

# Unraveling the SARS-CoV-2 Main Protease Mechanism Using Multiscale Methods

Carlos A. Ramos-Guzmán, J. Javier Ruiz-Pernía,\* and Iñaki Tuñón\*

Cite This: *ACS Catal.* 2020, 10, 12544–12554

Read Online

ACCESS |



Metrics &amp; More



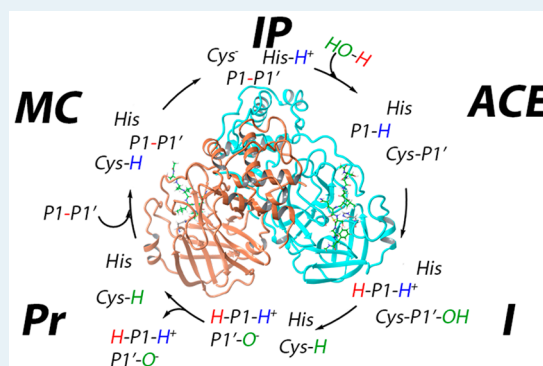
Article Recommendations



Supporting Information

**ABSTRACT:** We present a detailed theoretical analysis of the reaction mechanism of proteolysis catalyzed by the main protease of SARS-CoV-2. Using multiscale simulation methods, we have characterized the interactions established by a peptidic substrate in the active site, and then we have explored the free energy landscape associated with the acylation and deacylation steps of the proteolysis reaction, characterizing the transition states of the process. Our mechanistic proposals can explain most of the experimental observations made on the highly similar ortholog protease of SARS-CoV. We point to some key interactions that may facilitate the acylation process and thus can be crucial in the design of more specific and efficient inhibitors of the main protease activity. In particular, from our results, the P1' residue can be a key factor to improve the thermodynamics and kinetics of the inhibition process.

**KEYWORDS:** 3CL<sup>pro</sup> protease, SARS-CoV-2, minimum free energy path, QM/MM, acylation, deacylation, molecular dynamics



## 1. INTRODUCTION

The recent outbreak of COVID-19, a pneumonia-like illness caused by a coronavirus named SARS-CoV-2,<sup>1</sup> has rapidly evolved into a pandemic as recognized by the World Health Organization. SARS-CoV-2 has been shown to be highly contagious, causing a large number of infections around the world. The absence of vaccines and specific treatments has contributed to a rapid spread of this disease and a fatal outcome in many cases. Furthermore, the existence of other similar viruses detected in animals opens the possibility of future similar diseases.<sup>2,3</sup> Thus, finding effective strategies for the identification of potential targets for new drugs to fight against SARS-CoV-2 and other CoV-like viruses is an urgent need. One of these strategies is based on the disruption of the activity of those enzymes that are crucial in the replication cycle of the virus using adequate compounds. In this sense, knowledge of the catalytic activity of the enzyme in atomistic detail is one of the more powerful tools for efficient and specific new drugs design. In particular, the characterization and analysis of the geometry and electronic properties of the reaction transition state (TS) can be used as a guide for the design of active site inhibitors. In this work we analyze the reaction mechanisms for the main protease of SARS-CoV-2, also referred to as 3C-like protease (3CL<sup>pro</sup>), using density functional theory (DFT)-based multiscale methods. This enzyme plays an essential role during the replication of the virus and has no closely related homologues in human beings, making it an attractive drug target.<sup>4</sup>

The 3CL<sup>pro</sup> enzyme of SARS-CoV-2 exists as a functional homodimer with two active sites in charge of cleaving the translated polyproteins into individual fragments to be used by the coronavirus.<sup>5</sup> As with other cysteine proteases, each of the active sites contains a Cys-His catalytic dyad in charge of the hydrolysis of the peptide bond at specific sites of a polypeptide chain. Several structures of this protease have been already resolved by means of X-ray crystallography and deposited in the Protein Data Bank (PDB), including the free protease (PDB codes 6Y2E<sup>2</sup> and 6Y84<sup>6</sup>) and inhibitor-bound proteases (PDB codes 6LU7,<sup>6</sup> 6Y2F,<sup>7</sup> and 6LZE<sup>8</sup>). The SARS-CoV-2 main protease has a structure that is virtually identical to the ortholog from SARS-CoV (96% identity). Even more, the main residues involved in catalysis, binding, and dimerization processes are fully conserved.<sup>9</sup> Consequently, these two ortholog enzymes display highly similar substrate preferences.<sup>10</sup> The substrate cleavage by the 3CL<sup>pro</sup> takes place between Gln at the P1 position and a Gly/Ala/Ser at the P1' position (P and P' identify the residues placed before and after the scissile bond, respectively), making the presence of Gln an essential requirement.<sup>11</sup>

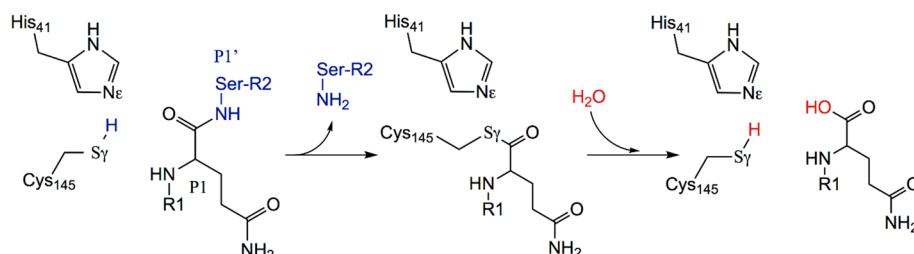
Received: August 5, 2020

Revised: September 26, 2020

Published: September 28, 2020



## Scheme 1. Reaction Mechanism in Cysteine Proteases



In principle, the reaction mechanism of cysteine proteases involves two basic steps (see Scheme 1).<sup>12</sup> In the first step, acylation, the peptide bond is broken, releasing the P' fragment of the peptidic substrate and forming an acyl-enzyme complex where the catalytic cysteine (Cys145 in the protease of SARS-CoV-2) is covalently bound to the carbon atom of the P1 residue of the target peptide. In a second step, deacylation, the acyl-enzyme is hydrolyzed, releasing the P fragment and recovering the enzymatic active site for another catalytic cycle. Covalent inhibitors of the protease activity form acyl-enzyme complexes that cannot be hydrolyzed and thus remain bonded into the active site.<sup>6,13</sup> In this work we take advantage of the similarities between the proteases of SARS-CoV and SARS-CoV-2 viruses and the existence of ligand-bound structures to build a structural model of a peptide substrate-enzyme Michaelis complex and to study the reaction mechanism using computational simulations.

## 2. METHODS

**2.1. Classical Molecular Dynamics Simulations.** The crystal structures with PDB codes 6Y2F<sup>7</sup> and 3AW0<sup>14</sup> were used as starting points to build the Michaelis complex. The former corresponds to the holoprotease of SARS-CoV-2, and the latter is the crystallographic structure of the SARS-CoV ortholog cocrystallized with the peptidic aldehyde inhibitor Ac-Ser-Ala-Val-Leu-His-aldehyde. To build the Michaelis complex corresponding to the 3CL<sup>pro</sup> protease of SARS-CoV-2, the two protein structures were aligned and then the cocrystallized ligand in 6Y2F was replaced with the crystallized ligand in 3AW0 in the two active sites of the homodimer (protomers A and B). The peptidomimetic Ac-Ser-Ala-Val-Leu-His-aldehyde inhibitor was elongated using the Maestro tool<sup>15</sup> until we built the substrate-like sequence Ac-Ser-Ala-Val-Leu-Gln-Ser-Gly-Phe-NMe. We placed the substrate in the two active sites, in agreement with the observation that the X-ray structures (6LZE, 6M0K, 6WNP, 6LU7, and 7BQY) contain inhibitor molecules in both of them. The absent hydrogen atoms were added using the Protein Preparation Wizard tool of Maestro, and PROPKA3.0<sup>16</sup> was used to calculate the protonation states of titratable residues at pH 7.4.

The tleap tool from AmberTools18<sup>17</sup> was used to prepare the simulation systems. The Michaelis complexes, described with the ff14SB force field,<sup>18</sup> were solvated into a box with a buffer region of at least 12 Å from any protein/substrate atom to the limits of the simulation box. TIP3P water molecules<sup>19</sup> were used. Na<sup>+</sup> atoms were added to neutralize the charge. The resulting system was minimized using 500 steps of the steepest descent method followed by the conjugate-gradient method, until the root-mean-square of the gradient was below 10<sup>-3</sup> kcal mol<sup>-1</sup> Å<sup>-1</sup>. The system was then heated from 0 to 300 K using a heating rate of 1.7 K-ps<sup>-1</sup>. The backbone heavy atoms were

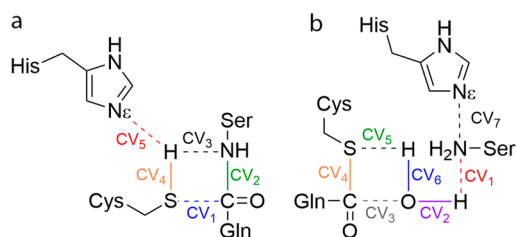
restrained in a Cartesian space using a harmonic potential with a force constant of 20 kcal mol<sup>-1</sup> Å<sup>-2</sup>. Along the equilibration step, the positional restraint force constant was changed from 15 to 3 kcal mol<sup>-1</sup> Å<sup>-2</sup>, decreasing by 3 units every 1.25 ns, and after 6.25 ns the positional restraints were removed and the systems continued their equilibration until 7.5 ns of NPT (300 K and 1 bar) simulation was completed. Then, 8 μs of NVT simulation at 300 K was performed with a 2 fs time step using SHAKE.<sup>20</sup> The particle mesh Ewald method was employed to describe the long-range electrostatic interactions;<sup>21,22</sup> for the short-range interactions, a cutoff of 10 Å was used. The pressure was controlled by a Berendsen barostat, and the temperature was controlled by a Langevin thermostat. For all the simulations, periodic boundary conditions were employed. The AMBER19 GPU version of pmemd<sup>23,24</sup> was used to run the classical molecular dynamic simulations. To sample a reasonable configurational space of the protein, two additional replicas of the Michaelis complex model with different initial velocities were run during 1 μs.

**2.2. QM/MM Calculations.** Exploration of the free energy surfaces associated with the peptide bond-breaking process have been carried out using quantum mechanics/molecular mechanics (QM/MM) simulations. In most simulations (see Table S1 for details), the side chains of the catalytic dyad (Cys145 and His41) and a fragment of the peptide substrate of protomer A were included in the QM region, while the rest of the system was described at the MM level as explained earlier. The interaction of Asp187 with His41 is mediated by a water molecule, as observed in the X-ray structures of the inhibited enzyme (6LZE, 6M0K, 6WNP, 6LU7, and 7BQY) and the apo form (1UJ1). The presence of this water molecule in between His41 and Asp187 limits possible polarization and charge-transfer effects, and the latter residue was not included in the QM region. The part of the substrate described at the QM level includes the two residues involved in the peptide bond to be broken (Gln-P1 and Ser-P1') and the previous and next peptide bonds up to the C<sup>α</sup> atoms of Leu-P2 and Gly-P2'. In the exploration of the hydrolysis of the acyl-enzyme, we also included a water molecule in the QM region. To describe the QM subsystem, we used the B3LYP functional<sup>25,26</sup> and D3 dispersion corrections.<sup>27</sup> Calculations were performed with the 6-31+G\* basis set unless otherwise indicated (we also employed the 6-31G\* basis set). This level of theory has been shown to be one of the best combinations to describe enzymatic reactions.<sup>28</sup> In addition, this computational description for the QM region (B3LYPD3/6-31+G(d)) provides a gas-phase enthalpic change for the proton-transfer reaction between imidazole and methanethiol (the motifs of Cys and His side chains), in agreement with the experimentally derived value (see the Supporting Information). A systematic study for the cysteine-histidine proton transfer also pointed to the B3LYP functional as the most robust one to obtain an

electronic description in agreement with higher-level methods.<sup>29</sup> Finally, a recent QM/MM analysis of the electronic properties of enzyme–substrate complexes in SARS-CoV-2 protease emphasizes the importance of choosing an adequate functional for a proper description of the process.<sup>30</sup> All calculations were run with a modified version of Amber18<sup>17,31</sup> coupled to Gaussian16<sup>32</sup> for DFT calculations. A cutoff radius of 15 Å was used for all QM/MM interactions, and the temperature was 300 K.

To explore the free energy landscape associated with the chemical reaction, we used our implementation of the string method, the adaptive string method (ASM).<sup>33</sup> In this method  $N$  replicas of the system (the nodes of the string) are evolved according to the averaged forces and kept equidistant, converging in such a way to the minimum free energy path (MFEP) in a space of arbitrary dimensionality defined by the collective variables (CVs). Once the string has converged (see an example in Figure S1), we define a single path-CV (a collective variable called  $s$ ) that measures the advance of the system along the MFEP. This path-CV is used as the reaction coordinate to trace the free energy profile associated with the chemical transformations under analysis. The MFEPs were explored on a free energy hypersurface defined by a set of CVs formed by those distances showing relevant changes during the process under study: H-Sy(Cys145), H-N $\epsilon$ (His41), H-N(P1'), Sy(Cys145)-C(P1), and C(P1)-N(P1') for the acylation step (see Scheme 2a) and H<sub>w1</sub>-O<sub>w</sub>, H<sub>w1</sub>-N(P1'), O<sub>w</sub>-C(P1),

**Scheme 2. Collective Variables Employed To Study the Acylation (a) and Deacylation (b) Reaction Mechanisms**



Sy(Cys145)-C(P1), H<sub>w2</sub>-O<sub>w</sub>, N(P1')-N $\epsilon$ (His41), and H<sub>w2</sub>-Sy(Cys145) for the deacylation step (see Scheme 2b). Different initial guesses, corresponding to different mechanistic proposals, were explored at the B3LYPD3/MM level using first the 6-31G\* basis set, and then the best candidates were recalculated using the 6-31+G\* basis set. Each of the strings was composed of at least 96 nodes (see Table S1) that were propagated with a time step of 1 fs until the RMSD of the string fell below 0.1 amu<sup>1/2</sup>·Å for at least 2 ps. The converged MFEPs are then averaged to define the  $s$  path-CV corresponding to each string. The free energy profiles along the path-CVs were obtained using an umbrella sampling algorithm<sup>34</sup> running simulations for at least 10 ps and were integrated using the WHAM technique.<sup>35</sup> The values of the force constants employed to bias the ASM simulations were determined on-the-fly to ensure a probability density distribution of the reaction coordinate that was as homogeneous as possible.<sup>33</sup> Replica exchange between neighboring string nodes was attempted every 50 steps to improve convergence.

For the proton transfer between Cys145 and His41, considering the proximity of the proton donor and acceptor atoms and the geometrical simplicity of the process, we traced

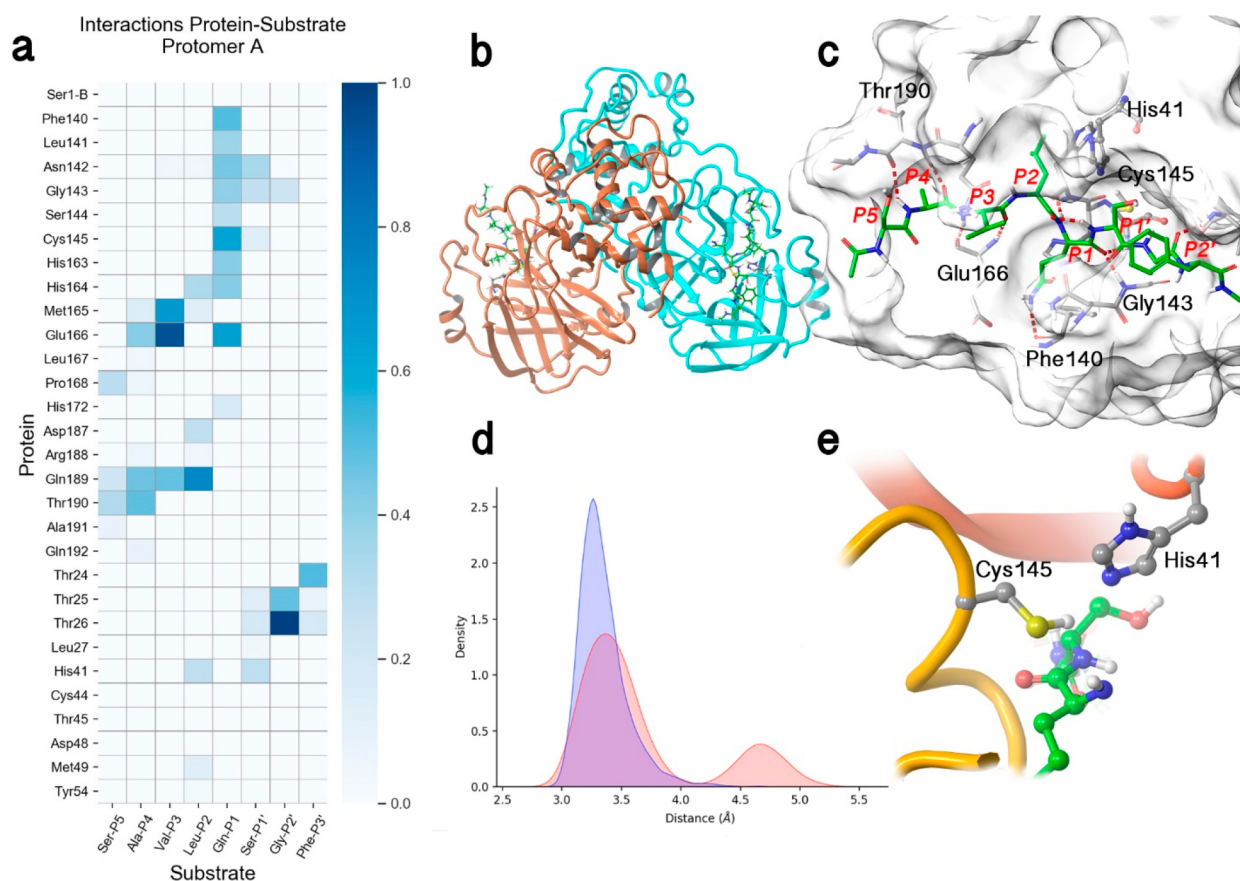
the free energy profile using umbrella sampling<sup>34</sup> along a simple proton-transfer coordinate defined as the antisymmetric combination of the distances of the proton to the donor and the acceptor atoms ( $d(\text{N}\epsilon\text{-H}) - d(\text{S}\gamma\text{-H})$ ). For this profile only the side chains of the two involved residues were included in the QM region (using the B3LYPD3/6-31+G\* level of theory). A total of 40 windows were used, corresponding to an increment of the reaction coordinate of 0.06 Å, with each of them composed of 10 ps of equilibration and 20 ps of data collection. The force constant employed to drive the reaction coordinate change was 600 kcal·mol<sup>-1</sup>·Å<sup>-2</sup>. All the rest of the details of the simulations were as described earlier. We also used umbrella sampling along distinguished coordinates to explore the free energy landscape associated with the separation between the first peptide fragment and the acyl-enzyme complex and between the two peptide fragments at the end of the deacylation process. Details of all the free energy simulations performed in this work are given in Table S1.

### 3. RESULTS

**3.1. Enzyme–Substrate Complex.** The time evolutions of the root-mean-square deviation (RMSD) values for each replica (2 protomers and 2 substrates in each replica) are shown in Figure S2. These values show that the protein structure is well-equilibrated and there are no large differences with respect to the initial structures, as prepared from the X-ray data. The observations made on the three replicas (one of 8  $\mu$ s and two of 1  $\mu$ s) are very similar in all cases. The substrate-binding pocket is divided into a series of subsites (denoted as S and S'), each accommodating a single residue of the substrate placed before (P) or after (P') the scissile peptide bond. The map of hydrogen-bond interactions observed during our molecular dynamics (MD) simulations of the Michaelis complex is given in Figure 1a, a general view of the substrate in the two active sites of the dimer formed by protomers A and B is shown in Figure 1b, and an insight into the active site of protomer A is provided in Figure 1c. The 3CL<sup>pro</sup> of SARS-CoV-2 (as is also the case of the SARS-CoV ortholog) presents a high specificity for Gln at the P1 position.<sup>11,36</sup> As seen in Figure 1a, the P1 residue is the one establishing more hydrogen-bond interactions with the enzyme. The O and N main-chain atoms of Gln-P1 are found to make hydrogen bonds with main-chain atoms of Gly143, Ser144, and His164 (unless indicated, all the residues of the protein belong to the same protomer A). The side chain of Gln-P1 is accommodated into the S1 subsite through hydrogen-bond contacts with the main-chain atoms of Phe140 and Leu141, with the N $\epsilon$  atom of His163 and the O $\epsilon$  atoms of Glu166. This last residue is in turn hydrogen-bonded to the terminal NH group of Ser1 from protomer B. In fact, the N-terminal fragment (N-finger) of protomer B plays an active role in preorganizing the active site of protomer A for catalysis.<sup>37</sup> Dimerization is an essential condition for catalysis in the protease of a related coronavirus,<sup>36,38,39</sup> and consequently, those mutants lacking the N-finger fragment are almost completely inactive.<sup>40</sup> The side chain of Leu-P2 is surrounded by the side chains of His41, Met49, His164, Met165, and Asp187, while the main-chain amide group is hydrogen-bonded to the O $\epsilon$  atom of Gln189. The side chain of Val-P3 is solvent-exposed, while main-chain N and O atoms are hydrogen-bonded to main-chain atoms of Glu166.

The binding subsite for Ala-P4 is constituted of main-chain interactions with Gln189 and Thr190, while the side chain of





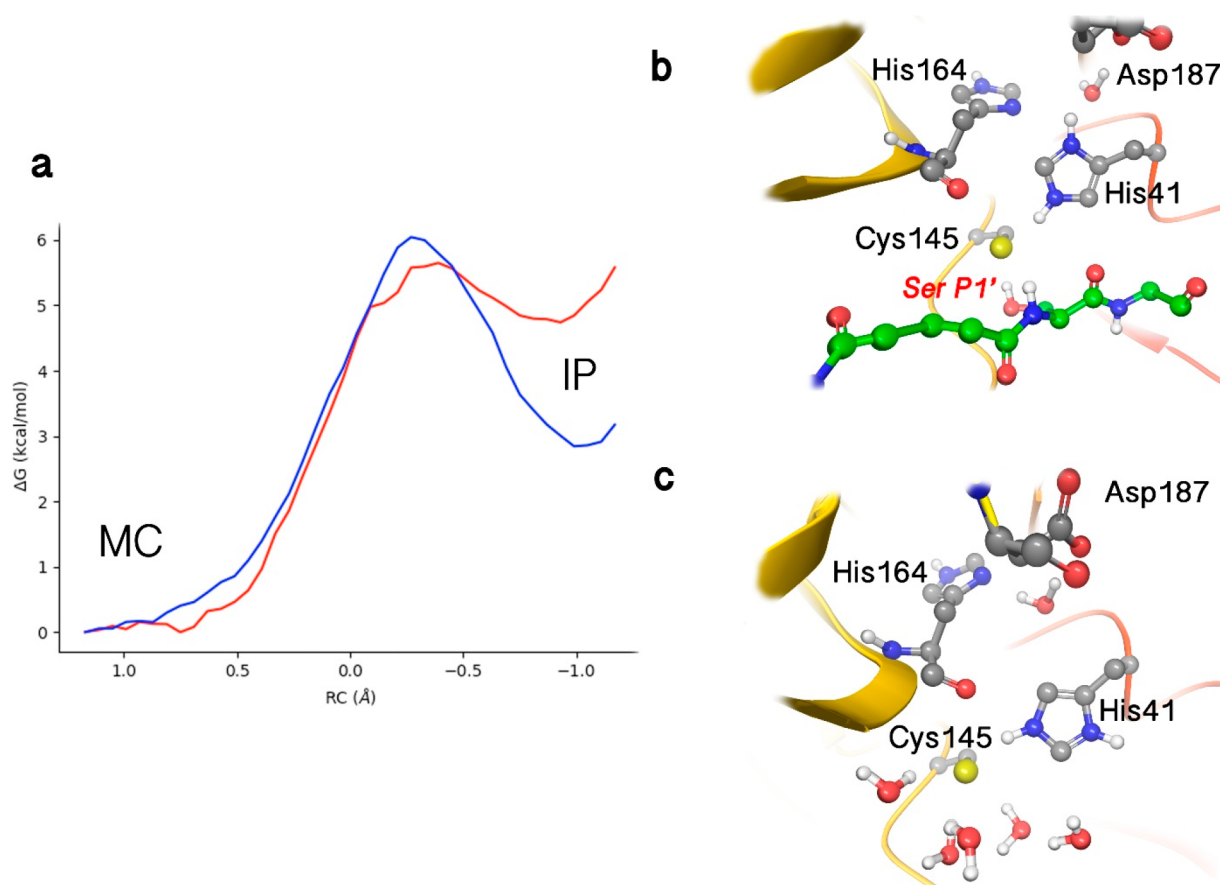
**Figure 1.** Results of the molecular dynamics simulation of 3CL<sup>pro</sup> of SARS-CoV-2 in complex with the substrate with sequence Ac-Ser-Ala-Val-Leu-Gln-Ser-Gly-Phe-NMe. (a) Fraction of hydrogen-bond contacts between the residues of the substrate and those of the protease found during the trajectory of the Michaelis complex. A hydrogen-bond contact is counted when the donor–acceptor distance is  $<3.8$  Å and the hydrogen-bond angle is  $>120^\circ$ . (b) General overview of the substrate–enzyme complex, showing the dimeric nature of the protease with two identical active sites occupied by the substrate. Note that the N-finger of each protomer is close to the active site of the neighbor protomer. (c) Insight into the binding pose of the peptide substrate into the active site, showing the most important active-site residues and the positions occupied by the P and P' residues of the substrate. (d) Probability densities of the distances from the Cys145-S<sub>γ</sub> atom to the carbonyl carbon atom of the substrate (C(P1)), in red, and to the N<sub>ε</sub> atom of His41, in blue. The bimodal distribution of S<sub>γ</sub>-C(P1) distances correspond to the trans (shorter distances) and gauche (longer distances) conformations of Cys145. (e) Disposition of the substrate in the vicinity of the catalytic dyad when Cys145 is present in the trans conformation. Note the proximity between S<sub>γ</sub> and C(P1) atoms and the orientation of the sulfhydryl proton toward the N<sub>ε</sub> atom of His41.

this residue is placed between the side chains of Met165, Leu167, and Pro168. The S5 subsite is formed by the side chain of Pro168 and by main-chain atoms of Thr190. This description of the S1–S5 interaction subsites observed in our simulations agrees with the characteristics found in the X-ray structures of SARS-CoV-2 3CL protease with inhibitors bound in the active site.<sup>6,8</sup> Our MD simulations of the substrate–enzyme complex offer, in addition, a detailed description of the S' subsites, those that accommodate the P' residues placed after the scissile peptide bond. The main-chain O atom of Ser-P1' establishes a hydrogen bond with the amide group of Gly143 and the side chain of Asn142. The hydroxyl group of the P1' side chain can contact with the catalytic dyad (Cys145 and His41), while the CH<sub>2</sub> group is packed between Thr25, Thr26, and Leu27. Gly-P2' is stabilized through main-chain contacts with Thr25 and Thr26. Finally, the side chain of the Phe-P3' residue is packed against the side chain of Thr24. This structural information can be useful in order to improve the binding and specificity of potential inhibitors of the protease activity because these structural findings are lost in the X-ray structures obtained from those inhibitors in which the

fragment corresponding to P' residues either is released during the formation of the acyl–enzyme complex<sup>13</sup> or is smaller than in our substrate.<sup>6,8</sup>

We have also analyzed the substrate–enzyme interactions in protomer B (see Figure S3). The interactions averaged over the  $\mu$ s-time-scale simulations show only small differences between the two protomers, indicating that the poses found in the two active sites are consistent with the reported data.

**3.2. Catalytic Dyad.** The reaction mechanism of cysteine proteases involves the nucleophilic attack of the S<sub>γ</sub> atom of a cysteine (Cys145 in our case) to the C(P1) atom of the peptide bond. Figure 1d shows the probability distribution of S<sub>γ</sub>-C(P1) distances found during our MD simulations. The distribution is clearly bimodal, with two peaks centered at 3.4 and 4.7 Å. These two peaks correspond to two different conformations of the side chain of Cys145, which can be present in trans and gauche conformations. This is in agreement with the observations made on the X-ray structure of the ortholog protease of SARS-CoV.<sup>37</sup> The most probable conformation corresponds to the trans conformer in which the S<sub>γ</sub> sulfur atom is closer to the substrate (see Figure 1e). In



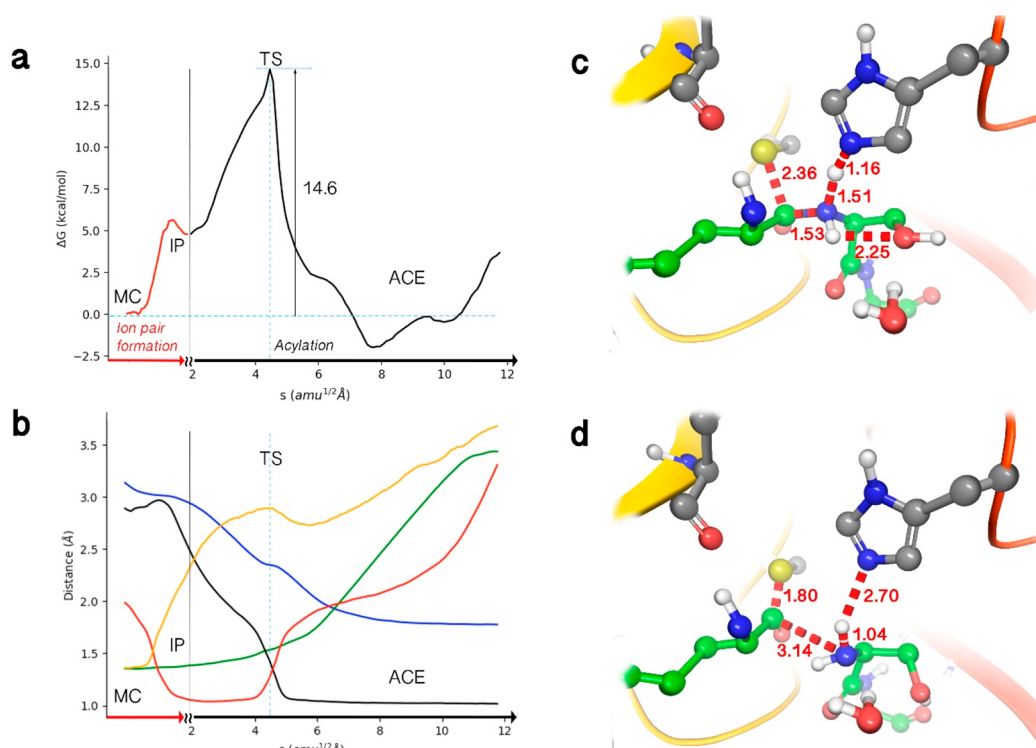
**Figure 2.** Analysis of the formation of an ion pair (IP) catalytic dyad ( $\text{Cys145}^- \cdots \text{His41H}^+$ ) from the neutral form ( $\text{Cys145H} \cdots \text{His41}$ ) found in the Michaelis complex (MC). (a) B3LYPD3/6-31+G\*/MM free energy profile associated with the proton-transfer coordinate from the  $S\gamma$  atom of Cys145 to the  $N\epsilon$  atom of His41 ( $d(N\epsilon\text{-H}) - d(S\gamma\text{-H})$ ) in the apo (blue line) and holo (red line) enzymes. (b) Representation of the ion pair in the holo enzyme, showing those interactions that stabilize the charged states of the catalytic dyad, in particular the hydroxyl group of Ser(P1'). (c) Representation of the ion pair in the apo enzyme, showing the presence of water molecules that stabilize the charged catalytic dyad when the substrate is absent.

both conformations the catalytic dyad remains hydrogen-bonded, with the most probable distance between Cys145- $S\gamma$  and His41- $N\epsilon$  being about 3.3 Å (see Figure 1d). Interestingly, His41 is, in turn, hydrogen-bonded through a highly conserved crystallographic water molecule to Asp187. This interaction can raise the  $pK_a$  of the histidine, increasing its ability to work as a base and abstract the proton from Cys145.

Considering that the 3CL protease can be asymmetrical,<sup>38</sup> we also analyzed a 1  $\mu\text{s}$  MD simulation of the dimer with only one of the protomers occupied by the substrate. Figure S4 shows that the distribution of  $S\gamma\text{-C}(P1)$  distances is very similar to that obtained when the substrate is present in the two active sites. A small decrease in the population of the unproductive Cys145 gauche conformer is observed when only one protomer is occupied, which could contribute to the activity of the enzyme. In our simulations of the reaction mechanism, we always selected the Cys145 trans conformation as the initial structure.

Taking into account the short distance observed between Cys145 and His41 and the possible activation of this last residue by the nearby Asp187, we explored the possibility to find the catalytic dyad forming an ion pair ( $\text{Cys}^-/\text{HisH}^+$  or IP in Figure 2) instead of the neutral form modeled in the Michaelis complex ( $\text{CysH}/\text{His}$  or MC in Figure 2). We evaluated the free energy difference between these two forms

of the dyad by means of free energy profiles associated with the proton-transfer coordinate from Cys145 to His41 obtained at the B3LYPD3/MM level (see Methods). The proton-transfer free energy profiles (Figure 2a) were obtained for the holo and apo forms of 3CL<sup>Pro</sup>. According to Figure 2a the catalytic dyad is more stable in its neutral form, for both the apo and holo forms. The IPs are 2.9 and 4.8 kcal·mol<sup>-1</sup> above the neutral form in the apo and holo enzymes, respectively. The anionic Cys145 can be stabilized by the presence of water molecules and by the hydroxyl group of Ser-P1' (see Figure 2b). In an experimental analysis of the substrate specificity of the ortholog SARS-CoV protease, it was found that mutation of Ser-P1' to another small residue (such as Gly or Ala) diminishes the rate constant by about 1 order of magnitude.<sup>36</sup> A similar effect on the rate constant was observed for SARS-CoV-2 protease when the leaving group was changed to *p*-nitroaniline.<sup>41</sup> These variations in the rate constant could be attributed, in part, to a stabilizing effect of the hydroxyl group of Ser-P1' on the ion pair of about 1.4 kcal·mol<sup>-1</sup>. Note that the requirement of a small residue at the P1' position by SARS-CoV and SARS-CoV-2 main proteases could be due in part to the fact that in these cases water molecules can access the active site, contributing also to the stabilization of the ion pair. Bulkier residues at the P1' position could hinder the access of water molecules and thus destabilize the ionized form



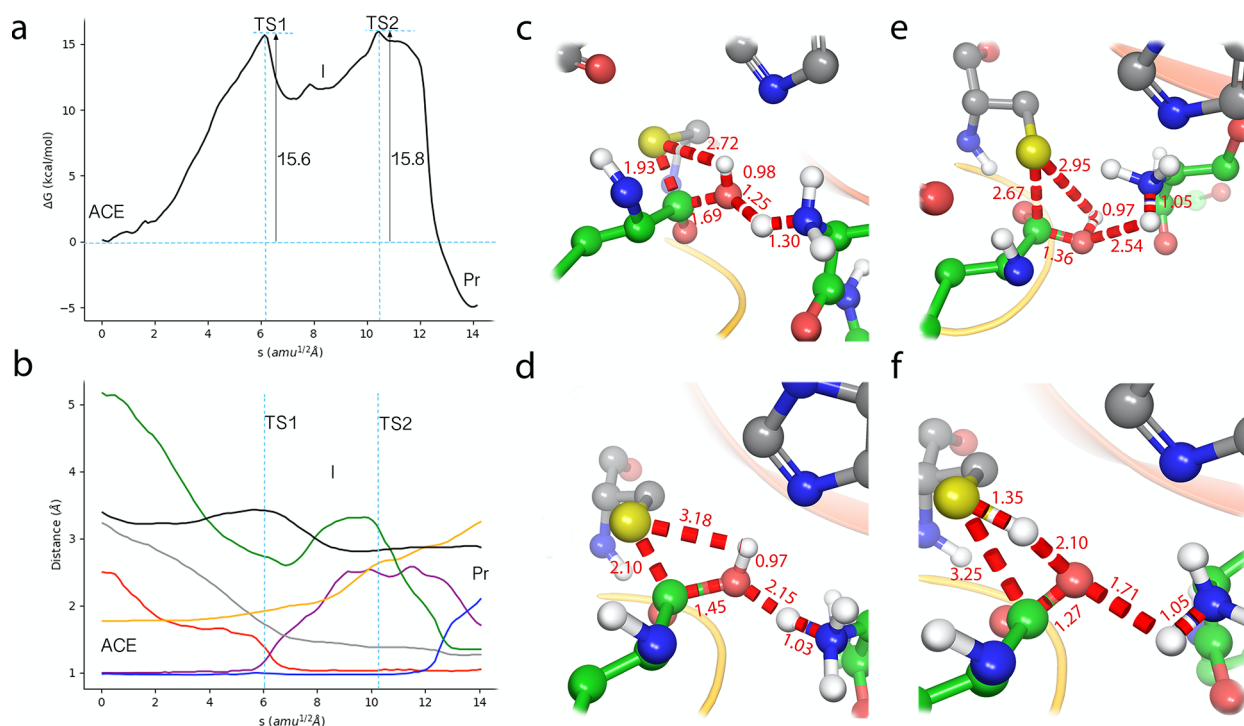
**Figure 3.** Simulation of the acylation reaction taking place through the formation of the ion pair. (a) B3LYPD3/6-31+G\*/MM free energy profile along the path-CV for the acylation reaction after the formation of the ion pair (IP) from the Michaelis complex (MC). The reaction takes place with a single transition state (TS) that yields the acyl-enzyme (ACE), which can be present in two conformations. (b) Evolution of the distances selected as collective variables (CVs) along the minimum free energy path (MFEP).  $S\gamma$ -H is in yellow, H-N $\epsilon$  is in red, C(P1)-N(P1') (the scissile peptide bond) is in green, H-N(P1') is in black, and  $S\gamma$ -C(P1) is in blue. (c) Representation of the TS for the acylation process. This TS corresponds to the proton transfer from His41 in the ion pair catalytic dyad to the nitrogen atom of the peptide bond (N(P1')) with the approach of the  $S\gamma$  atom of Cys145 to the carbonyl carbon atom (C(P1)) and the lengthening of the peptide bond. The values of the distances correspond to the coordinates of the MFEP at the TS, except the intramolecular distance between the hydroxyl and NH group of Ser(P1') that has been averaged over the trajectory of the corresponding string node. (d) Representation of the acyl-enzyme complex formed between the enzyme and the P fragment of the peptide, with a water molecule hydrogen-bonded to the N-terminal group of the P' fragment. The free energy profile shows two minima for the acyl-enzyme complex, differing in the distance between the P and P' fragments, as shown in Figure S5.

of the catalytic dyad. According to our free energy profiles shown in Figure 2a, the IP form is better stabilized with respect to the neutral dyad in the apo enzyme than in the holo one (by almost 2 kcal·mol<sup>-1</sup>), which can be related to the better solvation of the catalytic dyad in the former. In both the holo and apo enzymes, the free energy barrier associated with the transfer of the proton between His41 to Cys145 is small, revealing a fast equilibrium between the ion pair and neutral versions of the dyad, with the latter being the predominant form. The existence of a low-lying IP dyad is compatible with the experimental observations made in the kinetic characterization of the homologue 3CL<sup>pro</sup> of SARS-CoV, in which an ion pair mechanism for the proteolysis was proposed on the basis of the pH-inactivation profile with iodoacetamide and the analysis of solvent isotope effects.<sup>41</sup>

The picture derived from our QM/MM free energy profiles contrasts with the results obtained from the exploration of the proton transfer potential energy surface in the SARS-CoV 3CL<sup>pro</sup>.<sup>42</sup> This study suggests a scenario where substrate binding could reduce the energy cost of forming the IP. However, the reported energy difference between the IP and neutral complexes was significantly higher than our free energy estimation, about 11 kcal·mol<sup>-1</sup>, and difficult to reconcile with the observed rate constants (see below) and pK<sub>a</sub> determinations.<sup>43</sup>

**3.3. Acylation Step.** We explored the free energy landscape for the formation of the acyl-enzyme starting from the catalytic dyad IP at the B3LYPD3/MM level. The converged MFEP is shown in Figure 3a and b. According to our simulations, after IP formation the acylation proceeds by means of a proton transfer from His41 to the N(P1') atom followed by the nucleophilic attack of Cys145- $S\gamma$  on the C(P1) atom and the simultaneous breaking of the C(P1)-N(P1') peptide bond. These elementary events take place in a concerted but asynchronous way. The transition state (TS) found for this mechanism (see Figure 3c) is associated with the proton transfer from His41 to the amide nitrogen atom of the peptide bond N(P1'). At the TS the  $S\gamma$  atom of Cys145 approaches the C(P1) atom, reducing the interatomic distance from 3.11 to 2.34 Å. This approach is accompanied by a moderate lengthening of the peptide bond (the C(P1)-N(P1') distance being lengthened from 1.36 to 1.54 Å). According to the free energy path shown in Figure 3a, the total free energy barrier associated with the acylation process, including the free energy cost of the ion pair formation, is 14.6 kcal·mol<sup>-1</sup>. This value is compatible with the activation free energies derived from the steady-state rate constants measured at 25 °C for peptides cleaved at the Gln-Ser bond by the highly similar ortholog 3CL<sup>pro</sup> of SARS-CoV (between 16.2 and 17.2 kcal·mol<sup>-1</sup>).<sup>41</sup> It must be noticed that in this proteolysis the acylation step is not considered to be the rate-limiting one and



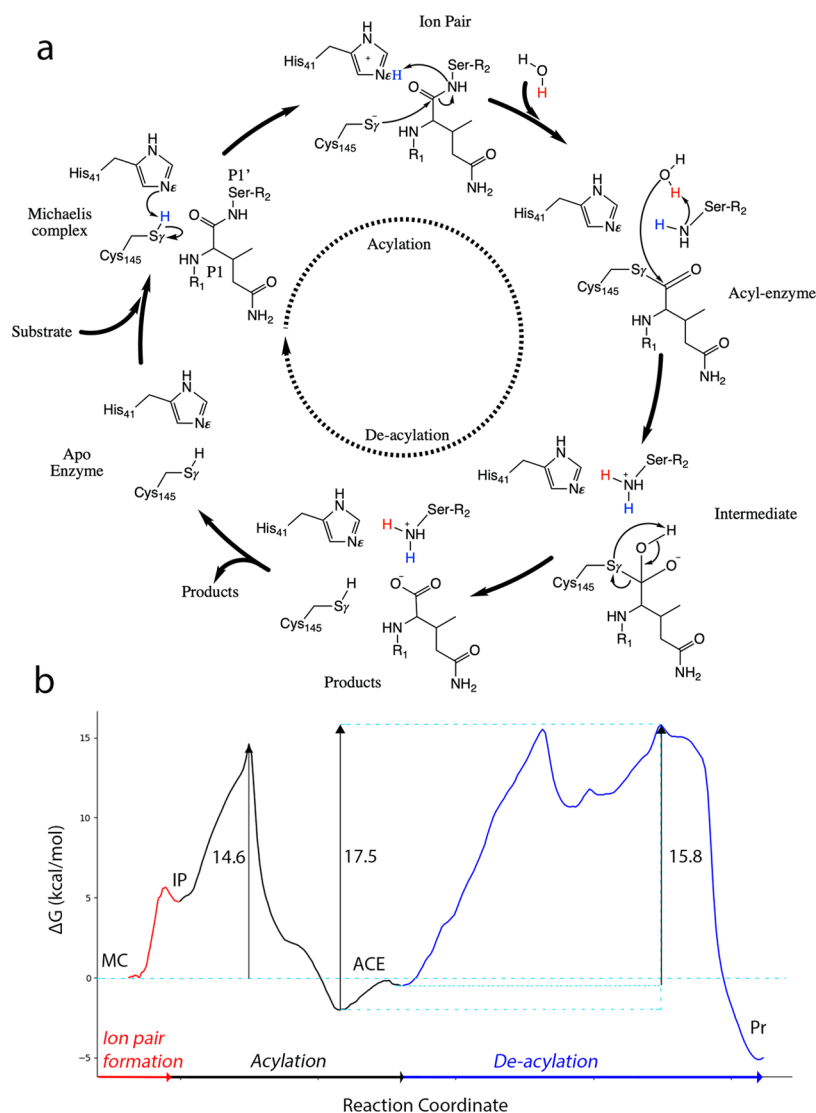


**Figure 4.** Simulation of the deacylation reaction with the N-terminal group of the P' fragment acting as a general base in charge of water activation. (a) B3LYPD3/6-31+G\*/MM free energy profile along the path-CV for the deacylation reaction. The process takes place in two steps. In the first one the water activated by the N-terminal group attacks the acyl–enzyme complex (ACE) to form a thiodiolate intermediate (I) through the first TS (TS1). In the second one the reaction product (Pr) is obtained after breaking the acyl–enzyme bond in the second TS (TS2). (b) Evolution of the distances selected as collective variables along the minimum free energy path.  $S\gamma$ -H<sub>w2</sub> is in green,  $N\epsilon$ -N(P1') is in black,  $O_w$ -C(P1) is in gray,  $N$ (P1')-H<sub>w1</sub> is in red,  $S\gamma$ -C(P1) is in yellow,  $O_w$ -H<sub>w1</sub> is in purple, and  $O_w$ -H<sub>w2</sub> is in blue. (c) Representation of TS1 where the water molecule is transferring a proton to the N-terminal group of the P' fragment and the resulting hydroxyl anion attacks the carbonyl carbon atom of the P fragment (C(P)). The values of the distances correspond to the coordinates of the MFEP at TS1. (d) Representation of the thiodiolate intermediate (I). (e) Representation of TS2 corresponding to the breaking of the  $S\gamma$ -C(P1) bond and the proton transfer from the carboxylic terminal group to the leaving sulfur atom. (f) Representation of the reaction products (Pr) with the P-COO<sup>-</sup> and P'-NH<sub>3</sub><sup>+</sup> peptide fragments in the active site of the protease.

then these experimental activation free energies provide an upper limit for the acylation barrier.<sup>41</sup> Figure 3c shows that this TS is stabilized by means of a hydrogen-bond interaction with the hydroxyl group of Ser-P1', indicating, also in agreement with the experimental observations, that the leaving group plays an important role in catalysis. The TS stabilization provided by the hydroxyl group could also contribute to the larger rate constant observed for the scission of the Gln-Ser bond.<sup>36</sup> Remarkably, the proposed reaction mechanism is also in good agreement with the experimental proton inventory results, which indicate that there are two protons in flight during the acylation, one at the TS and another one at earlier stages.<sup>41</sup> In our picture these two proton-transfer events correspond to the proton transferred from His41 to the N(P1') atom of the substrate at the TS and the proton transfer from Cys145 to His41 during IP formation. Regarding the acyl–enzyme product (Figure 3d), our free energy profile shows two possible conformations that differ in the presence of a water molecule that plays a key role during deacylation (see Figure S5). Finally, in our free energy profile the formation of the acyl–enzyme (see Figure 3d) is almost thermoneutral, with a reaction free energy of about  $-1$  kcal·mol<sup>-1</sup>.

Very recently an interesting QM/MM study of the same protease with a different substrate (where the leaving group was not a peptide fragment but a fluorescent tag, 7-amino-4-carbamoylmethylcoumarin) has been reported.<sup>44</sup> In that work the proton transfer from Cys145 to His41 was found to be

concomitant with the nucleophilic attack of the  $S\gamma$  atom on the carbonyl carbon atom, forming a thiohemiketal intermediate, and the cleavage of the peptide bond takes place in a subsequent step assisted by the proton transfer from His41. The differences with respect to our results could be due to the use of a different substrate and/or the use of different theoretical descriptions (that work used a combination of semiempirical and DFT methods with the M06-2X functional). The description of the acylation step can be highly dependent on the selection of the QM level, as demonstrated in Figure S6 where we compare the free energy profiles obtained with different Hamiltonians and basis sets. In any case, that QM/MM study obtained an activation free energy of 19.9 kcal·mol<sup>-1</sup>, in excellent agreement (within 0.5 kcal·mol<sup>-1</sup>) with the value derived from the experimental rate constant obtained for that substrate.<sup>10</sup> Interestingly, this rate constant ( $0.050$  s<sup>-1</sup>)<sup>10</sup> is significantly smaller than the value reported for the hydrolysis of the Gln-Ser bond by the ortholog enzyme of SARS-CoV ( $1.5$ – $8.5$  s<sup>-1</sup>).<sup>41</sup> The gap between these two experimental rate constant values could be due to differences in the preparation and purification of the enzyme or to genuine mechanistic differences between substrates in the main protease.<sup>41,45</sup> In this sense, as discussed earlier, the presence of a hydroxyl group at the P1' position could play an important role in the acylation process, improving the binding and kinetics of a hypothetical inhibitor. To explore other possible mechanisms,<sup>46</sup> we also studied a reaction path that does not



**Figure 5.** Proteolysis mechanism in 3CL<sup>pro</sup> of SARS-CoV-2 as deduced from our simulations. (a) Schematic representation of the reaction mechanism. (b) Free energy profile associated with the reaction mechanism.

involve the formation of an ion pair, although the associated free energy barrier is considerably higher and incompatible with the experimental rate constant (see Figure S7).

**3.4. Deacylation Step.** Regarding the deacylation step, the standard mechanistic proposal suggested for related enzymes<sup>47</sup> assumes that, once the neutral P'-NH<sub>2</sub> peptide fragment has left the active site, a water molecule activated by His41 attacks the C(P1)-S $\gamma$  bond, releasing the P-COOH peptide (also with a neutral terminal group) and, finally, regenerating the enzyme after a proton transfer from His41 to Cys145 (see Scheme 1). In our simulations of the acylation product, we found that a water molecule can be placed in between the P'-NH<sub>2</sub> leaving fragment and the acyl-enzyme complex, being correctly oriented to perform the hydrolysis of the acyl-enzyme (see Figures 3d and S5). This configuration suggests an alternative reaction mechanism that can yield the two peptide fragments with correct protonation states in the terminal groups and regenerate the enzymatic active site in its most stable state (the neutral catalytic dyad). An additional advantage of this novel mechanism is that it involves water activation by means of the N-terminus of the P' fragment, which is known to be a better base than histidine side chains (the average pK<sub>a</sub> values are

about 7.7 and 6.6, respectively).<sup>48</sup> Finally, the proposed mechanism can be at the origin of the mechanistic differences observed when the scissile bond is an amide instead of an ester (in that case a basic N-terminal group is not formed after the acylation).<sup>41</sup> We obtained the MFEP corresponding to such a mechanism at the B3LYPD3/MM level (see Figure 4). This process is stepwise, presenting two TSs (see Figure 4a). The first TS (TS1 in Figure 4c) corresponds to the proton transfer from the water molecule to the N(P1') atom, resulting in the formation of the P'-NH<sub>3</sub><sup>+</sup> peptide fragment. The free energy barrier associated with this step is 15.6 kcal·mol<sup>-1</sup>. This proton transfer is concomitant to the attack of the hydroxyl group on the C(P1) carbonyl carbon atom, resulting in the formation of an intermediate thiodiolate (Figure 4d). After rotation of the hydroxyl group to orient the proton toward the sulfur atom, the reaction proceeds with the cleavage of the C(P1)-S $\gamma$  bond. The second TS observed during the deacylation (TS2, Figure 4e) corresponds to the separation of the S $\gamma$  atom (with the C(P1)-S $\gamma$  distance being 2.67 Å). The free energy barrier associated with this second step from the acyl-enzyme complex is very close to the first one, 15.8 kcal·mol<sup>-1</sup> (17.5 kcal·mol<sup>-1</sup> from the most stable acyl-enzyme conformer; see



Figure 5). This value is in excellent agreement with the values derived from the reaction rate constants for the ortholog protease of SARS-CoV (from 16.2 to 17.2 kcal·mol<sup>-1</sup>).<sup>41</sup> Afterward, the leaving cysteine is stabilized by means of a proton transfer from the C-terminal group to the S $\gamma$  atom, regenerating the enzyme in its more stable protonation state (a neutral catalytic dyad) and yielding the P peptide fragment with a terminal unprotonated carboxylate (the product is represented in Figure 4f). The proposed mechanism, in which the general base is a N-terminal group, displays a smaller barrier than the standard mechanism in which His41 acts as the general base activating the water molecule, as expected from the relative pK<sub>a</sub> values (see Figure S8).

The deacylation step presents a free energy change of about -5 kcal·mol<sup>-1</sup>. This exergonic character can be increased with the release of the reaction products to the bulk. It is worth noticing that the peptide fragments obtained from this mechanistic proposal present a salt bridge between the charged C-terminal and N-terminal groups that must be broken during products release. The separation of the two terminal groups implies that water molecules must be placed, tightly bounded, between these two charged groups. Those configurations could contribute to an inverse solvent isotope effect observed under steady-state conditions only when the scissile bond is an amide and not an ester (this is only when a N-terminal group is available to act as the general base).<sup>41</sup>

#### 4. CONCLUSIONS

We have presented a detailed analysis of the Michaelis complex and the proteolysis mechanism in the 3CL<sup>pro</sup> of SARS-CoV-2 using DFT/MM computational simulations. Our study has identified key interactions established between the protein and a peptide substrate and the detailed reaction mechanism. The complete reaction cycle is shown in Figure 5a, while the associated free energy profile is given in Figure 5b. The reaction process involves the formation of a catalytic dyad ion pair from which the acylation step can proceed (see Video 1 in the Supporting Information). In this acylation, the TS involves the proton transfer from His41 to the nitrogen atom followed by the nucleophilic displacement of the peptide bond by the S $\gamma$  atom of Cys145. This finding can be relevant for the design of inhibitors with improved kinetics, stabilizing this proton transfer through, for example, an adequate choice of the P1' group. For the deacylation step (see Video 2 in the Supporting Information), we have proposed a completely novel mechanism where the N-terminal group of the first peptide fragment acts as the general base catalyzing the hydrolysis of the acyl-enzyme complex.

The free energy profile associated with the proposed reaction mechanism is shown in Figure 5b. According to this profile, the acylation step is almost thermoneutral, while the deacylation is more exergonic, in agreement with the observed irreversibility of the process. Notably, our mechanistic proposal for the deacylation step can explain the experimental differences observed between the hydrolysis of amide and ester substrates in the highly similar protease of SARS-CoV, which had remained unexplained until now. The predicted global free energy barrier, 17.5 kcal·mol<sup>-1</sup>, is in excellent agreement with the values derived from the observed rate constants for the hydrolysis of the Gln-Ser peptide bond in the highly similar SARS-CoV protease (between 16.2 and 17.2 kcal·mol<sup>-1</sup>).<sup>41</sup> Note that the rate constants for the hydrolysis of different substrates with a fluorescent tag as the leaving group

have been also measured for the SARS-CoV-2 protease.<sup>10</sup> The observed rate constants were smaller than those measured for the Gln-Ser hydrolysis in the SARS-CoV protease,<sup>41</sup> and the activation free energies derived from these rate constants measured for the SARS-CoV-2 protease, between 18.6 and 19.4 kcal·mol<sup>-1</sup>, are somewhat larger than our calculated value. However, because of the differences between experimental procedures, it is difficult to assess if the observed disparity among rate constants in the two proteases is due to intrinsic properties of the two enzymes (which is unlikely due to the identity between active sites) or, more likely, due to differences in the preparation of the enzyme (which can affect the measured rate constants) or to differences in the substrate (which can change the rate constant by 1 order of magnitude or more). We thus preferred to use as a reference for our study the detailed kinetic characterization made on the SARS-CoV protease with the same substrate as that employed in our simulations.<sup>41</sup>

Our study can also be useful to assist in the design of new and more potent inhibitors of SARS-CoV-2 main protease. Apart from the well-known interactions established by the P1-P4 groups in the active site, our MD simulations of the enzyme-substrate complex stress the role of the interactions established by the P1' and P2' groups. In particular, the hydroxyl group at the P1' position could be important not only for binding but also during the reaction process because it could assist the formation of the catalytic dyad ion pair and the acylation transition state. Interestingly, during the preparation of this manuscript a preprint has shown that a hydroxymethylketone derivative can be a potent inhibitor of SARS-CoV-2 3CL protease.<sup>49</sup> In this compound the hydroxyl group can establish a direct contact with the catalytic dyad.<sup>50</sup>

#### ■ ASSOCIATED CONTENT

##### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acscatal.0c03420>.

Plot of the string method convergence; RMSD plots of MD simulations; comparison between the active sites of the two protomers; distribution of S-C distances in the dimer with one and two substrates; a free energy profile corresponding to the acyl-enzyme complex; study of the impact of QM level on the acylation profile; alternative mechanisms for the acylation; deacylation mechanism when His41 is the general base; details of free energy calculations; and a gas-phase evaluation of the performance of the selected functional and basis set (PDF)

Formation of a catalytic dyad ion pair from which the acylation step can proceed (MP4)

Novel mechanism where the N-terminal group of the first peptide fragment acts as the general base catalyzing the hydrolysis of the acyl-enzyme complex (MP4)

#### ■ AUTHOR INFORMATION

##### Corresponding Authors

J. Javier Ruiz-Pernía – Departamento de Química Física, Universidad de Valencia, 46100 Burjassot, Spain; [orcid.org/0000-0002-4640-0419](https://orcid.org/0000-0002-4640-0419); Email: [j.javier.ruiz@uv.es](mailto:j.javier.ruiz@uv.es)

Iñaki Tuñón – Departamento de Química Física, Universidad de Valencia, 46100 Burjassot, Spain; [orcid.org/0000-0002-6995-1838](https://orcid.org/0000-0002-6995-1838); Email: [ignacio.tunon@uv.es](mailto:ignacio.tunon@uv.es)

## Author

Carlos A. Ramos-Guzmán – Departamento de Química Física, Universidad de Valencia, 46100 Burjassot, Spain

Complete contact information is available at:  
<https://pubs.acs.org/10.1021/acscatal.0c03420>

## Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

The authors acknowledge financial support from Feder funds and the Ministerio de Ciencia, Innovación y Universidades (Project PGC2018-094852-B-C22). We also acknowledge PRACE for awarding us access to MareNostrum based in Spain at Barcelona Supercomputing Center (BSC). The support of Alejandro Soriano from Servei d'Informàtica de la Universitat de València and Cristian Morales and David Vicente from BSC for the technical work is gratefully acknowledged. The authors also deeply acknowledge Dr. Kirill Zinovjev for assistance in the adaptation of the code and helpful discussions about the use of the string method.

## REFERENCES

- (1) Zhou, P.; Yang, X.-L.; Wang, X.-G.; Hu, B.; Zhang, L.; Zhang, W.; Si, H.-R.; Zhu, Y.; Li, B.; Huang, C.-L.; Chen, H.-D.; Chen, J.; Luo, Y.; Guo, H.; Jiang, R.-D.; Liu, M.-Q.; Chen, Y.; Shen, X.-R.; Wang, X.; Zheng, X.-S.; Zhao, K.; Chen, Q.-J.; Deng, F.; Liu, L.-L.; Yan, B.; Zhan, F.-X.; Wang, Y.-Y.; Xiao, G.-F.; Shi, Z.-L. A Pneumonia Outbreak Associated with a New Coronavirus of Probable Bat Origin. *Nature* **2020**, *579*, 270–273.
- (2) Lau, S. K. P.; Woo, P. C. Y.; Li, K. S. M.; Huang, Y.; Tsoi, H.-W.; Wong, B. H. L.; Wong, S. S. Y.; Leung, S.-Y.; Chan, K.-H.; Yuen, K.-Y. Severe Acute Respiratory Syndrome Coronavirus-like Virus in Chinese Horseshoe Bats. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 14040–14045.
- (3) Cheng, V. C. C.; Lau, S. K. P.; Woo, P. C. Y.; Yuen, K. Y. Severe Acute Respiratory Syndrome Coronavirus as an Agent of Emerging and Reemerging Infection. *Clin. Microbiol. Rev.* **2007**, *20*, 660–694.
- (4) Pillaiyar, T.; Manickam, M.; Namasivayam, V.; Hayashi, Y.; Jung, S.-H. An Overview of Severe Acute Respiratory Syndrome–Coronavirus (SARS-CoV) 3CL Protease Inhibitors: Peptidomimetics and Small Molecule Chemotherapy. *J. Med. Chem.* **2016**, *59*, 6595–6628.
- (5) Bangham, C. R. M. The Immune Control and Cell-to-Cell Spread of Human T-Lymphotropic Virus Type 1. *J. Gen. Virol.* **2003**, *84*, 3177–3189.
- (6) 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. X.; You, T.; Liu, X. X.; Yang, X. X.; Bai, F.; Liu, H.; Liu, X. X.; Guddat, L. W.; Xu, W.; Xiao, G.; Qin, C.; Shi, Z.; Jiang, H.; Rao, Z.; Yang, H. Structure of Mpro from COVID-19 Virus and Discovery of Its Inhibitors. *Nature* **2020**, *582*, 289–293.
- (7) 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  $\alpha$ -Ketoamide Inhibitors. *Science* **2020**, *368*, 409–412.
- (8) Dai, W.; Zhang, B.; Jiang, X.-M.; Su, H.; Li, J.; Zhao, Y.; Xie, X.; Jin, Z.; Peng, J.; Liu, F.; Li, C.; Li, Y.; Bai, F.; Wang, H.; Cheng, X.; Cen, X.; Hu, S.; Yang, X.; Wang, J.; Liu, X.; Xiao, G.; Jiang, H.; Rao, Z.; Zhang, L.-K.; Xu, Y.; Yang, H.; Liu, H. Structure-Based Design of Antiviral Drug Candidates Targeting the SARS-CoV-2 Main Protease. *Science* **2020**, *368*, 1331–1335.
- (9) Chen, Y. W.; Yiu, C.-P. B.; Wong, K.-Y. Prediction of the SARS-CoV-2 (2019-nCoV) 3C-like Protease (3CLpro) Structure: Virtual Screening Reveals Velpatasvir, Ledipasvir, and Other Drug Repurposing Candidates. *FI000Research* **2020**, *9*, 129.
- (10) Rut, W.; Groborz, K.; Zhang, L.; Sun, X.; Zmudzinski, M.; Pawlik, B.; Mlynarski, W.; Hilgenfeld, R.; Drag, M. Substrate Specificity Profiling of SARS-CoV-2 Main Protease Enables Design of Activity-Based Probes for Patient-Sample Imaging. *bioRxiv* **2020**, DOI: 10.1101/2020.03.07.981928.
- (11) Hilgenfeld, R. From SARS to MERS: Crystallographic Studies on Coronaviral Proteases Enable Antiviral Drug Design. *FEBS J.* **2014**, *281*, 4085–4096.
- (12) Brocklehurst, K.; Willenbrock, F.; Sauh, E. Cysteine Proteinases. In *Hydrolytic Enzymes*; Elsevier: New York, 1987; pp 39–158; DOI: 10.1016/S0167-7306(09)60016-8.
- (13) Jin, Z.; Zhao, Y.; Sun, Y.; Zhang, B.; Wang, H.; Wu, Y.; Zhu, Y.; Zhu, C.; Hu, T.; Du, X.; Duan, Y.; Yu, J.; Yang, X.; Yang, X.; Yang, K.; Liu, X.; Guddat, L. W.; Xiao, G.; Zhang, L.; Yang, H.; Rao, Z. Structural Basis for the Inhibition of SARS-CoV-2 Main Protease by Antineoplastic Drug Carmofur. *Nat. Struct. Mol. Biol.* **2020**, *27*, 529–532.
- (14) Akaji, K.; Konno, H.; Mitsui, H.; Teruya, K.; Shimamoto, Y.; Hattori, Y.; Ozaki, T.; Kusunoki, M.; Sanjoh, A. Structure-Based Design, Synthesis, and Evaluation of Peptide-Mimetic SARS 3CL Protease Inhibitors. *J. Med. Chem.* **2011**, *54*, 7962–7973.
- (15) *Schrödinger Release 2016-3: Maestro*; Schrödinger, LLC: New York, 2016.
- (16) Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical PKa Predictions. *J. Chem. Theory Comput.* **2011**, *7*, 525–537.
- (17) Case, D. A.; Cerutti, D. S.; Cheatham, T. E. I.; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Greene, D.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D.; Merz, K. M.; Monard, G.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Simmerling, C. L.; Bötello-Smith, W. M.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Xiao, L.; York, D. M.; Kollman, P. A. *Amber 2018*; University of California: San Francisco, 2018.
- (18) 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.
- (19) 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.
- (20) 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.
- (21) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N-log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (22) 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.
- (23) 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.
- (24) Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9*, 3878–3888.
- (25) Becke, A. D. Density-functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, S648–S652.
- (26) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785–789.
- (27) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion

Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.

(28) Shimato, T.; Kasahara, K.; Higo, J.; Takahashi, T. Effects of Number of Parallel Runs and Frequency of Bias-Strength Replacement in Generalized Ensemble Molecular Dynamics Simulations. *PeerJ. Phys. Chem.* **2019**, *1*, No. e4.

(29) Paasche, A.; Schirmeister, T.; Engels, B. Benchmark Study for the Cysteine–Histidine Proton Transfer Reaction in a Protein Environment: Gas Phase, COSMO, QM/MM Approaches. *J. Chem. Theory Comput.* **2013**, *9*, 1765–1777.

(30) Khrenova, M. G.; Tsirelson, V. G.; Nemukhin, A. V. Dynamical Properties of Enzyme–Substrate Complexes Disclose Substrate Specificity of the SARS-CoV-2 Main Protease as Characterized by the Electron Density Descriptors. *Phys. Chem. Chem. Phys.* **2020**, *22*, 19069–19079.

(31) Zinovjev, K. *String-Amber*; <https://github.com/kzinovjev/string-amber> (accessed Jun 24, 2020).

(32) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16*, Revision C.01; Gaussian, Inc.: Wallingford, CT, **2016**.

(33) Zinovjev, K.; Tuñón, I. Adaptive Finite Temperature String Method in Collective Variables. *J. Phys. Chem. A* **2017**, *121*, 9764–9772.

(34) Torrie, G. M.; Valleau, J. P. Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *J. Comput. Phys.* **1977**, *23*, 187–199.

(35) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. THE Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method. *J. Comput. Chem.* **1992**, *13*, 1011–1021.

(36) Fan, K.; Wei, P.; Feng, Q.; Chen, S.; Huang, C.; Ma, L.; Lai, B.; Pei, J.; Liu, Y.; Chen, J.; Lai, L. Biosynthesis, Purification, and Substrate Specificity of Severe Acute Respiratory Syndrome Coronavirus 3C-like Proteinase. *J. Biol. Chem.* **2004**, *279*, 1637–1642.

(37) 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. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 13190.

(38) Chen, H.; Wei, P.; Huang, C.; Tan, L.; Liu, Y.; Lai, L. Only One Protomer Is Active in the Dimer of SARS 3C-like Proteinase. *J. Biol. Chem.* **2006**, *281*, 13894–13898.

(39) Barrila, J.; Bacha, U.; Freire, E. Long-Range Cooperative Interactions Modulate Dimerization in SARS 3CLpro. *Biochemistry* **2006**, *45*, 14908–14916.

(40) Anand, K.; Palm, G. J.; Mesters, J. R.; Siddell, S. G.; Ziebuhr, J.; Hilgenfeld, R. Structure of Coronavirus Main Proteinase Reveals Combination of a Chymotrypsin Fold with an Extra  $\alpha$ -Helical Domain. *EMBO J.* **2002**, *21*, 3213–3224.

(41) Solowiej, J.; Thomson, J. A.; Ryan, K.; Luo, C.; He, M.; Lou, J.; Murray, B. W. Steady-State and Pre-Steady-State Kinetic Evaluation of Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) 3CLpro Cysteine Protease: Development of an Ion-Pair Model for Catalysis. *Biochemistry* **2008**, *47*, 2617–2630.

(42) Paasche, A.; Zipper, A.; Schäfer, S.; Ziebuhr, J.; Schirmeister, T.; Engels, B. Evidence for Substrate Binding-Induced Zwitterion

Formation in the Catalytic Cys-His Dyad of the SARS-CoV Main Protease. *Biochemistry* **2014**, *53*, 5930–5946.

(43) Huang, C.; Wei, P.; Fan, K.; Liu, Y.; Lai, L. 3C-like Proteinase from SARS Coronavirus Catalyzes Substrate Hydrolysis by a General Base Mechanism †. *Biochemistry* **2004**, *43*, 4568–4574.

(44) Świderek, K.; Moliner, V. Revealing the Molecular Mechanisms of Proteolysis of SARS-CoV-2 M pro by QM/MM Computational Methods. *Chem. Sci.* **2020**; DOI: 10.1039/D0SC02823A.

(45) 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.

(46) Elsässer, B.; Zauner, F. B.; Messner, J.; Soh, W. T.; Dall, E.; Brandstetter, H. Distinct Roles of Catalytic Cysteine and Histidine in the Protease and Ligase Mechanisms of Human Legumain As Revealed by DFT-Based QM/MM Simulations. *ACS Catal.* **2017**, *7*, 5585–5593.

(47) Szawelski, R. J.; Wharton, C. W. Kinetic Solvent Isotope Effects on the Deacylation of Specific Acyl-Papains. Proton Inventory Studies on the Papain-Catalysed Hydrolyses of Specific Ester Substrates: Analysis of Possible Transition State Structures. *Biochem. J.* **1981**, *199*, 681–692.

(48) Grimsley, G. R.; Scholtz, J. M.; Pace, C. N. A Summary of the Measured PK Values of the Ionizable Groups in Folded Proteins. *Protein Sci.* **2008**, *18*, 247–251.

(49) Boras, B.; Jones, R. M.; Anson, B. J.; Arenson, D.; Aschenbrenner, L.; Bakowski, M. A.; Beutler, N.; Binder, J.; Chen, E.; Eng, H.; Hammond, J.; Hoffman, R.; Kadar, E. P.; Kania, R.; Kimoto, E.; Kirkpatrick, M. G.; Lanyon, L.; Lendy, E. K.; Lillis, J. R.; Luthra, S. A.; Ma, C.; Noell, S.; Obach, R. S.; O'Brien, M. N.; O'Connor, R.; Ogilvie, K.; Owen, D.; Pettersson, M.; Reese, M. R.; Rogers, T. F.; Rossulek, M. I.; Sathish, J. G.; Steppan, C.; Ticehurst, M.; Updyke, L. W.; Zhu, Y.; Wang, J.; Chatterjee, A. K.; Mesecar, A. D.; Anderson, A. S.; Allerton, C. Discovery of a Novel Inhibitor of Coronavirus 3CL Protease as a Clinical Candidate for the Potential Treatment of COVID-19. *bioRxiv* **2020**; DOI: 10.1101/2020.09.12.293498.

(50) Hoffman, R.; Kania, R. S.; Brothers, M. A.; Davies, J. F.; Ferre, R. A.; Gajiwala, K. S.; He, M.; Hogan, R. J.; Kozminski, K.; Li, L. Y.; Lockner, J. W.; Lou, J.; Marra, M. T.; Mitchell, L. J., Jr.; Murray, B. W.; Nieman, J. A.; Noell, S.; Planken, S. P.; Rowe, T.; Ryan, K.; Smith, G. J., III; Solowiej, J. E.; Steppan, C. M.; Taggart, B. The Discovery of Ketone-Based Covalent Inhibitors of Coronavirus 3CL Proteases for the Potential Therapeutic Treatment of COVID-19. *ChemRxiv* **2020**; DOI: 10.26434/chemrxiv.12631496.v1.