (<1%) when reacting with a 12-mer peptide. On the other hand, Hsu *et. al.* [9] have found that the first three *N*-finger residues have only a small influence both on the dimer stability and on the activity of the enzyme (K_d for dimer dissociation changes from 0.28 µM to 3.4 µM upon deletion while 76% of catalytic activity is retained). However, in contrast with Chen *et. al.*, the serial truncation experiments of Hsu *et. al.* have revealed a dramatic effect in the structure and activity of the enzyme after removing Arg_4 (K_d=57.5 µM and 1.3% of activity in the truncated enzyme) and the next residues.[9]

Crystallographic structures have shed light onto the relationship between dimerization and enzyme activity. In the homodimer structures, the *N*-finger of one protomer is squeezed between domains II and III of another protomer. This allows Ser_1 and Arg_4 from one protomer to shape the substrate-specificity pocket (*i.e.* the S_1 site) and the oxyanion loop in the other protomer.[12] Curiously, in the crystal structures for some monomers (mutants G11A, S139A, and R298A), [13-15] the oxyanion loop is partially folded as a 3_{10}-helix that hinders the access to the active site, particularly to the S_1 subsite. Therefore, it seems that both protein dimerization and the right placement of the *N*-finger residues are indispensable for maintaining an open active site. This interpretation has been supported by surface plasmon resonance experiments showing that the full-length enzyme binds a P_6-P_1 hexapeptide model (K_d~152 µM), but almost no binding exists to the *N*-finger deleted protease.[11]

From the structural and activity results enumerated above, it seems that the monomeric form of the 3CL^{pro} enzymes has little or none affinity for peptide substrate, presumably due to the blocking conformation of the oxyanion loop in the active site. However, this view has been challenged by isothermal titration calorimetry (ITC) assays showing that the wild type and

several monomeric mutants of SARS-CoV (R298A, R298L, and R298A/Q299A) present comparable binding affinities for a 6-mer peptide substrate.[16] Moreover, substrate binding to the single mutants induces the dimerization of the enzyme, and kinetic assays provided similar k_{cat} values for the single mutants and the wild-type protein. In contrast, dimer stabilization upon substrate binding does not occur for the double mutant, which exhibits null proteolytic activity.[16] Another intriguing fact is that some of the crystal structures obtained for the 3CL[pro] enzyme present a dimer with one of the monomers displaying a partially-collapsed or disordered oxyanion loop, showing thus that the blocking conformation is not exclusive of the monomeric state.[12, 15] Moreover, it has been reported that the N214A mutant remains mainly as a dimer without enzymatic activity, but structurally very close to the wild-type enzyme (*i.e.* no collapsed active site is observed in the crystal structure).[17] Therefore, it seems most likely that a complex interplay exists among enzyme dimerization, active site structure, substrate binding and catalysis.

Clearly, some of the questions regarding the actual role of the protease dimerization on the architecture and activity of the SARS-CoV-2 catalytic site are prone to computational examination by means of molecular dynamics (MD) simulations in explicit solvent. Similarly, computational studies can reveal important molecular details of the Michaelis complexes with peptide substrates. Hence, in this work, we examine various configurations of the SARS-CoV-2 3CL[pro] enzyme differing in the monomer or homodimer state and in the presence or not of a short peptide substrate. All the models are subject to a 2.0 μs MD simulation followed by intensive analysis in order to characterize many structural and dynamic features ranging from the tertiary and quaternary structure to specific inter-residue contacts at the protomer interface or in the active site. By comparing the results obtained for the various configurations, the differences and similarities between the monomeric and dimer forms are discussed in detail, revealing also some

effects induced by substrate binding on the interdomain arrangements and the pre-reactive organization of the catalytic site. In this way, our computational results may complement previous crystallographic studies on the SARS-CoV and SARS-CoV-2 enzymes. Considering also that the spread of the COVID-19 disease has undoubtedly sparked a flurry of computational research on this system, the present results may also contribute to reach a consensus view about the actual flexibility and structure of the active site supported by independent simulation studies. Eventually, the representative structures produced by our simulations could be of interest to undertake further computational work.

METHODS

pK$_a$ calculations

The protonation state for the titratable residues at pH 7 were assigned according to structure-based pK$_a$ calculations performed with the H++ web server (version 3.1).[18] In these calculations, four high-resolution X-ray structures recently deposited at the Protein Data Bank (PDB) were considered: two apo-enzyme structures 6Y84 (1.39 Å)[19] and 6M03 (2.0 Å)[20] and two enzyme-inhibitor complexes 6LU7 (2.16 Å)[5] and 5R82 (1.31 Å)[21] (coordinates of the inhibitor atoms were removed). The pK$_a$ calculations were performed for the protease monomers and for the catalytically-competent dimeric forms of the apo-enzyme structures. In addition, three different dielectric constants were selected for the interior of the protein (ε_{int}=4, 10, and 20) to evaluate the consistency of the results.

Initial structures and building of the enzyme/peptide complex

Starting coordinates for the SARS-CoV-2 main protease were taken from the 6LU7 (pH=6.0) PDB structure. The coordinates of the inhibitor were removed prior to the edition and simulation of the native form of the enzyme, and all the crystallographic water molecules were maintained in the models. We also simulated an enzyme/substrate complex as the SARS-CoV-2 main protease performs multiple cleavages at viral polyproteins by recognizing several peptide sequences with an absolutely conserved glutamine at the P_1 site (*i.e.* cleavage occurs at the $P_1 \sim P_1$' peptide bond). Hence, to generate a dynamic model for substrate binding within the 3CL^{pro} active site, we selected the sequence of the cleavage site between the so-called nonstructural protein 4 and the 3CL^{pro} enzyme that are successively encoded in the viral polyproteins (*i.e.* scissile peptide bond 3263~3264). According to the information pre-released in the UniProt database (codes P0DTC1 and P0DTD1),[22] the -P_4-P_3-P_2-$P_1 \sim P_1$'-sequence of this recognition site is -Ala-Val-Leu-Gln~Ser-. We built this short peptide sequence within the protease active site by using the 6LU7 crystal structure as a template. Thus, the peptidomimetic inhibitor in 6LU7 includes a peptide moiety with the sequence -Ala-Val-Leu- located within the S_4-S_3-S_2 binding sites and a consecutive side chain bound within S_1 that resembles a glutamine side chain (see Scheme 1). Hence, we kept this part of the inhibitor and we merely built the Ser backbone chain at the S_1' site using the *Chimera* program.[23] During the molecular edition of the non-covalent enzyme/pentapeptide complex, the Ser(P_1') side chain, the *N*-terminal acetyl- and *C*-terminal *N*-methyl amide capping groups, and all H atoms in the peptide substrate were added by the tLEaP program included in the AMBER18 suite of programs.[24, 25]
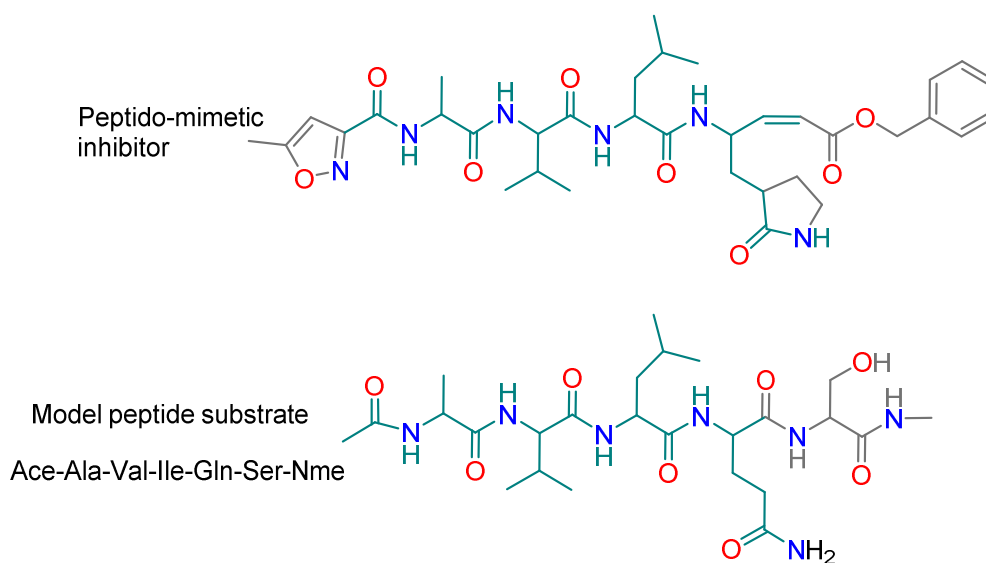
Scheme 1. Comparison between the molecular structures of the peptido-mimetic inhibitor and the peptide substrate selected for this work.

Molecular Dynamics Settings

Molecular Dynamics (MD) simulations in explicit solvent were run for the monomeric and dimeric forms of the enzyme, either in their free (unbound) state or in complex with a pentapeptide substrate. The coordinates of the solute atoms from the selected X-ray structure were processed with the tLEaP program in order to add the missing H atoms and to assign the molecular mechanics parameters. The systems, which were represented with the ff14SB version[26] of the all-atom AMBER force field,[27] were immersed in an octahedral water box that extended 18 Å (monomer) or 20 Å (dimer) from the protein atoms. The TIP3P potential[28] was used to represent the water molecules and Na^+ counterions[29] were added with the tLEaP program to neutralize the systems.

The solvent molecules and counterions were initially relaxed by means of energy minimizations and 100 *ps* of molecular dynamics (MD) using the SANDER program.[25] Then the

full systems were minimized and heated gradually to 300 K using 60 *ps* of constant volume (NVT) MD with a 1 fs time step and using the PMEMD program in AMBER18. Subsequently, the density was adjusted by means of 2.0 *ns* of constant pressure (NPT) MD with a 2 *fs* time step and using the Monte Carlo barostat as implemented in PMEMD. Langevin dynamics was employed to control the temperature (300 *K*) with a collision frequency of 2 ps^{-1}. The SHAKE algorithm [30] was selected to constraint all R-H bonds, and periodic boundary conditions were applied to simulate a continuous system at constant pressure (NPT). A non-bonded cutoff of 9.0 Å was used and the Particle-Mesh-Ewald method[31] was employed to include the contributions of long-range interactions. The production phase of the simulations at the NPT conditions extended up to 2.0 *μs* and coordinates were saved every 2.5 *ps* for analysis. The MD runs with a time step of 2.0 *fs* employed the GPU accelerated version of the PMEMD code included in AMBER18.[25;32]

Structural Analysis of the MD simulations

The CPPTRAJ software[33] in AMBER18 was used to compute the root mean squared deviation (RMSD) of the protein coordinates with respect to the reference X-ray structure along the MD trajectories. The coordinates of the models were also clustered using CPPTRAJ with the average-linkage clustering algorithm and a sieve of 250 frames. The distance metric between frames was calculated via best-fit coordinate RMSD using the coordinates of heavy atoms. The molecular surface of the whole systems and of the domain/monomer components were computed using the linear combination of pairwise overlaps (LCPO) method[34] as implemented in CPPTRAJ. Secondary structure assignment of the oxyanion loop was done using the 2002 CMBI version of the DSSP program.[35]

H-bond and vdW contacts were characterized using a specific software developed locally. H-bonds were identified based on a geometrical criterion (X···Y distance < 3.5 Å and X-H···Y angle > 120º), whereas hydrophobic interactions were scored by evaluating a dispersion attraction term[36] between pairs of atoms belonging to different hydrophobic groups. The criteria for assessing the occurrence of dispersion interactions between two groups were: (a) the total pairwise dispersion energy is larger than 0.5 kcal/mol in absolute value; (b) the distance between the centers of mass of the two interacting groups is below 12.0 Å. The Chimera visualization system[23] was employed to draw the ribbon/stick models of the systems.

Using a locally-developed FORTRAN code, the relative orientation of the protein domains and/or of the two protomers in the dimeric form, were monitored in terms of the Euler angles (*xyx* convention) that characterize the relative orientation of two rigid coordinate systems, which are placed at the center of mass of the considered fragments. Each coordinate system is defined by the principal inertia axes, which, in turn, are computed considering the coordinates of the backbone atoms located in the α-helical or β-strand elements within the selected region.

To further characterize the structure and shape of the active site region, we computed the radius of accessibility (r_{acc}) of various residue side chains and/or H-bond sites following a computational protocol that has been described in previous work.[37] For each atom or group of atoms, r_{acc} is defined as the maximum radius of a spherical ligand that can touch the desired target. The MSMS program[38] was used to carry out fast computations of molecular surfaces considering only the protein heavy atoms and probe spheres of varying radius. The r_{acc} values were calculated for a subset of 2000 snapshots for each simulation.

Electrostatic calculations

The electrostatic potential of the models in solution was computed on selected cluster representatives from the MD trajectories using the APBS software.[39] In the electrostatic Poisson-Boltzmann calculations, only the coordinates of the solute atoms were used, the atomic charges and radii being taken from the ff14SB AMBER representation. The non-linear PB equation was solved on a cubic lattice by using an iterative finite-difference method. The cubic lattice had a grid spacing of 0.33 Å and the points at the boundary of the grid were set to the sum of Debye-Hückel potentials. The dielectric boundary was the contact surface between the radii of the solute and the radius (1.4 Å) of a water probe molecule. The electrostatic potential was plotted onto the molecular surface computed by the MSMS program[38] using Chimera program.

Conformational Entropy

We estimated the conformational entropy ($S_{conform}$) of the backbone ϕ and ψ dihedral angles for residues located nearby the active site using the CENCALC program.[40] The entropy calculation relies on the discretization of the time evolution of the selected dihedral angles. To this end, the continuous probability density function (PDF) of each dihedral angle is represented by a von Mises kernel density estimator, which depends on a concentration parameter κ (a $\kappa=0.50$ value was chosen here). By finding the maxima and minima of the PDF, the time series containing the values of the corresponding dihedral angle during the MD simulation is transformed into an array of integer numbers labelling the accessible conformational states. This allows the estimation of the probability mass functions (P_i) of the individual dihedral angles from which the marginal (first-order) conformational entropy of the each dihedral is computed as,

$$S^1_{conform} = -R \sum P_i \ln P_i$$

Typically, μs-length MD simulations are required in order to obtain reasonably converged $S_{conform}$ values.[41]
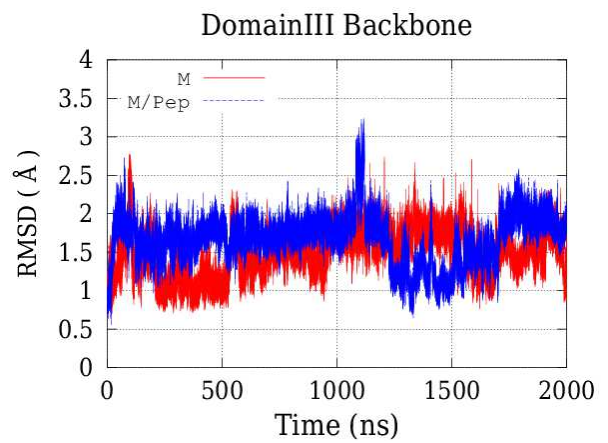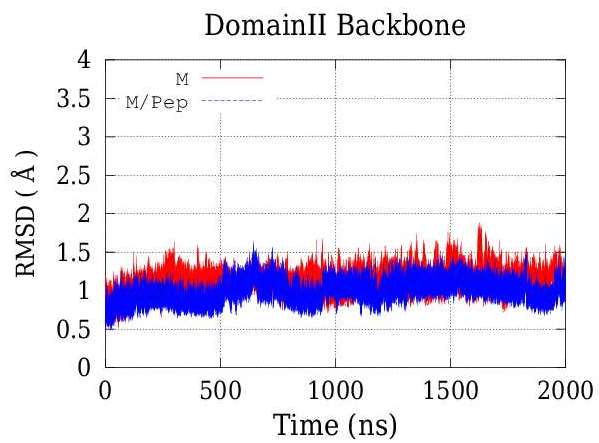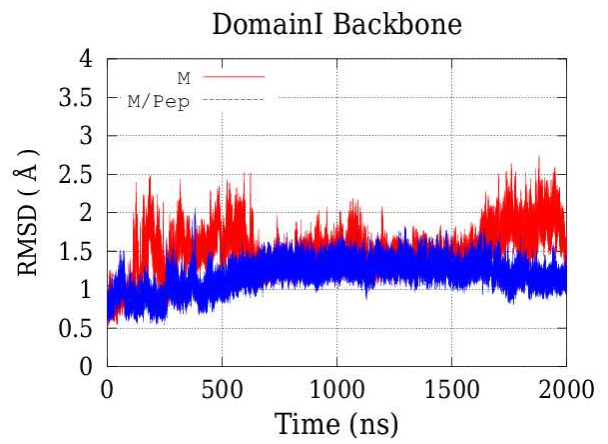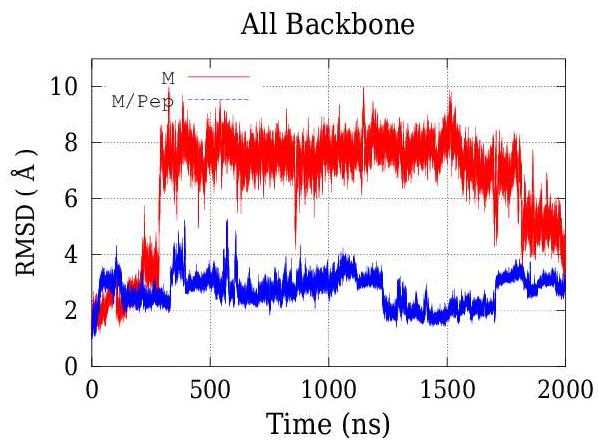
Results and Discussion

pK_a Calculations

Table S1 in the Supporting Information collects the computed pK_a values for all the titratable residues of the SARS-CoV-2 main protease. The crystal structures used in the calculations were selected according to their state (apo-enzyme or inhibited-enzyme), their atomic resolution (1.3-2.2 Å), and the experimental pH value (6.0-8.1). The results consistently predict that all lysines, arginines, aspartic and glutamic acids should be modelled in their charged state at pH 7. In contrast, all cysteines and tyrosines should remain neutral. For the imidazole histidine groups, their pK_a values are well below 7.0, suggesting thus that they will be mainly neutral at pH 7. Only the results obtained for His_{80} at high dielectric constants, more reliable for this solvent accessible residue, advice the coexistence of neutral and protonated forms. However, the computed pK_a values below 7.0 and the absence of negatively charged residues close to His_{80} prompted us to consider its neutral state. Then the most probable neutral state for the six histidines, with the side chain protonated at Nδ or Nε atoms, was selected after visual inspection of their contacts in the crystal structures. This resulted in His_{41} and His_{80} being protonated at Nδ whereas His_{64}, His_{163}, His_{164}, His_{172}, and His_{246} were protonated at Nε. We also note in passing that pK_a calculations on the dimer 3CL[pro] structures do not introduce any changes in the selected protonation states. In addition, the pK_a values obtained for the active site residues, particularly for the catalytic dyad Cys_{145}/His_{41}, support their neutral state within the pH range 6-8 corresponding to the selected crystal structures.
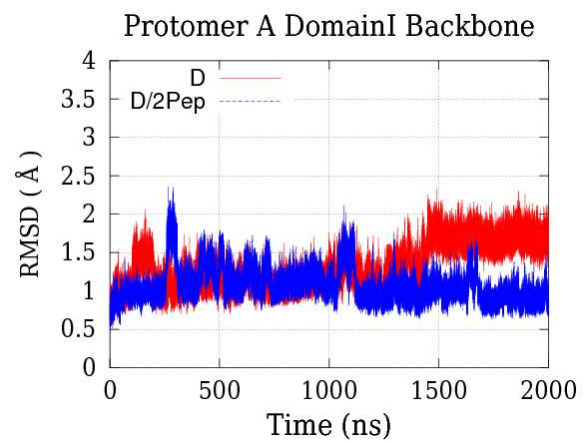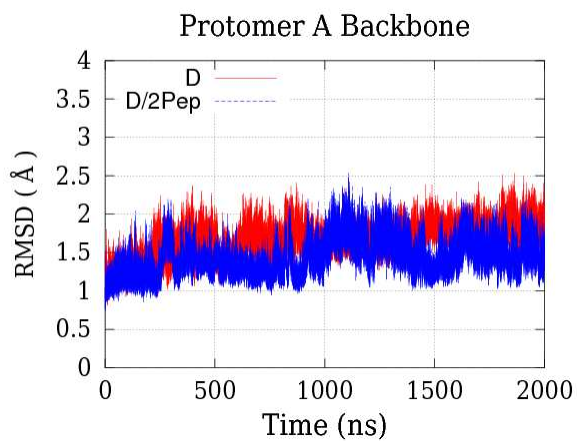
MD simulations

We examined various protease configurations that involve either the monomeric or the dimeric forms, which, in turn, simulate either the native (unbound) state or the bound state with the Ace-Ala-Val-Leu-Gln-Ser-Nme peptide substrate (see Table S2 in the Supporting Information). In particular, we run two 2-µs MD trajectories representing the native state that are labelled as **M** (monomer) and **D** (dimer). The enzyme-peptide complex was sampled by two MD trajectories labelled as **M/Pep** and **D/2Pep**, corresponding to the monomeric and dimeric forms, respectively. In **D/2Pep** the active site of each protomer A and B accommodates one peptide molecule. The MD results of the various simulations will be presented and analyzed in a comparative manner in order to better characterize the structure and flexibility of the protein domains, the amplitude of the inter-domain relative motions, the nature and stability of the protomer···protomer contacts, the shape and flexibility of the active site, the substrate mode of binding, etc. All these results can be also useful to outline more clearly the differences and similarities between the monomeric and dimeric forms of the enzyme in aqueous solution.

a)

All Backbone

DomainI Backbone

DomainII Backbone

DomainIII Backbone

b)

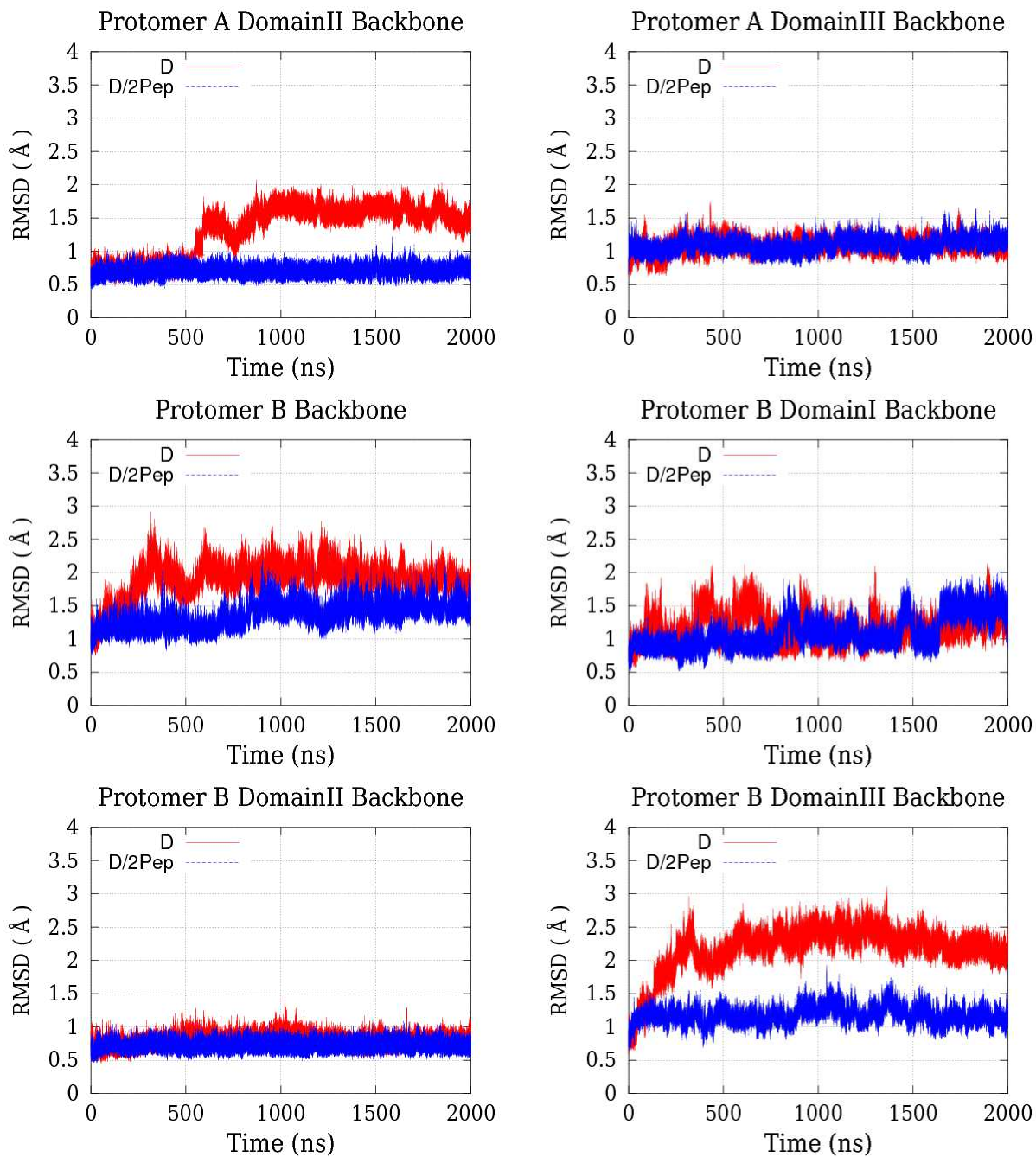Protomer A Backbone

Protomer A DomainI Backbone

Figure 1. (a) Time evolution along the **M** (in red) and **M/pep** (in blue) trajectories of the root mean squared deviation (RMSD) computed for selected backbone heavy atoms (in Å). (b) Time evolution along the **D** (in red) and **D/2Pep** (in blue) trajectories of the RMSD data (in Å).

Intra-domain structure and relative domain orientation

Changes in the internal geometry and in the relative position of the various protein domains were first analyzed by monitoring the time evolution of the RMSD of the coordinates of the heavy atoms with respect to the crystallographic structures (Figure 1; Table S3 presents mean RMSD values). In addition, the superposition of the solid-state and the MD-averaged ribbon models (Figures 2-3) reveals some differences in the positioning of specific loops and secondary structure elements.

The rotational motion of the helical domain III with respect to the central β-strand domain II along the **M** simulation is a remarkable result concerning the overall protein architecture (see Figure 1). In the coordinates of protomer A in the 6LU7 structure, domain III establishes only a few H-bond contacts with domain II (*e.g.* $Th_{111}\cdots Asp_{295}$ H-bond, $Arg_{131}\cdots Asp_{289}$ salt-bridge) and domain I (*e.g.* $Arg_4\cdots Gln_{299}$ H-bonds), and domains II/III are connected through a 14-long peptide linker (Gly_{183}–Asp_{197}). In contrast, the connecting loop between domains I and II comprises only 7 residues (Asp_{92}-Pro_{99}) including a central Pro_{96} residue and abundant interdomain contacts contribute to define the active site region. In fact the domain I/II arrangement was quite stable in all of our MD simulations.
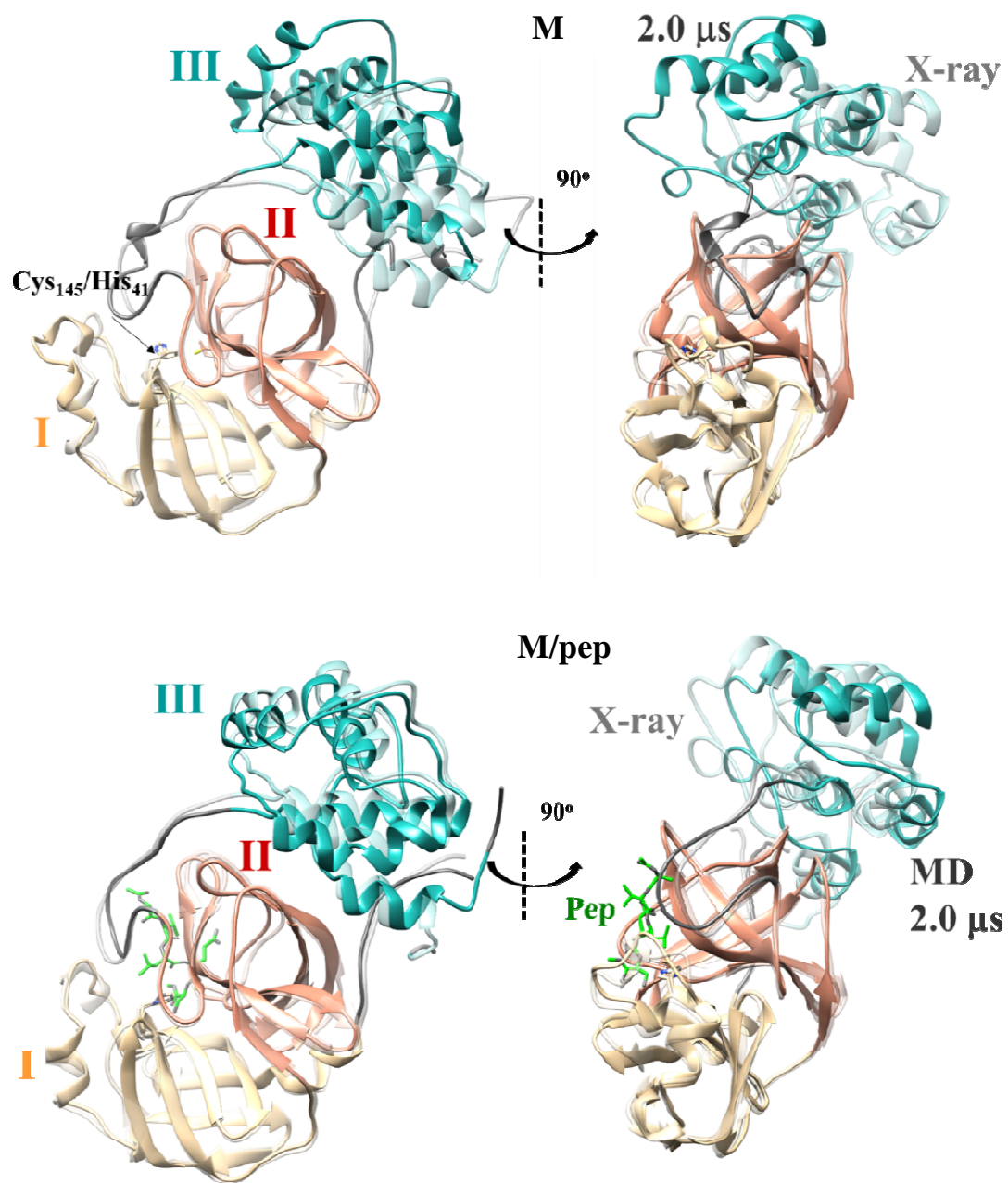
Figure 2 Different views (90º turned) for a ribbon representation of the average structure obtained from the last 50 *ns* of the **M** and **M/pep** simulations. The average structure was superposed over the 6LU7 X-ray structure (in lighter color) using the backbone coordinates of the domain II.

The relative motion of domain III is well described by the time evolution of the RMSD values for all the backbone atoms and by the center of mass (COM) distance between the domains II and III (see Figures 1 and S2). To characterize the interdomain orientation regardless of the COM separation, we also computed the Euler angles defined by a reference inertial system placed at the domain II and an analogous coordinate system located at domain III, the resulting ϕ, θ and ψ angles being shown in the form of polar plots collected in Figure S1. In the **M** simulation, the two domains slightly depart from each other (the COM distance elongates from ~26 to ~29 Å) at 200 *ns* without significant reorientation. During the ~200-300 *ns* interval, an ample reorientation occurred in less than 15 *ns* facilitated by a torsional change at Gly_{195} and the relative position remained stable during more than 1 μs. However, during the last 500 *ns* of the **M** simulation, domain III reorients again with respect to domain II adopting an intermediate pose in terms of the Euler angles (see Figure S1), which also departs significantly from the X-ray structure as shown in Figure 2.

To further assess whether or not the varying location of domain III is an intrinsic feature of the monomeric state in aqueous solution, we decided to run a second MD simulation started from a different X-ray structure (6Y84) with the same settings as those employed in the **M** simulation. For the sake of brevity, the structural details of this simulation, which was run for 1.0 μs, are not reported here. We just comment that a two-step interdomain reorientation occurred again within the 300-500 *ns* time interval, leading to an average MD structure that resembles the first configuration observed in the **M** trajectory (see Figure S4). Therefore, the independent MD trajectory added support to the results produced by the **M** simulation.
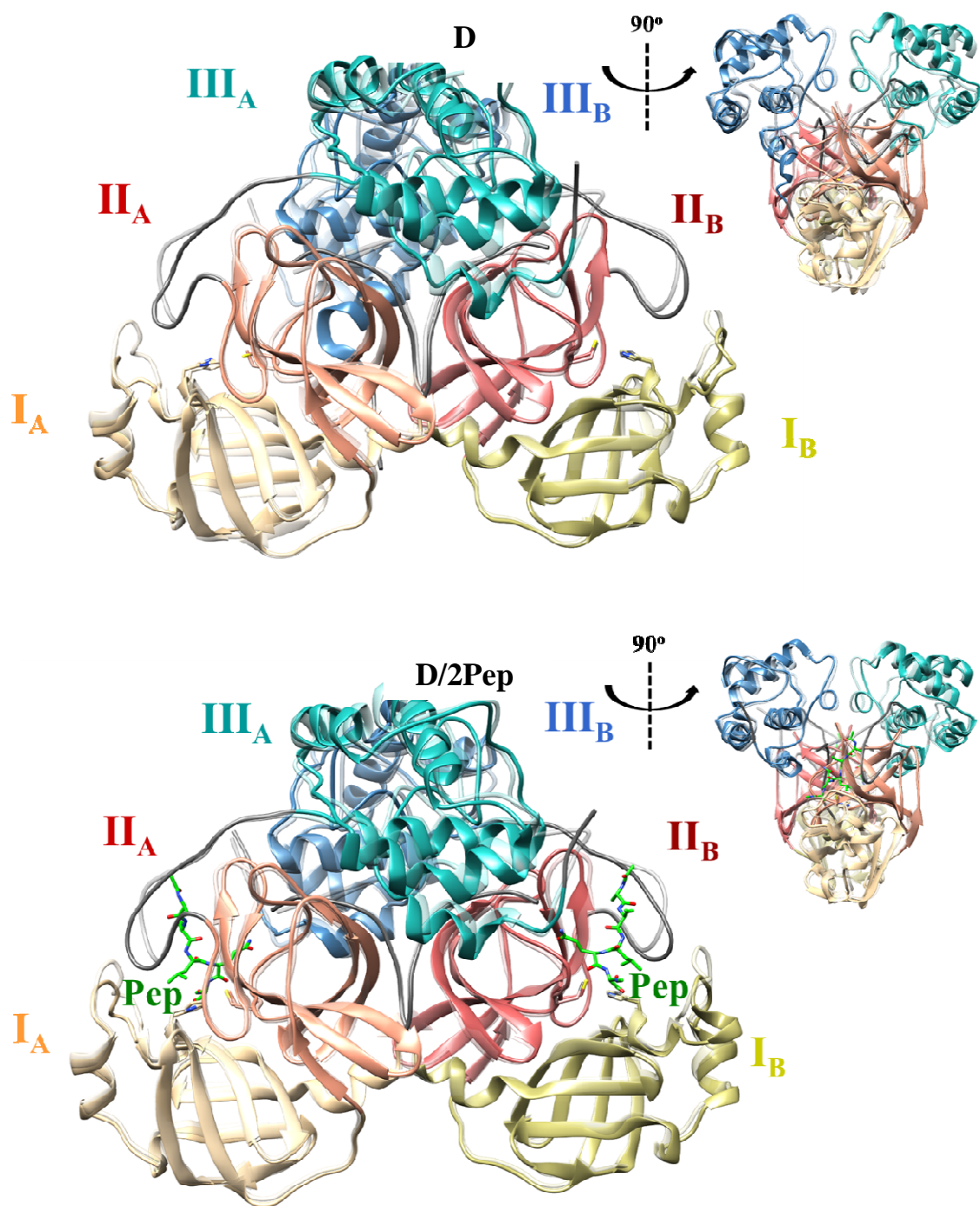
Figure 3 Different views for a ribbon representation of the average structure obtained from the last 50 *ns* of the **D** and **D/2Pep** simulations. The average structures were superposed over the 6LU7 X-ray structure (in lighter color) using the backbone coordinates of the domain II.

Interestingly, the wide domain II/III rearrangement was not observed in the presence of the peptide substrate (**M/Pep** simulation) although the COM separation and the reference Euler angles present significant fluctuations, specially in the first half of the simulation, indicating loosened interdomain contacts. The peptide substrate gives direct H-bond contacts with linker residues (*e.g.*, Gln_{189} and Gln_{192}), what could contribute to reduce the mobility of the II-III linker. On the other hand, the interdomain motions are largely dampened out in the dimer simulations (**D** and **D/2Pep**), which are characterized by a persistent interdomain disposition and stable interprotomer contacts (see below). Thus, it turns out that the destabilization of the interdomain orientation would occur only in the monomeric form.

Concerning the internal structure of the domains I-III, the most remarkable feature is the small structural deviations and low flexibility of the central domain II (see Figure 1 and Table S3). Thus, the mean RMSD values are only ~1.1±0.1 Å for the monomeric simulations (**M** and **M/Pep**) and even smaller for the dimer models (~0.8±0.1 Å) with the only exception of the domain IIA in the **D** simulation (1.52±0.21 Å; see below). The internal stability of domain II is in consonance with a certain buried character as only ~55-60% (monomer simulations) or ~40-45% (dimers) of its molecular surface is solvent-accessible (see Table S4). The terminal domains I and III are more hydrophilic, having around 70% and 85% of exposed surface, respectively, all along the MD simulations. They exhibit wider internal motions involving their helical and loop elements, which, however, do not induce dramatic changes in the domain structures, the largest RMSD values being lower than 2.5 Å. In general, the largest intradomain deviations arise in the native trajectories (**M** and **D**) as, for example, in the domain IIIB in **D** (RMSD=2.30±0.16 Å), whose *C*-terminal helix is largely displaced (see Figure 3). In this respect, it turns out that the local internal displacements and the actual flexibility in domains I/III are quite variable when

comparing among the data of the various simulations and/or of the A/B protomers in the **D** simulation. This variability suggests that the protein domains may access to different conformational states on the µs time scale. Finally, we note that the **D/2Pep** simulation consistently shows the smaller structural deviations and the lower fluctuations, both in terms of the intradomain RMSD data and the interdomain COM distances/Euler angles, revealing thus a rigidifying effect exerted by substrate binding.

Protomer···Protomer interactions

As in other 3CL^{pro} enzymes, only the SARS-CoV-2 main protease homodimer is considered to be catalytically active and, accordingly, targeting the interface between both protomers (A and B) might be an alternative therapeutic strategy for the treatment of COVID-19. Hence, we decided to analyze in detail the overall dimer architecture and the inter-residue contacts between the two protomers. According to the protomer···protomer distance in Figure S2, the dimer remains stable during the **D** and **D/2Pep** simulations. By means of molecular surface calculations (Table S4), we found that the A···B contact area consists of ~11% (~1400 Å^2) of the LCPO surface of the separated protomers and also remains quite stable along the **D** simulation. This contact area in **D** is similar to that in the X-ray structure (~1400 Å^2 in 6LU7), albeit lower than in the **D/2Pep** simulation (~1600 Å^2). The two protomers are oriented perpendicular to one another. Their relative orientation, as measured by the Euler angles between the corresponding inertial reference systems, is highly stable according to the MD simulations (Figure S1), what is anyway compatible with internal motions within the respective protomer domains. The stability of the dimer architecture is also evident in the good overall match between the average MD structures and their parent X-ray structures (see Figure 3). The global RMSD mean values further confirm

this agreement as they have moderate values of 2.91±0.12 Å (**D**), and 1.78±0.16 Å (**D/2Pep**), which, in turn, point out again the substrate binding effect.

In the crystallographic structures, the two protomers give symmetrical contacts (*e.g.*, the $Ser_1@O\cdots HN@Phe_{140*}$ and $Phe_{140}@NH\cdots O@Ser_{1*}$ H-bonds have the same interatomic distances; see Table S6). The A⋯B contacts mainly involve residues Ser_1-Gly_{11} in one protomer and several residues distributed in the three domains of the other protomer (*e.g.* $Ser_{10}\cdots Ser_{10}$, $Arg_4\cdots Lys_{137}$, $Met_6\cdots Val_{125}$, and $Arg_4\cdots Glu_{290}$). There are also a few non polar contacts between residues in domain II (*e.g.* $Val_{125}\cdots Val_{125}$) and between residues in domain III (*e.g.* $Ala_{285}\cdots Leu_{286}$). Most of these contacts are well maintained during the **D**, and **D/2pep** simulations (See Table S6). For instance, the $Gly_{11}@NH\cdots O\varepsilon@Glu_{14}$ H-bond presents a 100% of abundance and a short interatomic distance (2.9 Å) in the two dimer trajectories. But the percentage of occurrence of other polar interactions, mainly involving Ser_1 and Arg_4 residues in the *N*-finger, varies across the simulations and with respect to the crystal structure. Thus, the initial $Ser_1\cdots Glu_{166}$ H-bonds are missing for most part of the simulations (*i.e.* the highest percentage of occurrence is below 50%). Similarly, the $Ser_1\cdots His_{172}$ contact is weakened in the simulated dimers, and the $Ser_1\cdots Phe_{140}$ interaction is not abundant for the **D** trajectory. With respect to Arg_4, the salt bridge with Glu_{290} is well preserved but the $Lys_{137}(A)\cdots Arg_4(B)$ interaction is clearly destabilized (see Table S6A). On the other hand, the simulations also assess the relevance of non-polar interactions for the stability of the dimer. Thus, the Met_6/Tyr_{126}, Pro_9/Pro_{122}, Pro_9/Val_{125}, and Pro_9/Leu_{115} contacts, which connect the *N*-finger in one protomer to domain II in the other protomer, have a high abundance and a large scoring energy. Similarly, the Val_{125}/Val_{125} contact glues the domain IIA and IIB in all the simulations, while the hydrophobic packaging of the Ala_{285} and Leu_{286} residues in domains IIIA and IIIB also

contributes to fix the highly mobile *C*-terminal domains. This hydrophobic cluster involving residues Ala_{285}/Leu_{286} is a particularly important inter-protomer contact spot. The $Ala_{285}\cdots Leu_{286}$ average MD distances are clearly shorter for the **D/2Pep** trajectory (see Table S6B), which indicates that the long linker in domain III containing Ala_{285} and Leu_{286} is more packed in the dimer in the presence of the peptide substrate. This would explain the increase in the contact area between the protomers along **D/2Pep**.

Domain II in one protomer is also linked to domain III in the other protomer thanks to the $Ser_{123}\cdots Arg_{298}$ and $Ser_{139}\cdots Gln_{299}$ H-bonds. The $Ser_{123}\cdots Arg_{298}$ contact is mainly mediated by a water molecule during the simulations and it is clearly desestabilized along **D/2Pep**. In contrast, the presence of the substrate contributes to stabilize the $Ser_{139}\cdots Gln_{299}$ interactions (see Table S6A) that connect the oxyanion loop in one protomer to the *C*-terminal helix in the other protomer.

Structure and dynamics of the active site in the unbound form of the enzyme

The 3CL^{pro} active site is located at a shallow crevice in the I/II inter-domain region. The nucleophilic Cys_{145} belongs to the oxyanion loop, which is an *S*-shaped loop (Gly_{138}-Gly_{146}) in the domain II. The amide nitrogens of Cys_{145} and Gly_{143} define the "oxyanion hole", which binds the carbonyl group of the scissile peptide bond in the substrates. Adjacent to the oxyanion loop, a β-strand segment (His_{163}-Pro_{168}) comprises other residues that play an important role in substrate binding (*e.g.*, His_{163}, Glu_{166}), which, in turn, are close to a loop segment (Gly_{183}-Ala_{193}) placed at the beginning of the domain II/III linker that contributes to border the active site region.

(a) Cluster-1 (16 %)   **(b) Cluster-2 (14 %)**

(c)   **(d)**

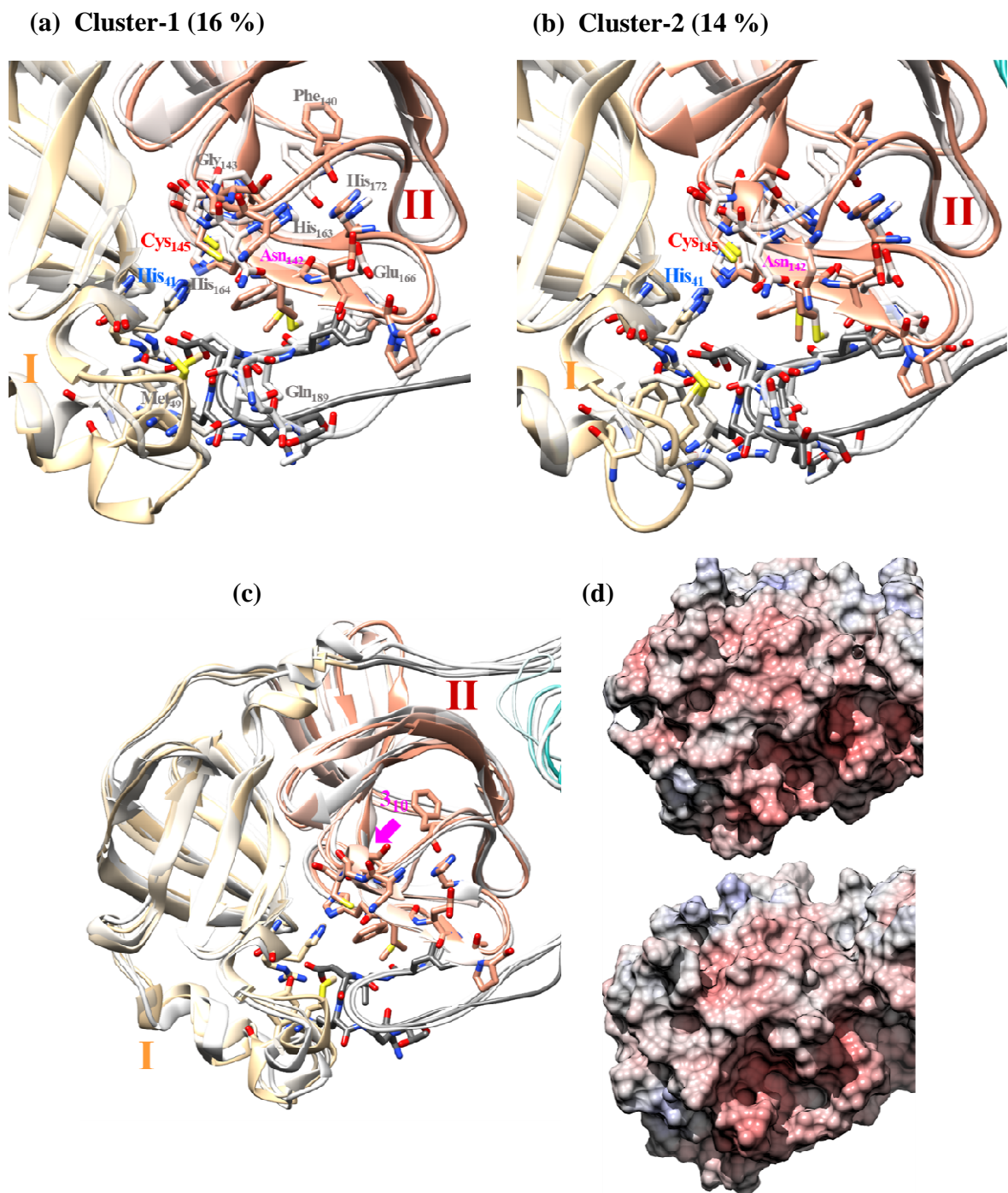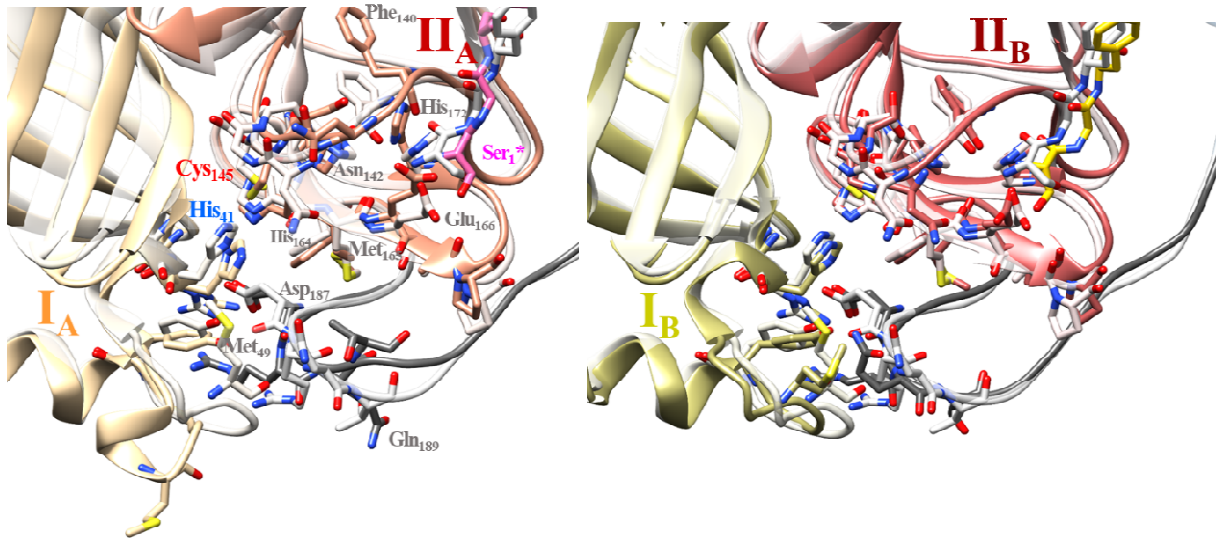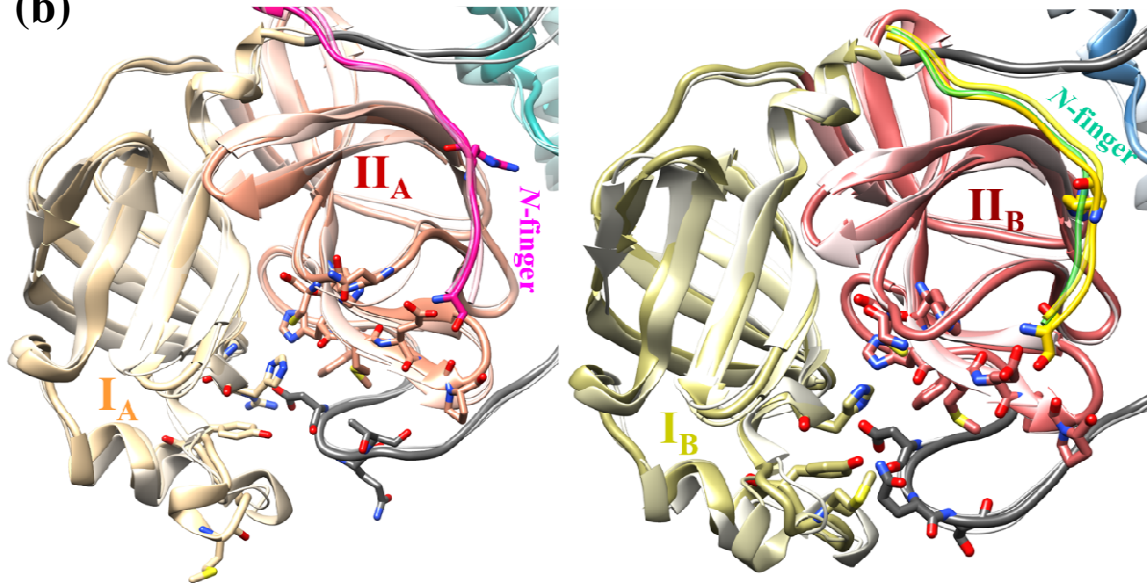**Figure 4** (a-b) View of the active site in the **M** simulation as shown by the superposition of the two most important cluster onto the X–ray structure (shown in lighter colors) (c) Superposition of top 5 cluster representatives (coil thickness and color intensity are proportional to cluster abundance). (d) Electrostatic potential mapped onto the surface in the two most populated clusters.
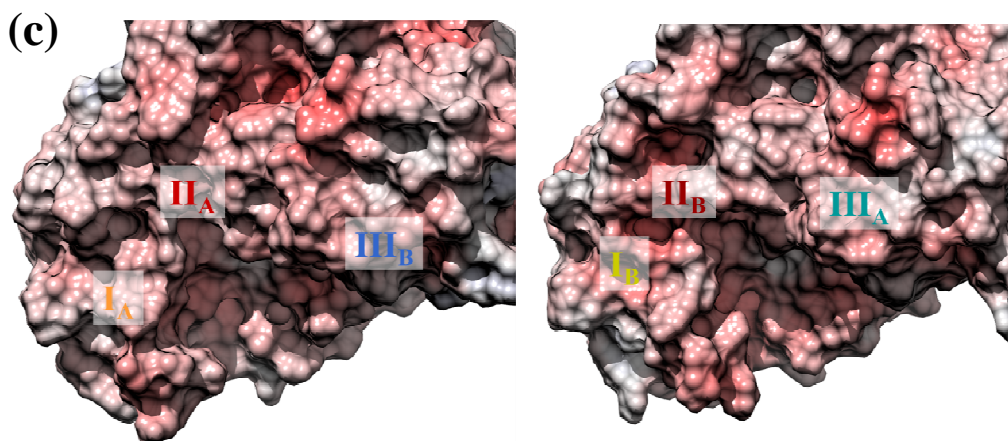
(a)

(b)

(c)

Figure 5 (a) View of the active site region in each protomer (A and B) along the **D** simulation as shown by the superposition of the top-1 clusters onto the X–ray structure (shown in lighter colors) (b) Superposition of the cluster representatives (coil thickness and color intensity are proportional to cluster abundance). (c) Electrostatic potential mapped onto the solvent-accessible protein surfaces.

Besides Cys_{145}, the catalytic dyad also includes His_{41} that is supposed to act as a base during the activation of the nucleophile. His_{41} belongs to a small helix that is placed within a long and highly helical connection loop in domain I. The positioning of His_{41} is assisted by a network of H-bonds including a $His_{41}\cdots Asp_{187}$ interaction mediated by a conserved water molecule and a salt-bridge between the nearby Arg_{40} guanidinium and the Asp_{187} carboxylate at the domain II/III linker. The active site is completed by several inter-domain non-polar interactions among Pro_{39}, Met_{49}, Leu_{27} in domain I and His_{164}, Met_{165} in domain II that constitute a hydrophobic pocket (site S_2) for accommodating the peptide substrate.

To describe the structure and flexibility of the active site in the MD simulations of the unbound configurations, we performed first clustering calculations that yield the population and representative structures, which allow us to visualize structural deviations with respect to the X-ray geometries. A more detailed description is provided by the statistical analysis of selected H-

bond/vdW contacts, the secondary structure analysis of the oxyanion loop, etc. We also measured the accessibility of the catalytic residues and binding sites in terms of the average accessibility radii (Table 1). The main results are displayed in Figures 4-7 while Tables S7-S8 collect statistical data on inter-residue contacts. Some details of the clustering calculations and of the $S_{conform}$ calculations are given in Table S5 and Figure S3.

Table 1. Average values (in Å) of the radii of accessibility (r_{acc}) at different sites from the last 1.5 μs of the MD simulations. Standard deviations are given in parentheses. The r_{acc} data for the X-ray structure are also included.

	6LU7	M		M/Pep		D_A		D_B		D_A/2Pep		D_B/2Pep	
Cys$_{145}$ (Sγ)	3.30	2.68	(1.16)	2.63	(1.21)	2.50	(1.28)	3.60	(0.88)	3.45	(1.03)	2.99	(1.16)
His$_{41}$ (imidazol)	3.45	3.95	(1.24)	3.43	(0.78)	4.37	(0.99)	3.78	(0.84)	3.90	(0.67)	3.62	(0.71)
Cys$_{145}$ (NH)	1.28	1.21	(0.44)	1.83	(0.43)	2.44	(0.60)	1.32	(0.28)	1.22	(0.31)	1.35	(0.36)
Gly$_{143}$ (NH)	3.10	3.94	(1.46)	2.76	(0.79)	5.18	(1.38)	3.47	(0.96)	3.11	(0.82)	2.99	(0.83)
Glu$_{166}$ (NH)	3.35	2.16	(1.17)	3.48	(0.71)	3.59	(0.76)	3.34	(0.83)	2.67	(0.62)	2.86	(0.73)
Glu$_{166}$ (C=O)	4.03	4.51	(1.08)	5.59	(0.60)	5.55	(0.74)	5.30	(0.77)	5.07	(0.67)	5.00	(0.84)
His$_{164}$ (C=O)	3.28	2.77	(0.67)	3.87	(0.65)	3.24	(0.77)	3.31	(0.80)	4.20	(0.49)	3.92	(0.57)
His$_{163}$ (imidazol)	2.28	1.63	(0.68)	2.55	(0.31)	2.02	(0.56)	2.26	(0.33)	2.37	(0.17)	2.38	(0.18)
Met$_{49}$ (Side chain)	3.45	3.82	(1.14)	5.67	(0.52)	5.94	(0.26)	5.74	(0.61)	5.89	(0.34)	5.57	(0.77)
His$_{164}$ (side chain)	0.68	1.42	(0.52)	1.43	(0.28)	1.26	(0.72)	1.09	(0.21)	1.33	(0.16)	1.33	(0.21)
Met$_{165}$ (side chain)	3.35	3.51	(0.83)	3.69	(0.97)	4.12	(1.11)	3.83	(0.83)	4.52	(0.57)	3.96	(0.89)
Leu$_{167}$ (side chain)	1.90	1.98	(0.50)	2.38	(0.47)	1.74	(0.56)	2.14	(0.59)	2.47	(0.37)	2.41	(0.42)

The active site region remains quite accessible during the MD simulations as expressed in terms of the mean r_{acc} values collected in Table 1. According to this structural index, which measures the size of the largest spherical probes contacting a particular group of protein atoms, the catalytic Cys$_{145}$ thiol group, the His$_{41}$ imidazol and the Gly$_{143}$ amide group are not sterically blocked during the **M/D** simulations with r_{acc} values above 2.7, 3.7 and 3.5 Å, respectively, while the backbone Cys$_{145}$ position is partially buried with smaller r_{acc} between 1.3 and 2.0 Å. The majority of the binding spots located at the hydrophobic shallow pocket are well solvent-exposed and have r_{acc} values quite similar to those in the X-ray structures with the exception of the Glu$_{166}$ backbone and His$_{163}$ side chain in the **M** simulation (see below). Such large accessibility is observed again in the molecular surface drawings (see Figure 4), which in addition reveal that a local concentration of negative electrostatic potential is another feature of the active-site surface patch.

Although the global form of the active region is comparable during the **M** and **D** simulations, there are significant differences concerning the amplitude of the fluctuations and deviations with respect to the initial coordinates. For example, clustering analysis considering both backbone and side chain atoms results in 80 different clusters for the **M** trajectory, the top-2 clusters having moderate abundances of 16% and 14%, respectively. The same clustering settings yield 33/32 clusters for protomers A/B in the dimer, the top **D$_A$**/**D$_B$** clusters being more populated (33% for the first cluster and 14/17% for the second cluster; Table S5). When the average-linkage clustering is based on the backbone RMSD values, there are fewer clusters with higher abundances as expected, but the flexibility measurement is somehow modified. Thus, the monomeric **M** simulation still has many clusters (25), the two most populated ones accounting for 39% and 16% of the analyzed MD frames. For each protomer in the **D** simulation, the

backbone-only clustering results in 7 ($\mathbf{D_A}$) and 4 ($\mathbf{D_B}$) clusters, but the population distribution is asymmetrical given that the top-2 $\mathbf{D_A}$ clusters have 37% and 33% whereas the backbone D_B is quite rigid, the top-1 cluster having 89% abundance. In terms of the T-weighed conformational entropy values (Figure S3), the flexibility of the backbone dihedral angles in the vicinity of the active site amounts to -11, -13 and -7 kcal/mol for the $\mathbf{M}$, $\mathbf{D_A}$ and $\mathbf{D_B}$ catalytic sites, respectively, which seem in consonance with the backbone clustering analysis. Therefore, it is clear that the active site in the monomeric state exhibits a significant flexibility both in terms of backbone and side chain atoms. However, the active site region may also retain some plasticity in the dimer state as shown in the $\mathbf{D_A}$ active site.

Inspection of the cluster representatives confirms that the catalytic site in the isolated monomer (Figure 4c) is more ductile than that in protomers A and B of the dimer (Figure 5b). Moreover, a close examination of the top-2 $\mathbf{M}$ clusters (Figure 4) reveals that the oxyanion loop, the domain-II/III linker segment and the interhelical connecting loop in domain I tend all to be distorted with regards to the X-ray structure. For the oxyanion loop, a short 3_{10} helix conformation is detected in residues Asn_{142}-Ser_{144} which, in turn, is associated with the placement of the Asn_{142} side chain over the shallow S_1 subsite, blocking access to the Glu_{166} NH and CO groups. However, the time evolution of the r_{acc} indexes indicates that the S_1 "collapse" in the $\mathbf{M}$ simulation is reversible (Figure S5) given that access to the Glu_{166}@NH and His_{163} sites fluctuates on the ns time scale between blocked phases with r_{acc} around ~1.0 Å and solvent-exposed phases with r_{acc} above 2 Å. The r_{acc} plots are indeed correlated with the placement of the Asn_{142} side chain and the 3_{10} helical distortion at the oxyanion loop. According to secondary structure analyses using the DSSP method, the central residues Asn_{142}-Gly_{143}-Ser_{144} show 3_{10} helical conformation in ~43% of the analyzed snapshots, with the backbone chain interchanging

between the helical and turn/coil conformations along the **M** trajectory (Figure S6), what is in consonance with the structural flexibility of the oxyanion-loop in this state.

The conformation of the oxyanion loop during the **D** simulation deserves also particular attention due to its role in substrate binding and catalysis. A dissimilar behavior was observed between the two protomers so that the D_B active site maintains an overall conformation that is closer to the crystallographic structure and clearly more rigid than that of the other protomer D_A (Figure 5). The major difference arises in the oxyanion loop because it adopts a β-strand conformation around the Phe_{140} residue during the central part of the **D** simulation (protomer A; Figure S6) without compromising the accessibility to the important binding sites. Therefore, the secondary structure analysis further confirms the larger flexibility of the oxyanion loop in D_A and reveals its complex conformational properties, which oscillate between flexible and quasi-static states in the absence of substrate molecules. However, the implications (if any) of this behavior for substrate binding are not clear.

The simulations allow us to establish unambiguously the relationship between the oxyanion loop dynamics and specific inter-residue contacts. Among such contacts, the catalytic Cys_{145} forms a persistent (94-98%) $Cys_{145}@C=O\cdots Asn_{28}@N\delta H$ interaction in all the simulations except D_A, in which such backbone$\cdots$side chain contact is less abundant (54 %). Thus, the $Asn_{28}\cdots Cys_{145}$ interaction is involved in the orientation and flexibility of the *C*-terminal portion of the oxyanion loop (Asn_{28} seems also important because of its additional contacts with Cys_{117} and Gly_{120} at the domain I/II interface). To analyze the contacts of the Cys_{145} thiol group, we selected geometric criteria ($S_\gamma\cdots$:X distance < 4.0 Å and $90 < S_\gamma H\cdots$:X angle $< 180°$) that take into account the larger size and more diffuse electron cloud of the sulfur atom.[42] The most abundant interaction of $Cys_{145}@S\gamma H$ occurs with the backbone carbonyl group of His_{164},

especially in the monomeric state (84% for **M**, 81% for protomer A and 65% for B in **D**). When looking at the central region of the oxyanion loop, its conformation is mainly held by means of a $Ser_{144}@O\gamma H\cdots Leu_{141}@O$ H-bond so that its rupture (*e.g.* in $\mathbf{D_A}$) leads to a different loop arrangement. Considering the first part of the oxyanion loop, it turns out that the Phe_{140} side chain largely determines its conformation. This group gives several non-polar contacts with Tyr_{126}, His_{163}, His_{172} and Val_{114} both in the X-ray structures and in the majority of the simulations. However, the destabilization of this hydrophobic clustering in **M** and $\mathbf{D_A}$, particularly of the Phe_{140}(phenyl)$\cdots His_{163}$(imidazol) $\pi-\pi$ stacking, would trigger the Phe_{140} rearrangement that results in the conformational change of the whole oxyanion loop.

The network of H-bond contacts around the catalytic His_{41} residue is also of particular interest. The simulations confirm the presence of water mediated $His_{41}@NH\cdots Wat\cdots Asp_{187}@O\delta$ and/or $His_{41}@N\delta H\cdots Wat\cdots Asp_{187}@O\delta$ contacts. Furthermore, the same water molecule connects the His_{41} and His_{164} side chains ($His_{41}@N\delta H\cdots Wat\cdots His_{164}@N\delta\sigma$). We note, however, that the bridging water molecule exchanges frequently with bulk solvent on the nanosecond time scale without disrupting its structural role. The His_{41} side chain also forms non-polar weak contacts with Leu_{27}, Pro_{39}, Met_{49}, and His_{164} (see Table S7B and S8B). The salt-bridge $Arg_{40}\cdots Asp_{187}$ interaction is also observed in all the simulations and contributes to maintain in the right place the short helix containing His_{41} and the first part of the II/III connecting loop (residues 189-191).

Substrate binding

As previously mentioned, the binding of the peptide substrate within the active site exerts a certain rigidifying effect in the inter-domain dynamics, the global structure of the enzyme becoming more compact. Peptide binding can also induce specific effects on the oxyanion loop

conformation with respect to the unbound form of the enzyme. Thus, the structural analysis, clustering calculations and secondary structure assignments point out clearly that the oxyanion-loop conformation is nearly-frozen in the presence of the substrate and remains relatively close to the X-ray conformation. Hence, the MD simulations suggest that a stable positioning of the oxyanion loop partially induced by the substrate would be required for optimal binding and catalysis.
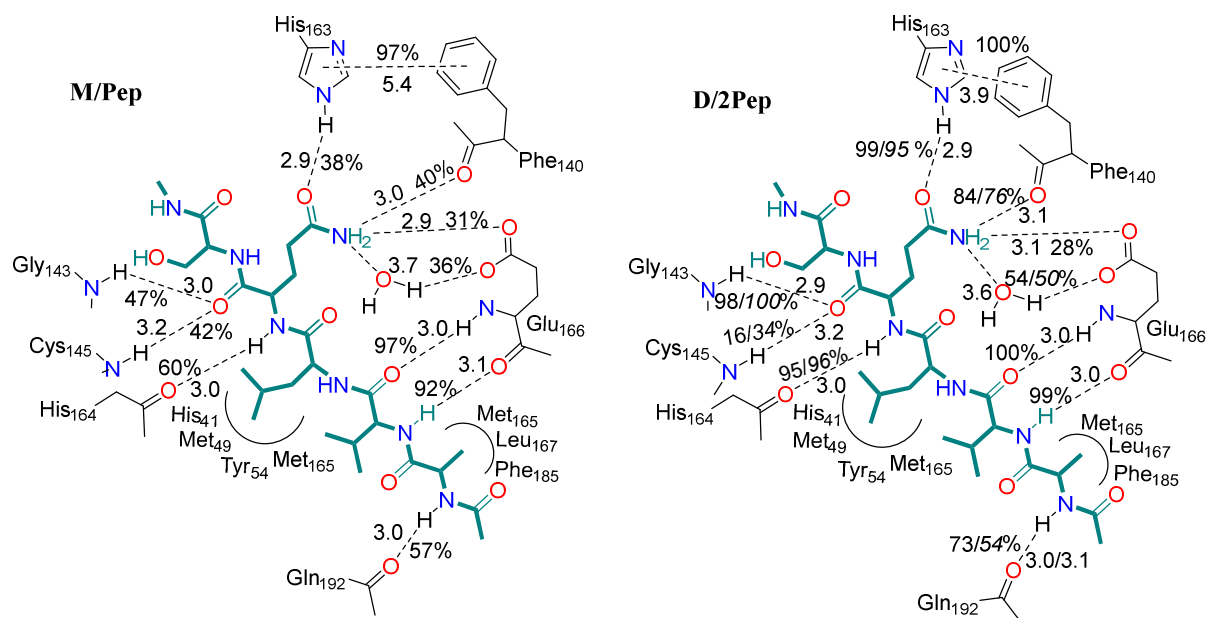
Scheme 2 Schematic representation of enzyme-substrate interactions. Average values of heavy-atom separation (Å) and % of abundances are indicated for selected contacts. Some abundances are segregated into protomer A and B (in italics).
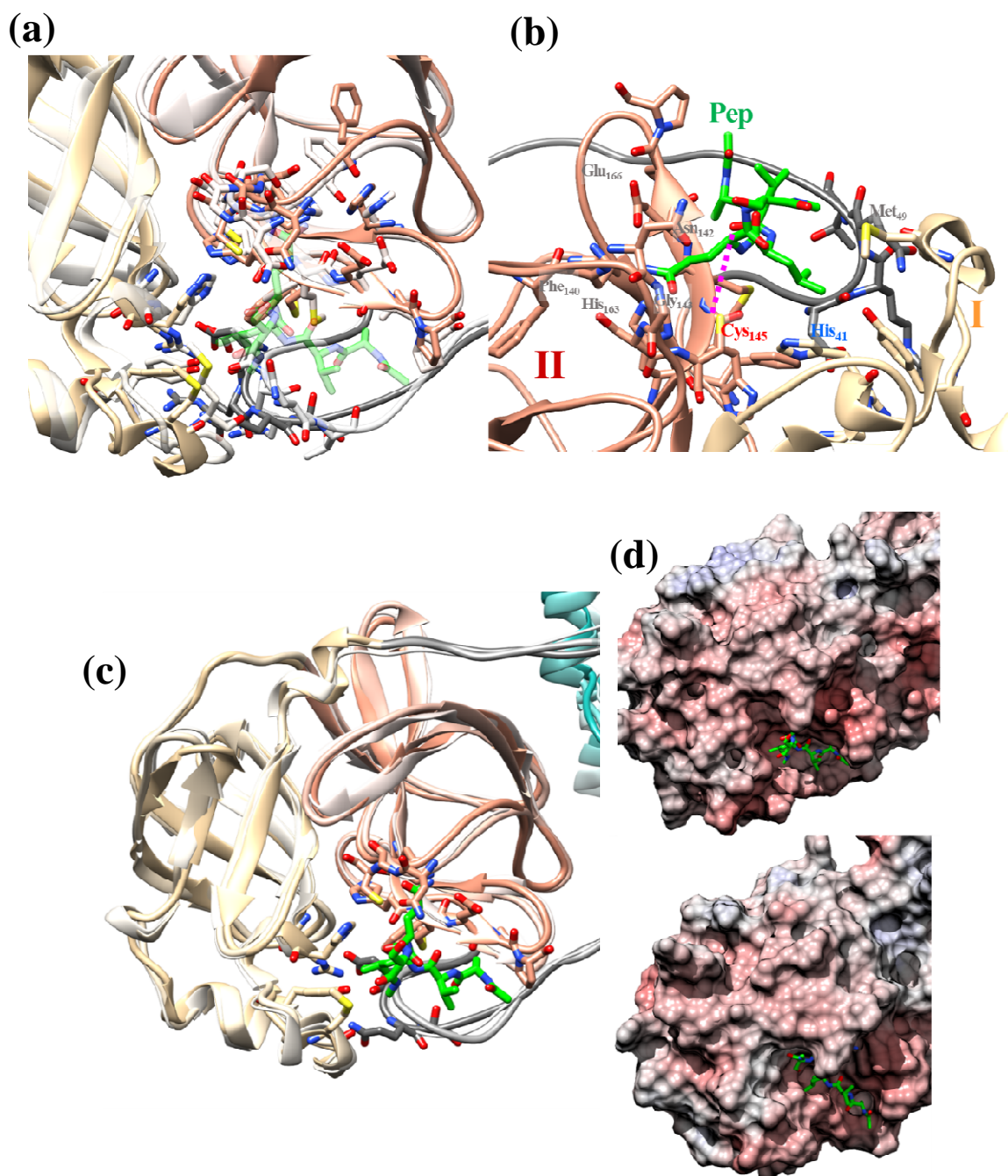
Figure 6 (a) View of the active site as shown by the superposition of the most important **M/Pep** cluster representative on the X–ray structure (shown in light colors). (b) Closer view of the active site region showing the peptide substrate and the catalytic dyad. (c) Superposition of the cluster representatives (d) Electrostatic potential mapped onto the solvent-accessible protein surfaces.
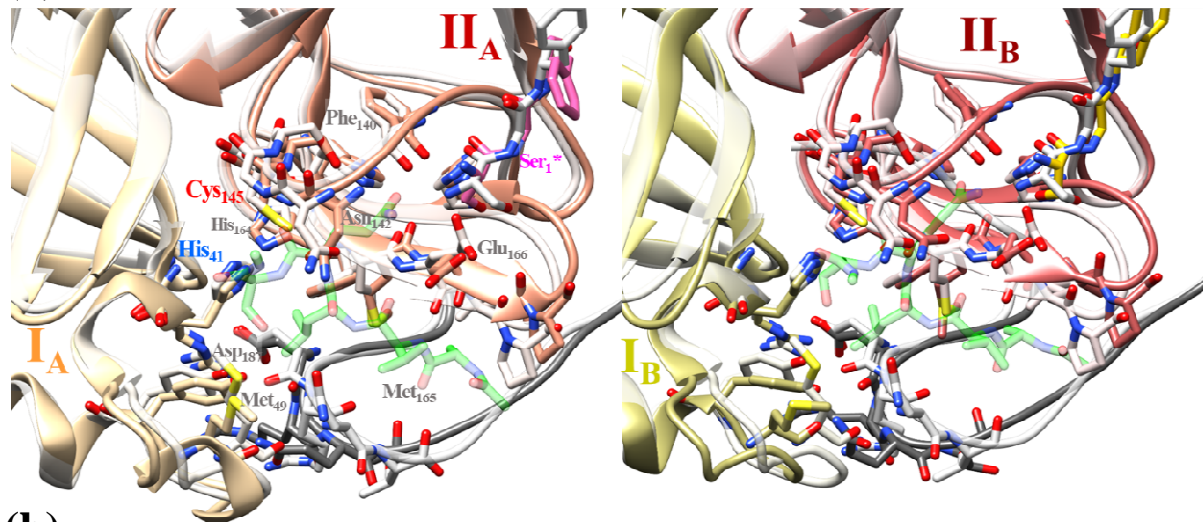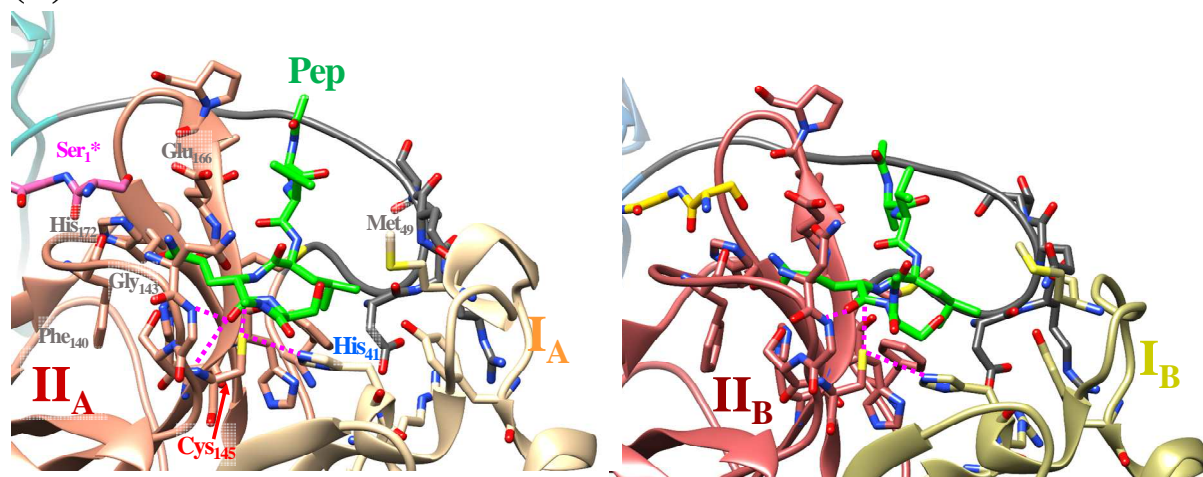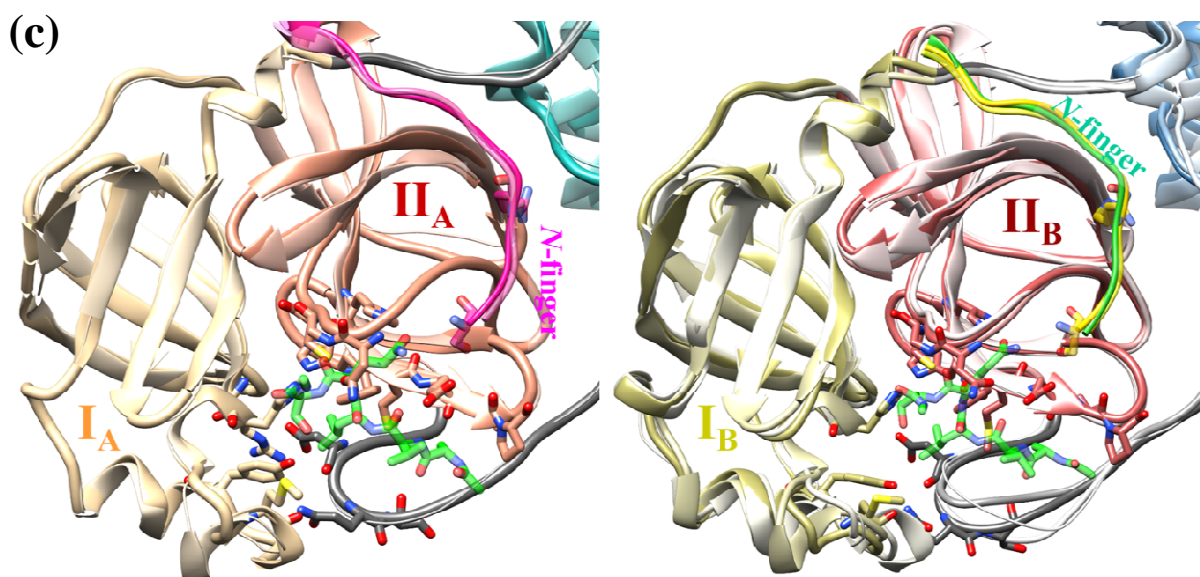
Figure 7. (a) View of the active site as shown by the superposition of the most important **D/2Pep** cluster representative on the X–ray structure (shown in light colors). (b) Closer view of the active site region showing the peptide substrate and the catalytic dyad. (c) Superposition of the cluster representatives.

Concerning the enzyme-peptide binding determinants, we note first that the short peptide Ace-Ala(P_4)-Val(P_3)-Leu(P_2)-Gln(P_1)~Ser(P_1')-Nme aligns antiparallel to the terminal part of the long II β-strand (see Scheme 2 and Figures 6-7). The analysis of the **M/Pep** and **D/2Pep** simulations shows that the alignment occurs from P_4 to P_2 thanks to two highly stable (*i.e.* 90%-100%) H-bond contacts involving the backbone groups of Glu_{166} in the β-strand and that of the substrate Val(P_3) residue. At the *N*-terminal end of the peptide, some H-bonds also occur with residue Gln_{192} in the II-III connection loop, although these contacts are less abundant (~ 60%). The cleavage sites selected by the SARS-CoV-2 main proteinase includes Leu, Phe, Val, or Met at the P_2 site, all of them well suited for being accommodated at the hydrophobic S_2 subsite of the enzyme. According to the calculated dispersion energy scorings, the Leu(P_2) side chain

placed at S_2 mainly interacts with the side chains of His_{41}, Met_{49}, and Tyr_{54} in domain I, and Met_{165} in domain II (Table S9).

Coronavirus main protease invariably recognizes peptide cleavage sites within polyproteins with a strictly conserved glutamine residue in P_1. Taking into account that the hydrolysis reaction catalyzed by this enzyme occurs at the P_1-P_1' amide bond, it is clear that the $Gln(P_1)$ residue should be correctly placed within the active site. According to our simulations, the backbone amino group of $Gln(P_1)$ interacts with the backbone carbonyl group of His_{164} (60% for **M/Pep** and ~96% for **D/2Pep**), whereas the $Gln(P_1)$ backbone carbonyl group is placed within the oxyanion hole defined by the amino groups of Gly_{143} and Cys_{145}. At this point, we note a remarkable difference between the **M/Pep** and **D/2Pep** simulations. For the monomer, both $Gly_{143}@NH\cdots Gln(P_1)@O$ and $Cys_{145}@NH\cdots Gln(P_1)@O$ H-bonds present a low abundance (*i.e.* < 50 %), which suggests that the scissile peptide bond is not well positioned within the active site. In contrast, the $Gly_{143}@NH\cdots Gln(P_1)@O$ H-bond is well maintained during the whole **D/2Pep** trajectory in the active site of protomers A and B. Some differences also arise at other contacts formed by the P_1 side chain that build up the specificity of the S_1 site. Our simulations confirm that residues Phe_{140} and His_{163} are important in defining the S_1 subsite via $His_{163}@N\varepsilon H\cdots Gln(P_1)@O\varepsilon$ and $Phe_{140}@O\cdots Gln(P_1)@N\varepsilon H$ H-bonds. However, these contacts are clearly more abundant in the two active sites of the dimer than in the monomer state, pointing out that the *N*-finger of the second protomer contributes to organize the S_1 subsite (see Scheme 2 and Table S9). Glu_{166} also participates in the binding of the $Gln(P_1)$ side chain but, according to the **M/Pep** and **D/2Pep** simulations, the $Glu_{166}@O\varepsilon\cdots Gln(P_1)@N\varepsilon H$ H-bond contact is preferentially mediated by a solvent molecule.

In the most likely hydrolysis mechanism assisted by the coronavirus main protease, the reactive events would involve the proton transfer from $Cys_{145}@S\gamma H$ to $His_{41}@N\varepsilon$ and the nucleophilic attack of the $S\gamma$ atom to the carbonyl group of $Gln(P_1)$. In this respect, the simulations show that, compared to the native form, the Cys_{145} side chain shifts to preferentially interact with the His_{41} side chain when the peptide substrate is bound within the active site. Thus, a catalytically relevant $Cys_{145}@S\gamma H \cdots N\varepsilon@His_{41}$ contact is present along the **D/2Pep** simulation (80% for A and 81% for B), although it is less abundant for **M/Pep** (49 %). The H-bond network involving His_{41}, His_{164}, and Asp_{187} and the structural water molecule is well preserved in the presence of the peptide substrate, the water exchange with bulk solvent being reduced to just one (**M/Pep** and **D$_B$/Pep**) or none (**D$_A$/Pep**) events. Hence, the relocation of the Cys_{145} thiol group and the stable positioning of His_{41} result in an average $S\gamma \cdots N\varepsilon$ distance and average $S\gamma H \cdots N\varepsilon$ angle of 3.9±0.5 Å and 95±43º for **M/Pep**, and 3.5±0.3 Å and 125±39º for **D/Pep** (equal values for protomers A and B). These values suggest that the homodimer state favors geometrically the activation of the nucleophile by His_{41}. We also measured the average $Cys_{145}@S\gamma \cdots C@Gln(P_1)$ distance / $S\gamma \cdots C \cdots O$ angle, which have values of 5.1 ± 0.8 Å / 84±32º in **M/pep** , 3.7 ± 0.6 Å / 83±13º in **D/2Pep** for protomer A, and 4.0 ± 0.7 Å / 82±11º in **D/2Pep** for protomer B. In this case, the shorter $Cys_{145}@S\gamma \cdots C@Gln(P_1)$ distances in the **D/2Pep** simulation suggest again that the dimer state exhibits an enzyme-substrate orientation favorable for catalysis.

DISCUSSION

Most of the crystallographic and biochemical studies reported to date about the 3CL^{pro} proteases have been devoted to the SARS-CoV enzyme. Although their major conclusions are reasonably expected to be valid for the SARS-CoV-2 protease due to their high degree of homology, there is, of course, a growing research activity aimed specifically at SARS-CoV-2. Nevertheless, detailed molecular descriptions of the structure and dynamics of the native 3CL^{pro} enzymes in aqueous solution and of its binding determinants are still lacking, what prompted us (and other researchers) to computationally examine the SARS-CoV-2 protein by means of extensive MD simulations. In this scenario, our results may complement currently-available structural data given that several protein configurations were simulated and subject to comparative analysis in order to clarify some questions regarding the monomer structure, the dimer stability, and so on. It must be noted, however, that the significance of the present 2.0 μs MD simulations will depend on their critical comparison with results produced by other theoretical studies.

With regard to the tertiary structure of the monomeric 3CL^{pro} protein, a remarkable observation is the domain III rearrangement occurring during the **M** simulation. Thus, the monomer enzyme can adopt alternative conformations in aqueous solution. Interestingly, the crystallographic structures of three mutants of the SARS-CoV protein (G11A, S139A, and R298A) reveal a similar domain II /domain III pose.[13] These mutations cause the dimer dissociation by disrupting critical contacts at the protomer interface and, thereby, the rotation of domain III has been considered a structural feature of the monomeric enzyme. In the most abundant domain III orientation observed in the **M** simulation, the alignment of domains II and III is stabilized by a few salt bridges that also involve *N*-finger residues (Arg$_4$···Asp$_{216}$, Lys$_5$

$\cdots$Glu$_{290}$ and Arg$_{131}\cdots$Asp$_{289}$). In the SARS-CoV mutants (2PWX structure),[13] the Lys$_5$ $\cdots$Glu$_{290}$ salt bridge is again observed while the domain III is slightly rotated and establishes different salt bridges (Arg$_{131}\cdots$Glu$_{240}$ and Lys$_{137}\cdots$Asp$_{289}$). The simulations also gave some clues about the domain III rotation event. At the beginning of the monomer MD simulation, the *N*-finger, which lacks the Arg$_4\cdots$Glu$_{290}$* contact characteristic of the dimer state, is anchored at the last α-helix of domain III both by polar (Arg$_4\cdots$Gln$_{299}$) and non-polar (Phe$_3\cdots$Phe$_{291}$) interactions. It turned out that the weakening of the $\pi-\pi$ contact preceded the rupture of the Arg$_4\cdots$Gln$_{299}$ contact, loosening then the *N*-finger hold onto domain III and enabling the torsional changes at the hinge Gly$_{195}$ residue for the wide rotation to occur. Hence, it seems reasonable to consider that the *N*-finger structural instability induces the domain III conformational change. We also note in passing that the domain III rearrangement may be especially unfavorable for the recognition and organization of the long polyprotein sequences processed *in vivo* by the main SARS-CoV proteases.

The MD simulations probe possible effects associated with the binding of the relatively small pentapeptide substrate. In general, we found that the amplitude of intra- and interdomain motions as well as the global flexibility are reduced in the presence of the substrate as measured in terms of various descriptors (RMSDs, molecular surface, cluster representatives, etc.). The dampening of the domain I and II mobility is well understood because the peptide molecule, which remains perfectly bound to the catalytic site, further interconnects domains I and II by forming a stable network of H-bonds and non-polar contacts. Similarly, the direct influence of the Gln(P_1) residue on the oxyanion loop conformation is equally clear. However, the allosteric mechanism(s) through which substrate binding can affect the conformation of distal enzyme regions and/or enhance dimerization are not obvious. This is apparently the case for the outer

wall of a channel[43] passing through the central region of the dimer constituted by the side chains of Ala_{285} and Leu_{286} from each of the protomers. The inter-residue contact analysis and the surface calculations indicate that this hydrophobic cluster is significantly more compact in the **D/2Pep** simulation than in **D**, what is likely connected with the lower mobility of domain III upon substrate binding. On the other hand, the fact that the domain III rotation did not occur during the monomeric **M/Pep** simulation, confirms again the rigidifying role played by the peptide molecule. More specifically, the residues of the domain II/III linker loop that border the catalytic site ($-Gln_{189}-Thr_{190}-Ala_{191}-Gln_{192}-$) become less flexible in the presence of the peptide molecule, what, in turn, may hamper the torsional change at the nearby Gly_{195} acting as the hinge residue for the domain III rearrangement. Therefore, in this way, substrate binding to monomeric proteases could favor the inter-domain orientation that is adequate for dimerization, helping thus clarify the cooperative effects between substrate binding and dimerization experimentally observed[16] in the monomeric mutants of SARS-CoV.

Clearly, the small impact on dimer stability and catalysis due to the truncation of residues 1-3 in the *N*-finger of SARS-CoV[9] agrees nicely with our MD simulations in aqueous solution, which reveal how the contacts involving the *N*-terminal Ser_1 in the starting crystallographic structures are largely weakened by water molecules (*e.g.*, $Ser_1 \cdots Glu_{166}$ salt bridge and $Ser_1 \cdots Phe_{140}$ backbone contact). Hence, we propose that Ser_1 plays only a secondary role in controlling the conformation of the oxyanion loop, which contrasts with the critically important Arg_4 residue. Experimentally, residue deletion up to Arg_4 results in a monomeric SARS-CoV protein with null or very weak enzymatic activity.[9] In fact, our simulations emphasize that Arg_4 has a twofold role. On one hand, it provides the $Arg_4 \cdots Glu_{290}*$ salt-bridge between the two protomers and helps fixing the position of domain III in the same protomer through a double H-

bond between the *N*-finger backbone and the Gln_{299} side chain. On the other hand, its guanidinium group contributes to clutch the oxyanion loop by H-bonding the Lys_{137}* amide group in the other protomer. However, in the native form of the enzyme, the latter interaction and the $Ser_{139} \cdots Gln_{299}$* contact with the C-terminal helix may be perturbed as seen in the **D** simulation, conferring some flexibility to the oxyanion loop. Other residues located in the *N*-coil also contribute to dimer stability, the $Gly_{11}@NH \cdots O\varepsilon@Glu_{14}$* contacts being especially important in terms of their abundance (100%) and steady interatomic distance (2.9 Å) all along the dimer simulations, what is in consonance with the inability of the $Gly_{11}Ala$ SARS-CoV mutant to dimerize.[13]

As mentioned in the Introduction, the absence of catalytic activity in the monomeric SARS-CoV mutants has been ascribed to a presumably collapsed form of the active site, in which a partially-3_{10} helical oxyanion loop would impede access to the binding sites. To our knowledge, such collapsed conformation of residues Ser_{139}-Phe_{140}-Leu_{141} has been observed both in the monomeric X-ray structures of the SARS-CoV enzyme and in some of its dimer states. In our simulations, the 3_{10} helical twist occurs exclusively during the monomeric **M** simulation at the Asn_{142}-Ser_{144} residues adjacent to Cys_{145}. Furthermore, this transition turns out to be *reversible* (twist and untwist events are observed) and only reduces the accessibility to the $Glu_{166}@NH$ group and His_{163} side chain. Therefore, the simulations suggest that, in the liquid phase, the active site of the monomeric state would be transiently, but not permanently, in the partially-collapsed conformation. The MD analysis also shows that the S_1 structural fluctuations associated with the oxyanion loop motions would be the direct consequence of the missing contacts with the *N*-finger of the protomer. Concerning the dynamics of the dimer unbound state, we find that all the binding sites remain perfectly solvent-exposed and the active site regions

tend to be more rigid than in the monomer form. The terminal residues (1-3) of the *N*-finger still have some conformational freedom and the oxyanion-loop in $\mathbf{D_A}$ shows flexibility around the Phe$_{140}$ position. Hence, we propose that the S_1 site of the 3CL^{pro} enzymes may retain some plasticity in the dimer form, which would be more accentuated in the monomer state reducing, but not abolishing, its peptide binding ability. On the contrary, we expect that the monomer SARS-CoV-2 protein would have a significant affinity for binding the examined peptide as indicated by some highly stable enzyme-substrate contacts along the **M/Pep** MD trajectory. Therefore, although the **M/Pep** complex may be not optimal for catalysis (see below) and the **M** active site exhibits varying oxyanion loop conformations, our simulations give no support to the hypothesis of a collapsed active site to explain the non-catalytic behavior of the monomer state. In fact, they seem more compatible with the report of calorimetric K_d values for the enzyme-peptide complexes involving monomeric mutants of SARS-CoV that are close to that of the wild-type dimer.[16]

The present MD simulations allow us to investigate the noncovalent binding between the SARS-CoV-2 enzyme and a model peptide, Ace-Ala-Val-Leu-Gln~Ser-Nme, reproducing the sequence of the first cleavage site catalyzed by 3CL^{pro} at the polyproteins. The construction of the initial complex was feasible thanks to the crystallographic adduct (6LU7) between SARS-CoV-2 and a peptido-mimetic inhibitor. The remarkable stability of the enzyme-peptide mode of binding all along the (unconstrained) simulations confirms that the active site is readily adapted to accommodate this sequence. As above discussed, the presence of the peptide molecule influences both globally and locally the structure and dynamics of the systems. In this respect, the alignment of the substrate chain along the binding crevice together with the accommodation of Gln(P_1) and Leu(P_2) in the S_1 and S_2 pockets, firmly locks the scissile peptide bond into the

Cys$_{145}$/Gly$_{143}$ oxyanion hole. Furthermore, we found that substrate is essential for the side chain of the nucleophilic Cys$_{145}$ to adopt a suitable orientation for proton transfer towards the His$_{41}$ imidazol and nucleophilic attack towards the Gln(P_1) carbonyl group. Therefore, we note that the catalytic competency of the enzyme should not be judged on the basis of the Cys$_{145}$···His$_{41}$ contacts in the native form of the enzyme. The average interatomic distances and angles relating Cys$_{145}$@Sγ, Gln(P_1)@C and His$_{41}$@Nε corroborate the nearly pre-reactive character of the Michaelis complexes in the **D/2Pep** simulation (*e.g.*, Sγ···C ~3.7-4.0 Å; Sγ···Nε ~3.5 Å). For the monomeric Michaelis complex, however, the measured pre-reactive contacts are much less favorable (*e.g.*, Sγ···C ~5.1 Å; Sγ···Nε ~3.9 Å).

Finally, the question arises about why the monomeric SARS-CoV-2 and SARS-CoV proteases have minimal or null activity against short peptide molecules. In this respect, our results reveal significant differences between the **M/Pep** and **D/2Pep** complexes that would undoubtedly favor the catalytic efficiency in the dimer state. On one hand, the larger flexibility of the monomeric active site could result in a certain entropic penalty as peptide binding would stifle the oxyanion-loop and the domain II/III linker residues delimiting the binding sites. On the other hand, the relative position between the nucleophile and the Ser(P_1')~Gln(P_1) peptide bond seems not adequate for catalysis in the monomeric Michaelis complex. This is most likely due to the above discussed changes in the S_1 site regarding the oxyanion loop position and the lack of the *N*-finger residues. In addition, other possible effects could positively affect catalysis in the dimer complexes. For example, the proximity of the whole domain III of the other protomer to the active site region could well provide a favorable environment for the electrostatic stabilization of transition states and reactive intermediates. More conclusive evidence about the catalytic

impairment of the monomeric proteases could be gained by means of further computational work aimed to determine the full catalytic pathway and free energy profiles.

SUPPORTING INFORMATION

Table S1: Electrostatic pK_a calculations. Table S2: Summary of the sampled systems and computational settings. Table S3: Average RMSD values. Table S4: LCPO molecular surface calculations. Table S5: Details of average-linkage clustering calculations. Table S6: Statistics of protomer⋅⋅⋅protomer contacts. Tables S7-8: Statistics of interesidue contacts in the active site. Table S9: Statistics of enzyme⋅⋅⋅peptide contacts. Figure S1: Dial plots of the Euler rotational angles of the domains I/III. Figure S2: time evolution of interdomain COM distances. Figure S3: Convergence plots of the first-order conformational entropy. Figure S4: Ribbon representations of average structure for the monomeric simulation started at 6Y84. Figure S5: Time evolution of selected r_{acc} (Å) values. Figure S6: Secondary structure of the oxyanion loop. A zip package containing the PDB files (solute atoms) of the most populated cluster representatives shown in Figures 4-7.

REFERENCES

1.   Zhu, N.; Zhang, D.; Wang, W.; Li, X.; Yang, B.; Song, J.; Zhao, X.; Huang, B.; Shi, W.; Lu, R.; Niu, P.; Zhan, F.; Ma, X.; Wang, D.; Xu, W.; Wu, G.; Gao, G. F.; Tan, W., A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N. Engl. J. Med.* **2020**, 382, 727-733.

2.   Gorbalenya, A. E.; Baker, S. C.; Baric, R. S.; de Groot, R. J.; Drosten, C.; Gulyaeva, A. A.; Haagmans, B. I.; Lauber, C.; Leontovich, A. M.; Neuman, B. W.; Penzar, D.; Perlman, S.; Poon, L. L. M.; Samborskiy, D. V.; Sidorov, I.; Sola, I.; Ziebuhr, J., The Species Severe Acute Respiratory Syndrome-Related Coronavirus: Classifying 2019-nCoV and Naming it SARS-CoV-2. *Nat. Microbiol.* **2020**, 5, 536-554.

3.   Ziebuhr, J., The Coronavirus Replicase. *Curr. Top. Microbiol. Immnunol.* **2005**, 287, 57-94.

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

5.   Jin, Z.; Du, X.; Xu, Y.; Deng, Y.; Liu, M.; Zhao, Y.; Zhang, B.; Li, X.; Zhang, L.; Peng, C.; Duan, Y.; Yu, J.; Wang, L.; Yang, K.; Liu, F.; Jiang, R.; Yang, X.; You, T.; Liu, X.; Yang, X.; Bai, F.; Liu, H.; Liu, X.; Guddat, L. W.; Xu, W.; Xiao, G.; Qin, C.; Shi, Z.; Jiang, H.; Rao, Z.; Yang, H., Structure of M^{pro} from COVID-19 Virus and Discovery of its Inhibitors. *Nature* **2020**, 582, 289-293.

6.   Hilgenfeld, R., From SARS to MERS: Crystallographic Studies on Coronaviral Proteases Enable Antiviral Drug Design. *FEBS J.* **2014**, 281, 4085-4096.

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

8.   Shi, J.; Wei, Z.; Song, J., Dissection Study on the Severe Acute Respiratory Syndrome 3C-Like Protease Reveals the Critical Role of the Extra Domain in Dimerization of the Enzyme. *J. Biol. Chem.* **2004**, 279, 24765-24773.

9.   Hsu, W.-C.; Chang, H.-C.; Chou, C.-Y.; Tsai, P.-J.; Lin, P.-I.; Chang, G.-G., Critical Assessment of Important Regions in the Subunit Association and Catalytic Action of the Severe Acute Respiratory Syndrome Coronavirus Main Protease. *J. Biol. Chem.* **2005**, 280, 22741-22748.

10.   Lin, P.-Y.; Chou, C.-Y.; Chang, H.-C.; Hsu, W.-C.; Chang, G., -G., Correlation between Dissociation and Catalysis of SARS-CoV Main Protease. *Arch. Biochem. Biophys.* **2008**, 472, 34-42.

11.   Chen, S.; Chen, L.; Tan, J.; Chen, J.; Du, L.; Sun, T.; Shen, J.; Chen, K.; Jiang, H.; Shen, X., 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. *J. Biol. Chem.* **2005**, 280, 164-73.

12.   Yang, H.; Yang, M.; Ding, Y.; Liu, Y.; Lou, Z.; Zhou, Z.; Sun, L.; Mo, L.; Ye, S.; Pang, H.; Gao, G. F.; Anand, K.; Bartlam, M.; Hilgenfeld, R.; Rao, Z., The Crystal Structures of Severe Acute Respiratory Syndrome Virus Main Protease and its Complex with an Inhibitor. *Proc. Nat. Acad. Sci. USA* **2003**, 100, 13190-13195.

13.   Chen, S.; Hu, T.; Zhang, J.; Chen, J.; Chen, K.; Ding, J.; Jiang, H.; Shen, X., Mutation of Gly-11 on the Dimer Interface Results in the Complete Crystallographic Dimer Dissociation of Severe Acute Respiratory Syndrome Coronavirus 3C-Like Protease. *J. Biol. Chem.* **2008**, 283, 554-564.

14.   Shi, J.; Sivaraman, J.; Song, J., Mechanism for Controlling the Dimer-Monomer Switch and Coupling Dimerization to Catalysis of the Severe Acute Respiratory Syndrome Coronavirus 3C-Like Protease. *J. Virol.* **2008**, 82, 4620–4629.

15.   Hu, T.; Zhang, Y.; Li, L.; Wang, K.; Chen , S.; Chen, J.; Ding, J.; Jiang, H.; Shen, X., Two Adjacent Mutations on the Dimer Interface of SARS Coronavirus 3C-Like Protease Cause Different Conformational Changes in Crystal Structure. *Virology* **2009**, 388, 324-334.

16.   Cheng, S.-C.; Chang, G.-G.; Chou, C.-Y., Mutation of Glu-166 Blocks the Substrate-Induced Dimerization of SARS Coronavirus Main Protease. *Biophys. J.* **2010**, 98, 1327-1336.

17.   Shi, J.; Han, N.; Lim, L.; Lua, S.; Sivaraman, J.; Wang, L.; Mu, Y.; Song, J., Dynamically-Driven Inactivation of the Catalytic Machinery of the SARS 3C-Like Protease by the N214A Mutation on the Extra Domain. *PLoS Comput. Biol.* **2011**, 7, e1001084.

18.   Anandakrishnan, R.; Aguilar, B.; Onufriev, A. V., H++3.0: Automating pK Prediction and the Preparation of Biomolecular Structures for Atomistic Molecular Modeling and Simulations. *Nucleic Acids Res.* **2012**, 40, W537-W541.

19.   Owen, C. D., Lukacik, P., Strain-Damerell, C.M., Douangamath, A., Powell, A.J., Fearon, D., Brandao-Neto, J., Crawshaw, A.D., Aragao, D., Williams, M., Flaig, R., Hall, D., McAauley, K., Stuart, D.I., von Delft, F., Walsh, M.A., COVID-19 Main Protease with Unliganded Active Site. **2020**. DOI: 10.2210/pdb6y84/pdb.

20.   Zhang, B.; Zhao, Y.; Jin, Z.; Liu, X.; Yang, H.; Rao, Z., The Crystal Structure of COVID-19 Main Protease in Apo Form. **2020**. DOI: 10.2210/pdb6m03/pdb.

21.   Fearon, D., Powell, A.J., Douangamath, A., Owen, C.D., Wild, C., Krojer, T., Lukacik, P., Strain-Damerell, C.M., Walsh, M.A., von Delft, F, PanDDA analysis group deposition -- Crystal Structure of COVID-19 main protease in complex with Z219104216. **2020**. DOI: 10.2210/pdb5r82/pdb.

22.   Consortium, T. U., UniProt: A Worldwide Hub of Protein Knowledge. *Nucleic Acids Res.* **2019**, 47, D506-D515.

23.   Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E., UCSF Chimera-A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, 25, 1605-1612.

24.   D.A. Case, I. Y. B.-S., S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, D. Ghoreishi, M.K. Gilson, H. Gohlke, A.W. Goetz, D. Greene, R Harris, N. Homeyer, S. Izadi, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D.J. Mermelstein, K.M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, J. Smith, R. Salomon-Ferrer, J. Swails, R.C. Walker, J. Wang, H. Wei, R.M. Wolf, X. Wu, L. Xiao, D.M. York, P.A. Kollman *AMBER 2018*, University of California: San Francisco, 2018.

25.   Salomon-Ferrer, R.; Case, D. A.; Walker, R. C., An Overview of the Amber Biomolecular Simulation Package. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2013**, 3, 198-210.

26.   Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C., ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, 11, 3696-3713.

27.   Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, 117, 5179-5197.

28.   Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, 79, 926-935.

29.   Joung, I. S.; Cheatham, T. E., Molecular Dynamics Simulations of the Dynamic and Energetic Properties of Alkali and Halide Ions Using Water-Model-Specific Ion Parameters. *J. Phys. Chem. B* **2009**, 113, 13279-13290.

30.   Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C., Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* **1977**, 23, 327-341.

31.   Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G., A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, 103, 8577–8593.

32.   Le Grand, S.; Götz, A. W.; Walker, R. C., SPFP: Speed without Compromise—A Mixed Precision Model for GPU Accelerated Molecular Dynamics Simulations. *Comput. Phys. Commun.* **2013**, 184, 374-380.

33.   Roe, D. R.; Cheatham, T. E., PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, 9, 3084-3095.

34.   Weiser, J.; Shenkin, P. S.; Still, W. C., Approximate Atomic Surfaces from Linear Combinations of Pairwise Overlaps (LCPO). *J. Comput. Chem.* **1999**, 20, 217-230.

35.   Kabsch, W.; Sander, C., Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers* **1983**, 22, 2577-637.

36.   Elstner, M.; Hobza, P.; Frauenheim, T.; Suhai, S.; Kaxiras, E., Hydrogen Bonding and Stacking Interactions of Nucleic Acid Base Pairs: A Density-Functional-Theory Based Treatment. *J. Chem. Phys.* **2001**, 114, 5149-5155.

37.   Suárez, D.; Field, M. J., Molecular Dynamics Simulations of Human Butyrylcholinesterase. *Proteins: Struct., Funct., Bioinf.* **2005**, 59, 104-117.

38.   Sanner, M. F.; Olson, A. J.; Spehner, J.-C., Reduced Surface: An Efficient Way to Compute Molecular Surfaces. *Biopolymers* **1996**, 38, 305-320.

39.   Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A., Electrostatics of Nanosystems: Application to Microtubules and the Ribosome. *Proc. Natl. Acad. Sci. USA* **2001**, 98, 10037-10041.

40.   Suárez, E.; Díaz, N.; Méndez, J.; Suárez, D., CENCALC: A Computational Tool for Conformational Entropy Calculations from Molecular Simulations. *J. Comput. Chem.* **2013**, 34, 2041-2054.

41.   Suárez, D.; Díaz, N., Sampling Assessment for Molecular Simulations Using Conformational Entropy Calculations. *J. Chem. Theory Comput.* **2014**, 10, 4718-4729.

42.   Zhou, P.; Tian, F.; Lv, F.; Shang, Z., Geometric Characteristics of Hydrogen Bonds Involving Sulfur Atoms in Proteins. *Proteins: Struct., Funct., Bioinf.* **2009**, 76, 151-163.

43.   Shi, J.; Song, J., The Catalysis of the SARS 3C-Like Protease Is under Extensive Regulation by its Extra Domain. *Febs J.* **2006**, 273, 1035-45.

For Table of Contents Use Only

SARS-CoV-2 Main Protease: A Molecular Dynamics Study

Dimas Suárez and Natalia Díaz