

Inhibition Mechanism of SARS-CoV-2 Main Protease with Ketone-Based Inhibitors Unveiled by Multiscale Simulations

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

Departamento de Química Física, Universidad de Valencia, 46100 Burjassot (Spain)

*To whom correspondence should be addressed:

ignacio.tunon@uv.es

j.javier.ruiz@uv.es

Abstract

We present the results of combined classical and QM/MM simulations for the inhibition of SARS-CoV-2 3CL protease by a recently proposed ketone-based covalent inhibitor, PF-00835231, that is under clinical trial. In the noncovalent complex formed after binding into the active site the carbonyl group of this inhibitor is placed in the oxyanion hole formed by the NH main chain groups of residues 143 to 145. The P1-P3 groups of the inhibitor establish similar interaction with the enzyme to those of equivalent groups in the natural peptide substrate, while the hydroxymethyl moiety of the inhibitor partly mimics the interactions established by the P1' group of the peptide in the active site. Regarding the formation of the covalent complex, the reaction is initiated after the proton transfer from Cys145 to His41. Formation of the covalent hemithioacetal complex takes place by means of the nucleophilic attack of the S γ atom of Cys145 on the electron deficient carbonyl carbon atom and a proton transfer from the catalytic His41 to the carbonyl oxygen atom mediated by the hydroxyl group. Our findings can be used as a guide to propose modifications of the inhibitor in order to increase its affinity by the 3CL protease.

Keywords: 3CL protease, SARS-CoV-2, Minimum Free Energy Path, QM/MM, Ketone-based inhibitor, Binding, SARS-CoV-2, PF-00835231, PF-07304814

Introduction

Inhibition of the activity of the 3CL protease (or main protease) of SARS-CoV-2 is one of the therapeutic strategies to treat the COVID-19 pandemic. This enzyme is essential in the vital cycle of this and other related coronaviruses, being in charge of the cleavage of the long polyproteins resulting from the translation of the viral genome, in order to produce the non-structural proteins needed for virus replication.¹ The 3CL or main protease of SARS-CoV-2 is a cysteine protease that uses a Cys-His catalytic dyad to hydrolyze peptide bonds at specific positions of the polyprotein chain. The proteolysis mechanism of the 3CL protease involves two steps: i) the formation of a thiohemiacetal intermediate resulting of the attack of the S_{γ} atom of the catalytic cysteine to the carbonyl carbon atom of the target peptide bond, accompanied by the release of the N-fragment and ii) the hydrolysis of the intermediate to release the C-terminal fragment and to recover the resting state of the active site (see Figure 1a).²⁻⁵ The 3CL protease exclusively cleaves the polyprotein after a glutamine residue, a cleavage specificity not shown by any human protease.⁵ This characteristic may facilitate the development of antiviral drugs with reduced side effects.

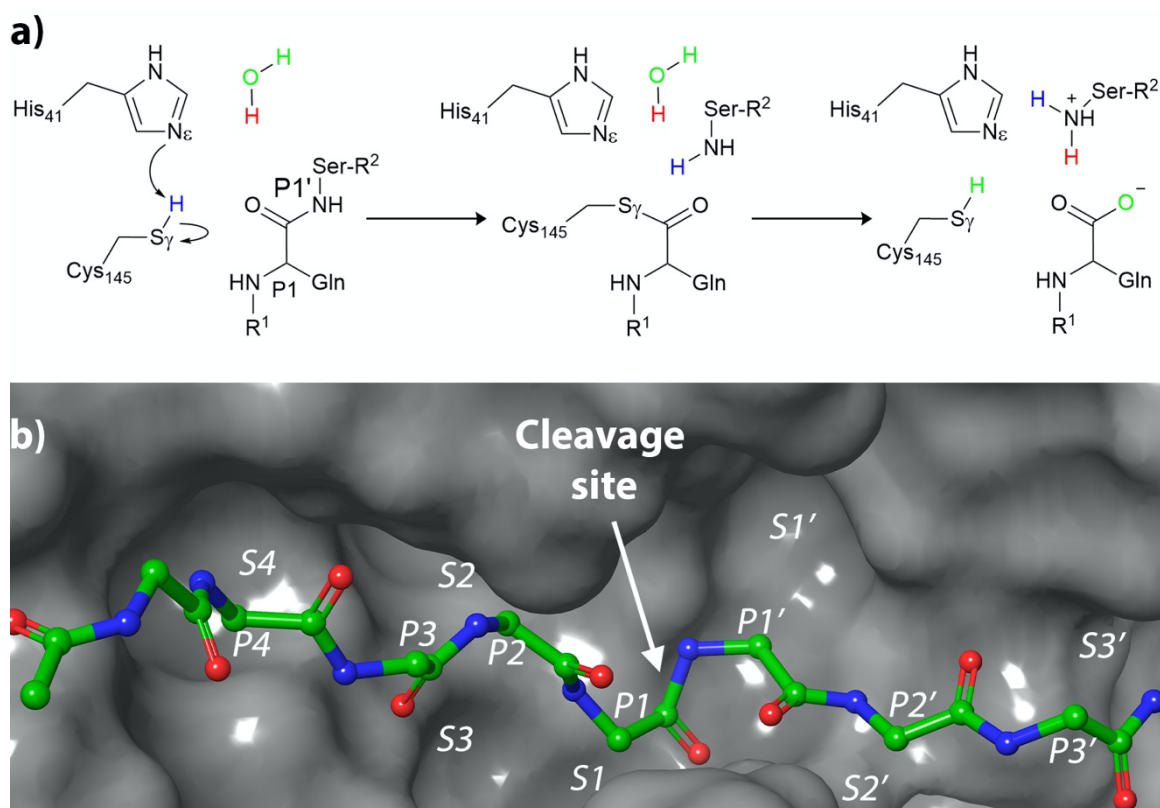
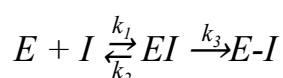


Figure 1. (a) Proposed proteolysis mechanism in SARS-CoV-2 3CL protease, Ramos et al.³ **(b)** Schematic representation of the interaction sites in the active site of proteases.

The reactive cysteine present in 3CL protease is an attractive target for the development of covalent inhibitors of this enzyme, using as warheads functional groups able to form a chemical bond with the S γ atom of this residue.⁶ Examples of warheads already used in inhibitors of the SARS-CoV-2 3CL protease are Michael acceptors,⁷ α -ketoamides,⁸ aldehydes,^{9–11} ketones¹² and others.¹³ In these inhibitors warheads are flanked by different groups that try to mimic the interactions established by the fragments of the peptide substrates placed before (Pi) and after (Pi') the cleaved peptide bond with the Si and Si' sites of the protease (see Figure 1b). These compounds first bind into the active site of the protease forming a complex (EI) governed by noncovalent interactions. After binding, the noncovalent EI complex react with the thiol group of the catalytic cysteine to yield the E-I covalent complex:



Covalent inhibitors can be reversible or irreversible depending on the relative stability of the E-I complex.¹⁴ In principle, reversibility could reduce the toxicity associated to treatments with these compounds due to off-target covalent modifications.¹⁵

One of the most promising family of inhibitors is constituted by aldehyde derivatives. At least three of these compounds presenting large inhibitory capacities have been already successfully tested in animals.^{9–11} These compounds, see Figure 2a, present a γ -lactam ring at the P1 position, taking advantage of the selectivity of this enzyme by a glutamine residue before the bond to be cleaved. Another common characteristic is the presence of a hydrophobic group at P2 position, reproducing the preference of the enzyme by a leucine residue in the natural substrate. Combination of kinetic and structural studies demonstrated that aldehydes react with the enzyme forming a reversible hemithioacetal complex, where the electrophilic carbon atom of the aldehyde group is bonded to the S γ atom of the cysteine.^{9–11} We recently performed computational simulations that showed that a water molecule is recruited in the reaction mechanism to participate in the proton transfer from the catalytic histidine to the aldehyde oxygen atom.¹⁶ Another promising inhibitor, already under clinical trial, is the ketone-based inhibitor developed by Pfizer and known as PF-00835231 (see Figure 2b) that inhibits the 3CL protease forming also a hemithioacetal complex.¹² PF-00835231 is a hydroxymethyl ketone formed after metabolization of the phosphate prodrug PF-07304814 that shows potent SARS-CoV-2 inhibition, good solubility and stability in antiviral assays, converting it in an excellent candidate for therapeutic treatment of COVID-19.¹² The hydroxymethyl group of this ketone inhibitor could potentially mimic the interactions established by a serine residue at the P1' position of the peptide substrate of the protease, interactions that are obviously absent in aldehyde inhibitors. In addition, the hydroxymethyl group could also actively participate in

the reaction mechanism, playing the role of the recruited water molecule in 3CL inhibition by aldehyde derivatives.¹⁶

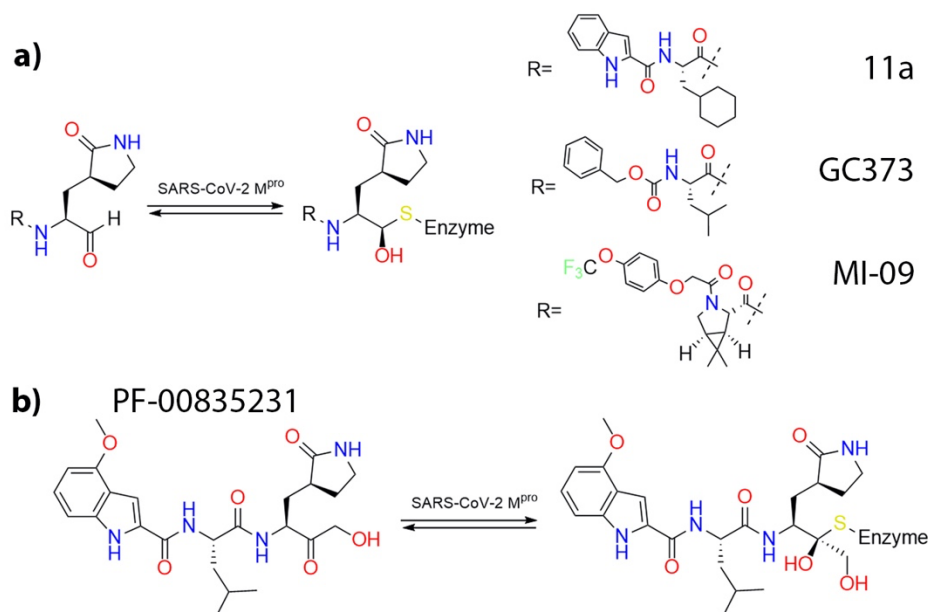


Figure 2. (a) Aldehyde inhibitors under development for treatment of COVID-19. **(b)** PF-00835231 ketone-based inhibitor.

In this work we present the results of classical and hybrid QM/MM Molecular Dynamics (MD) simulations of the ketone-based inhibitor PF-00835231 in the active site of the SARS-CoV-2 3CL protease. We have first carried out classical MD simulations of the noncovalent complex (EI) identifying the most relevant hydrogen-bond interactions established between the inhibitor and the active site residues. This interaction pattern is compared to that found for a peptidic substrate. Second, we used hybrid QM/MM methods to explore the reaction mechanism for the covalent inactivation of the enzyme, i.e., we explored the transformation of the noncovalent complex into the hemithioacetal product (the covalent E-I complex). As reported previously,^{3,4,16} the reaction process is initiated with the activation of the catalytic dyad: a proton transfer from Cys145 to His41 that results in the catalytic dyad ion pair (IP). After this proton transfer the reaction proceeds with the nucleophilic attack of the activated cysteine on the carbonyl carbon atom and the proton transfer from the catalytic histidine to the carbonyl oxygen atom. In this last step the hydroxyl group of the inhibitor acts as a proton relay, accepting the proton from His41 and giving a proton to the carbonyl oxygen atom. The structure obtained for the reaction product agrees with the x-ray structure of the 3CL protease inhibited with PF-00835231.¹² The simulations presented in this work show the atomistic details of the SARS-CoV-2 3CL

protease inhibition mechanism with a ketone-based inhibitor and could be used to improve, by rational design, future generations of antivirals.

Methodology

Classical Molecular Dynamics simulations

The PDB structure used to build the noncovalent EI complex formed was 6XHM (resolution 1.4 Å) that contains the dimeric form of the COVID-19 main protease covalently bonded to the PF-00835231 inhibitor.¹² The H-bond assignment was optimized using the protein Preparation Wizard tool from Maestro¹⁷ and the most probable protonation states at pH 7.4 were determined using its PROPKA3.0.¹⁸ For neutral histidine residues, the δ/ϵ protonation state was determined after visual inspection of the x-ray structure. The parameters for the inhibitor were obtained using the non-standard residue parameterization procedure implemented in Amber with the Antechamber program¹⁹ from the AmberTools18²⁰ package. The atomic charges for inhibitor atoms were obtained using the Restrained Electrostatic Potential (RESP) method²¹ at the HF/6-31G* level. Standard amino acids were described using the ff14SB forcefield.²² This complex was immersed in a box of TIP3P water molecules, using the tleap tool from AmberTools18,²⁰ in such a way that protein-inhibitor atoms were at least 12 Å from the limits of the simulation box. The charge of the enzyme-inhibitor system was neutralized by the addition of Na⁺ ions.

The noncovalent EI complex was prepared in the active site of protomer B of the x-ray structure. To remove bad contacts the complex was first minimized using 500 steps of steepest descent algorithm. Then the conjugate gradient method was used to find a minimum energy structure until the root mean square of the gradient was below 10^{-3} kcal·mol⁻¹·Å⁻¹. Then, the system was heated slowly from 0 to 300 K. A linear heating ramp was used along 120 ps followed by a 20 ps simulation at 300 K. To avoid dramatic changes in the atomic coordinates while heating, the position of the heavy atoms of the protein backbone were restrained using a harmonic restraint with a force constant of 20 kcal·mol⁻¹·Å⁻². Next, the system was equilibrated in the NPT ensemble (300K and 1 bar). The force constant of the positional restraints was reduced from 15 to 0 kcal·mol⁻¹·Å⁻², decreasing 3 units every 1.25 ns. Finally, the system ran free of restraints for 1.25 ns. In order to get enough sampling 5 replicas of 1 μs of the noncovalent enzyme inhibitor complex were simulated. Bond distances involving a hydrogen atom were constrained using SHAKE²³. Electrostatic interactions were treated using the Particle Mesh Ewald^{24,25} and a 10 Å cut-off radius was used for short-range interactions. During NPT simulations the Berendsen barostat and Langevin thermostat were used to control pressure and temperature respectively, while the time step was set equal to 2 fs. For all classical molecular dynamic simulations the AMBER19 GPU version of PMEMD^{26,27} was employed.

To study the free energy profile change to the rotation of the catalytic histidine side chain a classical potential of mean force (PMF) was performed using umbrella sampling.²⁸ The dihedral angle formed by the $C\alpha-C\beta-C\gamma-N\delta$ atoms of the residue was used as the distinguished coordinate. In order to obtain the starting structures for the PMF a potential energy scan was done along the reaction coordinate. Starting from -120° , initial structures were sequentially minimized under a harmonic restraint using a force constant of $100 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{rad}^{-2}$. For each minimization 2000 steps of steepest descent method were followed by the conjugate gradient method until the root mean square of the gradient was below $10^{-3} \text{ kcal}\cdot\text{mol}^{-1}\text{\AA}^{-1}$. This procedure was repeated to obtain a set of 49 structures separated by 5° from -120° to 120° . Then, a total of 52 ns of classical MD was performed for each of the 49 windows, the first 2 ns were run for relaxation followed by 50 ns of production. The total simulation time for the production stage of this PMF was longer than $2 \mu\text{s}$. The free energy profile was integrated using the Weighted Histogram Analysis Method (WHAM).²⁹

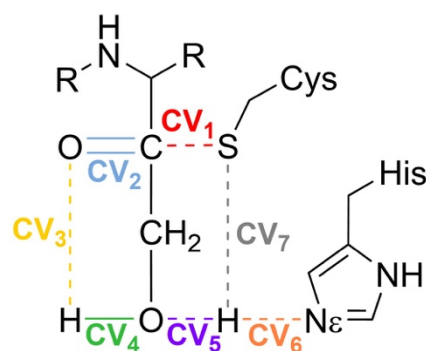
QM/MM calculations

The reaction under analysis is a multidimensional process that involves changes in several valence coordinates, preventing the use of free energy surfaces of low dimensionality. A better approach to study chemical reactions with complex reaction coordinates is Adaptive String Method (ASM).³⁰ With this method the minimum free energy pathway (MFEP) can be traced on a multidimensional free energy surface of arbitrary dimensionality and then an adequate reaction coordinate for the process can be defined as the advance along this path. It is worth to mention that in this method the number of collective variables describing the reaction coordinate can be as large as the system needs without implying an additional computational cost. Scheme 1 describe the set of 7 collective variables (CVs) used to describe the chemical transformation of the noncovalent complex to the hemithioacetal. In this case we used the distances of all the bonds whose formal order could be changed during the chemical transformation.

In our string calculations the structures of reactants and products were connected by 96 replicas of the system (or string nodes). Using QM/MM MD simulations, at every simulation step the nodes are moved according to their free energy gradient but keeping them redistributed equidistantly along the string. This procedure is continued until the string converges to the MFEP displaying a RMSD below $0.1 \text{ amu}^{1/2}\cdot\text{\AA}$ for at least 2 ps. To increase the convergence speed, replica exchange between nodes was used, with attempts done every 50 steps. After convergence, a path-CV (denoted as s) that measure the advance of the system along the MFEP is defined and used as reaction coordinate to trace the

corresponding reaction free energy profile. Then 8 ps of QM/MM simulations were run for every node and the sampled values were integrated with WHAM²⁹ to obtain the free energy profile along the path-CV. The values of the force constants employed to bias the ASM simulations were determined on-the-fly³⁰ in order to ensure a homogeneous probability density distribution of the reaction coordinate. In the string simulations the mass of the protons in flight was set equal to 2 amu and the time step was of 1 fs.

The B3LYP functional^{31,32} with a 6-31+G* basis set and D3 dispersion corrections³³ was selected to describe the QM region. This computational level has been shown to be a good methodological combination to obtain activation energies in agreement with the experimental data for the acylation of a peptide substrate³ and an inhibitor of the SARS-CoV-2 protease.¹⁶ QM/MM calculations were performed using a modified version of Amber18^{20,34} and Gaussian16³⁵ for Density Functional Theory calculations. For all the QM-MM interactions the cutoff-radius used was 15 Å. The QM region included the side chains of the catalytic dyad (His41 and Cys145), the backbone atoms of residues P1 and the hydroxymethyl P1' fragment in the PF-00835231 inhibitor. Any other atom was described at the MM level and the link atom approach was used to describe the boundary between the two subsystems.



Scheme 1. Representation of the Collective Variables (CVs) selected for the string calculations on 3CL inhibition by PF-00835231.

Results and Discussion

Binding of the PF-00835231 inhibitor. The starting point for our simulations is the x-ray structure with PDB code 6XHM. This structure corresponds to the dimeric enzyme with the two active sites inhibited by the formation of the hemithioacetal complex with the inhibitor PF-00835231. In both active sites (corresponding to chains A and B) the bond between the S γ atom of Cys145 and the carbonyl carbon atom is formed, with distances of 1.86 and 1.80

Å in chains A and B, respectively. In both cases the pose of the inhibitor is very similar (see Figure 3a) and the only significant difference appears in the rotameric state of the catalytic histidine (His41). In both active sites the N ϵ atom of His41 is close to the S γ atom of Cys145, displaying the same distance, 3.71 Å. However, in chain B the N ϵ atom is significantly closer to the inhibitor than in chain A, suggesting that this conformation would be more adequate for a subsequent proton transfer to the inhibitor. The distance from this atom to the hydroxyl oxygen atom of the inhibitor is only 2.65 Å in chain B, while in chain A the distance is increased to 3.80 Å. In the active site of this chain the N δ atom of His41 is only slightly closer, 3.67 Å. These two rotameric states, hereafter denoted as ϵ -rotamer (chain B) and δ -rotamer (chain A), are connected by means of a 180° rotation around the C β -C γ bond of His41. We thus started our simulations of the noncovalent EI complex studying the preferred rotameric state of the catalytic His41 (see Figure 3b). With this purpose, after the equilibration stage described in the Methodological section, we traced the free energy profile corresponding to His41 rotation. Results displayed in Figure 3c show that for the noncovalent EI complex the ϵ -rotamer is the preferred conformation for His41 when the PF-00835231 inhibitor is present in the active site, the ϵ -rotamer is more stable than the δ -rotamer by ca 8.0 kcal·mol⁻¹. This result agrees with the simulations performed for the noncovalent complex formed between the 3CL protease and the similar 11a aldehyde inhibitor shown in Figure 2. Also, in this case the ϵ -rotamer was found to be more stable than the δ -rotamer, but only by 3.2 kcal·mol⁻¹.¹⁶ The observed increase in the free energy difference between both rotamers is a first indication of the role played by the hydroxymethyl group of the PF-00835231 inhibitor, which seems to make more favorable contacts with the His41 residue in the ϵ -rotamer, in agreement with the His41-hydroxyl group distances observed in the 6XHM x-ray structure discussed above. It must be also noticed that the larger stability observed for the ϵ -rotamer in the noncovalent complexes, agrees with the more frequent presence of this conformer over the δ -one in the x-ray structures of the hemithioacetal complexes formed between the SARS-CoV-2 3CL protease and other inhibitors.¹⁶

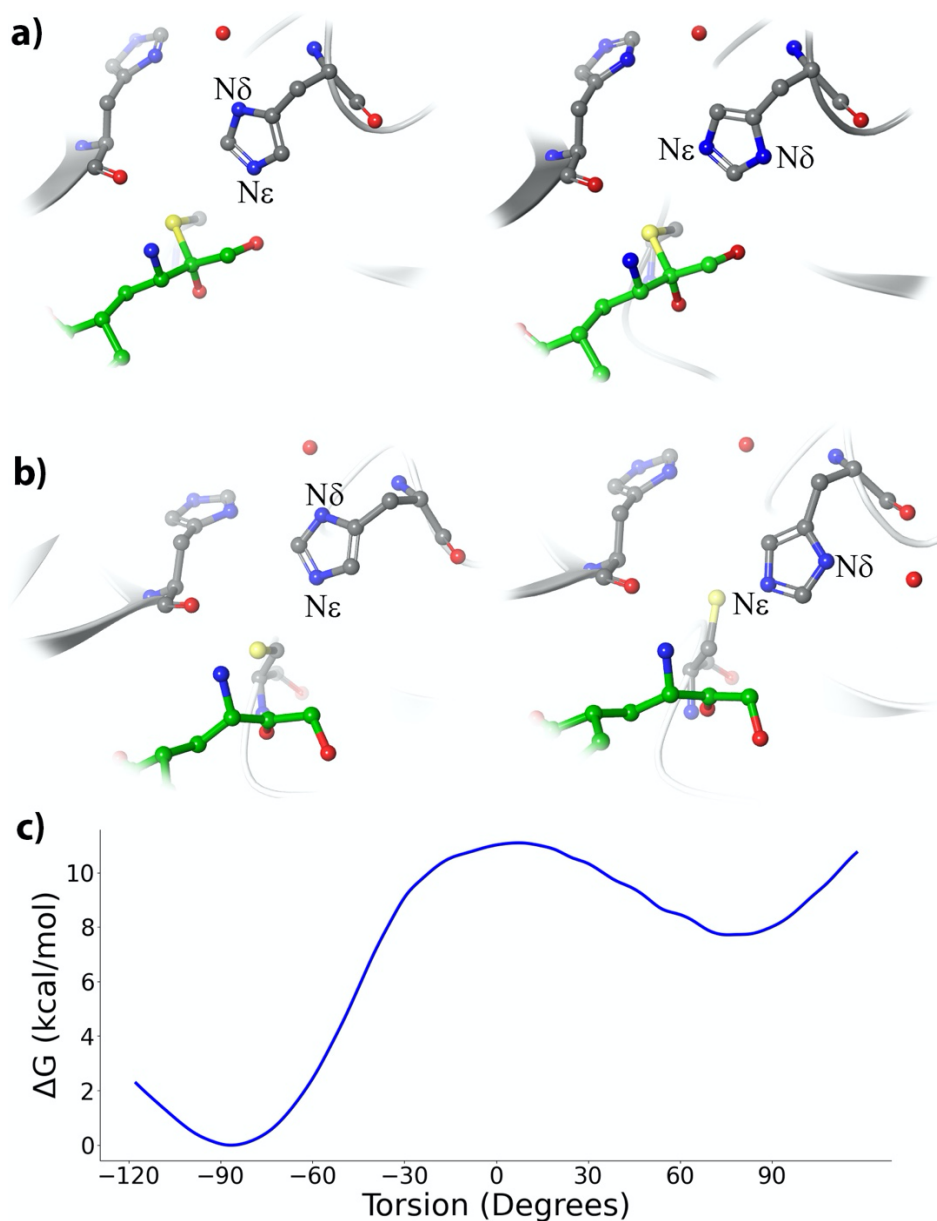


Figure 3. His41 rotameric state. **(3a)** Covalent product of PF-00835231 (carbon atoms in green) with SARS-CoV-2 3CL protease in the ϵ -rotameric state found in chain B of 6XHM (left) and in the δ -rotameric found in chain A of 6XHM (right). **(3b)** Representation of the noncovalent EI complex in the ϵ -rotameric state (left) and in the δ -rotameric state (right). **(3c)** Free energy profile for the rotation of His41 along the C_{γ} - C_{β} bond, the ϵ - and δ -rotameric states appear a dihedral angle values of approximately -90 and 90 degrees.

Once determined the preferred conformation for the catalytic histidine, we analyzed the binding pose of the inhibitor and the interactions established with the enzyme in the noncovalent EI complex by means of MD simulations (5 replicas of 1 μ s each). These MD simulations were stable in all cases (see RMSD time evolutions in Figure S1), showing a

binding pose consistent with the x-ray structure of the hemithioacetal complex (see Figure 4a). P1, P2 and P3 sites of the inhibitor present an interaction pattern similar to that of a peptide substrate with sequence -Val-Leu-Gln|Ser- (where the vertical line indicates the scissile bond)³ and also similar to those of other peptidyl inhibitors, such as Michael acceptors³⁶ and aldehyde derivatives.¹⁶ Figure 4b compares the fraction of hydrogen bonds established between the peptide substrate or the inhibitor and enzymatic residues during the MD simulations. The P1 γ -lactam ring is a recurrent group present in many inhibitors of SARS-CoV-2 and SARS-CoV 3CL proteases, exploiting the selectivity of these enzymes for substrates that present a glutamine residue at this position.⁵ This P1 group establishes hydrogen bonds with His163, Glu166 and Phe140. The isobutyl hydrocarbon group at the P2 position, identical to the side chain of leucine that is the preferred P2 residue for the 3CL protease, stacks with the His41 imidazole ring, interacting also with other nearby residues, such as His164, Met165 and Gln189. Finally, the P3 group of the PF-00835231 inhibitor is exposed to the solvent and stabilized by hydrogen bond interactions with main chain atoms of Met165, Glu166 and Glu189. The position of the carbonyl oxygen atom of the inhibitor is stabilized by means of hydrogen bond interactions with the main chain NH groups of Cys145 (2.3 ± 0.3 Å), Ser144 (2.8 ± 0.4 Å) and Gly143 (2.5 ± 0.3 Å), as seen in Figure 4a. These interactions are also observed in the x-ray structure of the hemithioacetal product.

One of the main novelties of the PF-00835231 inhibitor is the inclusion of a hydroxymethyl group at the P1' position. This group resembles a serine residue, which is one of the preferences of 3CL proteases at this position. In fact, comparison of the hydrogen bond interactions established by Ser-P1' in the peptide substrate and the hydroxymethyl group in the PF-00835231 inhibitor shows that this last is able to recover a fraction of the interactions established by the peptide. The hydroxyl group of the inhibitor mimics the interactions made by the serine side chain, in particular with the catalytic dyad, His41 and Cys145. The main difference is due to the larger conformational flexibility of the P1' group of the inhibitor, reflected in the fact that the hydroxyl group can also establish interactions with other residues of the active site, mainly with Gly143 and Asn142. In the peptide inhibitor, the presence of P2' and subsequent primed sites reduces the flexibility of the P1' group favoring the formation of more stable interactions with the catalytic dyad.

The formation of a covalent bond between the enzyme and the inhibitor requires the activation of the S_γ atom of Cys145 by means of a proton transfer from Cys145 to His41 and the subsequent nucleophilic attack of this atom on the electron deficient carbonyl carbon atom of the inhibitor, forming the hemithioacetal product (the covalent E-I complex).^{3,36} We thus monitored the distances of the Cys145 S_γ atom with the His41 N_ε atom and with the carbonyl carbon atom of the inhibitor (C). The probability distributions of these distances

are shown in Figure 4c and both of them display a bimodal distribution that can be attributed to the presence of *trans* and *gauche* conformers of the Cys145 side chain, conformers that have been already observed in the x-ray structure of the orthologue enzyme of SARS-CoV.³⁷ The S_{γ} -C and S_{γ} -N ϵ distributions are peaked at 3.4/5.4 and 3.3/5.2 Å, respectively; showing that a significant fraction of the noncovalent EI complex conformations observed during the MD simulations are ready to proceed to the formation of the hemithioacetal product.

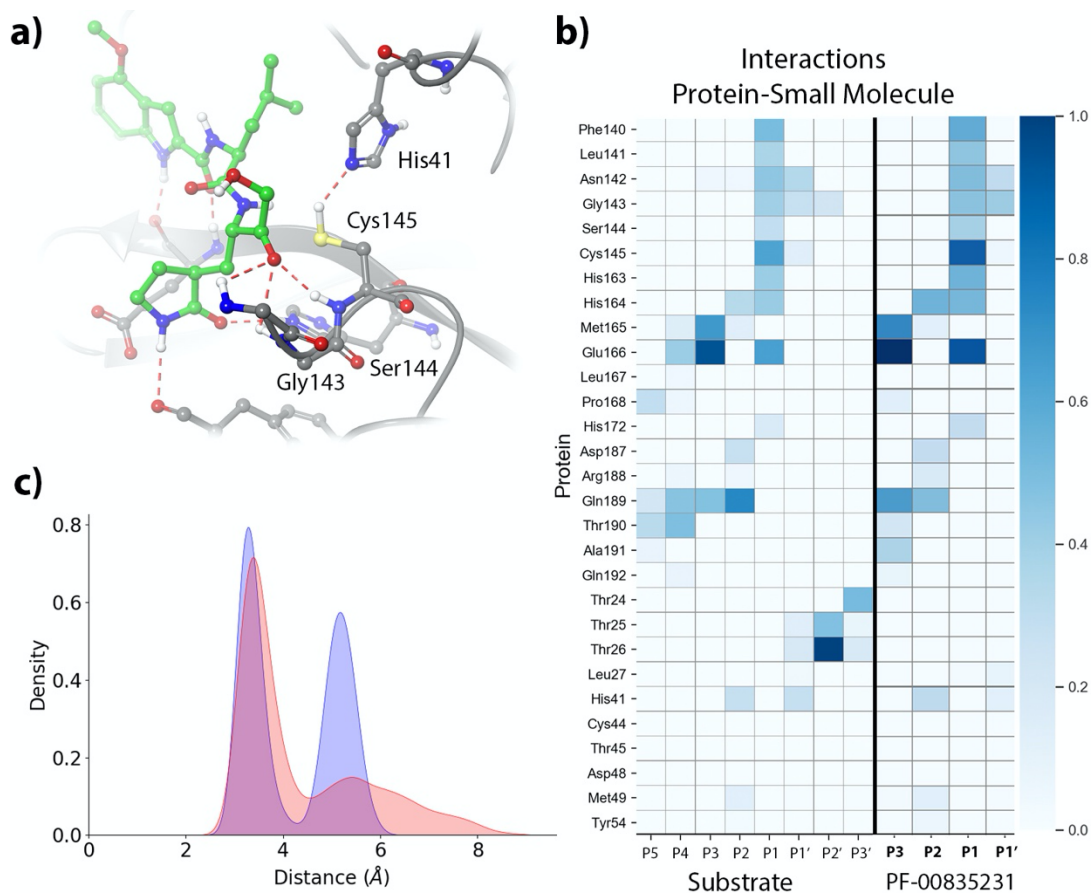


Figure 4. Noncovalent complex formed between PF-00835231 and the 3CL protease of SARS-CoV-2. **(4a)** Binding pose of the inhibitor in the active site of the protease, showing the location of the catalytic dyad and the oxyanion hole. Note that the carbonyl oxygen is accommodated into the oxyanion hole. **(4b)** Fraction of hydrogen bond contacts between residues of PF-00835231 and a peptide substrate³ and those of the protease. A hydrogen bond contact is counted when the donor-acceptor distance is < 3.8 Å and the hydrogen bond angle is $> 120^{\circ}$. **(4c)** Probability densities of the distances from the Cys145- S_{γ} atom to the C carbon atom of the substrate, in blue and from the N ϵ atom of His41 to the aldehyde oxygen atom, in red.

Formation of the covalent hemithioacetal product.

In order to explore the reaction mechanism for the formation of the covalent E-I complex from the noncovalent one, we traced the corresponding MFEP using the string method at the B3LYPD3/MM level. According to previous studies on the 3CL protease activity with peptide substrates^{3,4} and inhibitors^{16,36,38,39} formation of a covalent bond with the S γ atom of Cys145 requires its activation by means of the proton transfer to His41. After formation of the covalent bond with the inhibitor, the reaction must be completed with the proton transfer from His41 to the inhibitor. In our previous study with aldehyde derivatives, we found that this proton transfer takes place mediated by a water molecule that occupies the same position than the hydroxyl group in the PF-00835231 inhibitor, suggesting the participation of this group as a proton relay during the proton transfer from His41 to the carbonyl oxygen atom.¹⁶

Our string simulations converged to a two-steps reaction mechanism, as shown in Figures 5a and 5b. In the first step a metastable ion pair (IP) catalytic dyad is formed after the proton transfer from Cys145 to His41. The IP is found to be 10.7 kcal·mol⁻¹ above the noncovalent EI complex (see Figure 5a). This free energy difference is very close to the values reported for other inhibitors. In particular, formation of the IP in presence of the 11a aldehyde¹⁶ requires 9.3 kcal·mol⁻¹ and 10.3 in presence of a Michael acceptor.³⁶ This IP, represented in Figure 5c, is found in a free energy plateau, with a very low free energy barrier for the reverse process. At the IP the proton transfer from Cys145 to His41 has been completed (the N ϵ -H and S γ H distances are 1.04 and 2.61 Å, respectively) and His41 is now oriented towards the hydroxyl group of the inhibitor, forming a hydrogen bond with a N ϵ H-O distance of 1.93 Å. This first proton transfer from Cys145 to His41 is accompanied by the approach of the activated S γ atom to the carbonyl carbon distance: the distance between these two atoms being reduced from 3.09 to 2.25 Å when going from the noncovalent EI complex to the IP. From this state, the reaction proceed by means of the nucleophilic attack of the S γ atom of Cys145 to the carbonyl carbon atom of the inhibitor and the proton transfer from His41 to the hydroxyl group of the inhibitor and from this to the carbonyl oxygen atom (see evolution of the CVs in Figure 5b). The rate-limiting TS for the formation of the hemithioacetal is presented in Figure 5d. At the TS the enzyme-inhibitor covalent bond is almost completely formed, presenting a distance of 1.99 Å and the carbonyl double bond has been elongated from 1.22 (EI complex) to 1.35 Å. The two proton transfers events are not very advanced at the TS. The N ϵ -H distance has been slightly lengthened to 1.12 Å, while the distance of this proton to the hydroxyl oxygen atom of the inhibitor is 1.47 Å. Simultaneously, the proton transfer from the hydroxyl group to the carbonyl oxygen atom is also found at a slightly more advanced stage, being the distances of the proton to the donor and acceptor atoms of 1.16 and 1.39 Å, respectively. From this TS the reaction is

completed to yield the hemithioacetal product (see Figure 5e) where the two proton transfers have been completed. In our calculations, the covalent S_γ-C bond presents a distance of 1.88 Å in the product state, in excellent agreement with the x-ray observation (1.80/1.86 Å). The His41 residue remains at a hydrogen bond distance of the hydroxyl group of the inhibitor (the N_ε-O distance being 2.53 Å, also close to the x-ray value which is 2.65 Å). In general, the configuration obtained for the covalent E-I complex agrees very well with the experimental x-ray geometry observed in the 6XHM structure. The overlap between the two structures, calculated and x-ray, is shown in Figure 5e.

The free energy profile for the covalent inactivation of the 3CL protease with PF-00835231 can be compared with experimental and theoretical results obtained for the inhibition with aldehyde derivatives 11a and GC-373 (see Figure 2a). Recent analysis indicate that the PF-00835231 inhibitor has similar or higher potency against SARS-CoV-2 in human A549 cells than GC-376 (the prodrug of GC-373).⁴⁰ Our calculations predict that the formation of the hemithioacetal product from the noncovalent complex is exergonic, with a reaction free energy of -6.2 kcal·mol⁻¹. This value is larger (in absolute value) than the reaction free energy obtained for SARS-CoV-2 3-CL enzyme inhibition with the aldehyde derivative 11a using the same computational scheme, -2.8 kcal·mol⁻¹,¹⁶ but smaller than the value predicted for a Michael acceptor, -15.0 kcal·mol⁻¹.³⁶ These two inhibitors, aldehyde and Michael acceptor, are examples of covalent reversible¹⁰ and irreversible⁷ inhibitors, respectively. Thus, our calculations predict that the ketone-based PF-00835231 inhibitor would still be a covalent reversible inhibitor but with more irreversible character than aldehyde inhibitors. The predicted activation free energy for the process is 19.7 kcal·mol⁻¹. Unfortunately, there are not experimental measurements of the first order inactivation rate constant for this inhibitor, but we can again compare our results with the estimations obtained for the inhibition of the SARS-CoV-2 3CL protease by aldehyde derivatives. The rate constant for the inhibition of the protease with the aldehyde inhibitor GC373 has been measured to be 2.45·10⁻³ s⁻¹ at 30°C,⁴¹ a value that, according to Transition State Theory, can be translated into an activation free energy of 21.1 kcal·mol⁻¹. For 11a our simulations predicted an activation free energy of 18.5 kcal·mol⁻¹,¹⁶ a value similar to that obtained here for the PF-00835231 inhibitor. Thus, according to the results here presented, the PF-00835231 inhibitor would present similar kinetic properties for the covalent inhibition of the SARS-CoV-2 3CL enzyme than the aforementioned aldehyde-based inhibitors. However, as stressed above, the formation of the hemithioacetal would be more favorable from the thermodynamic point of view in the case of the ketone-based inhibitor.

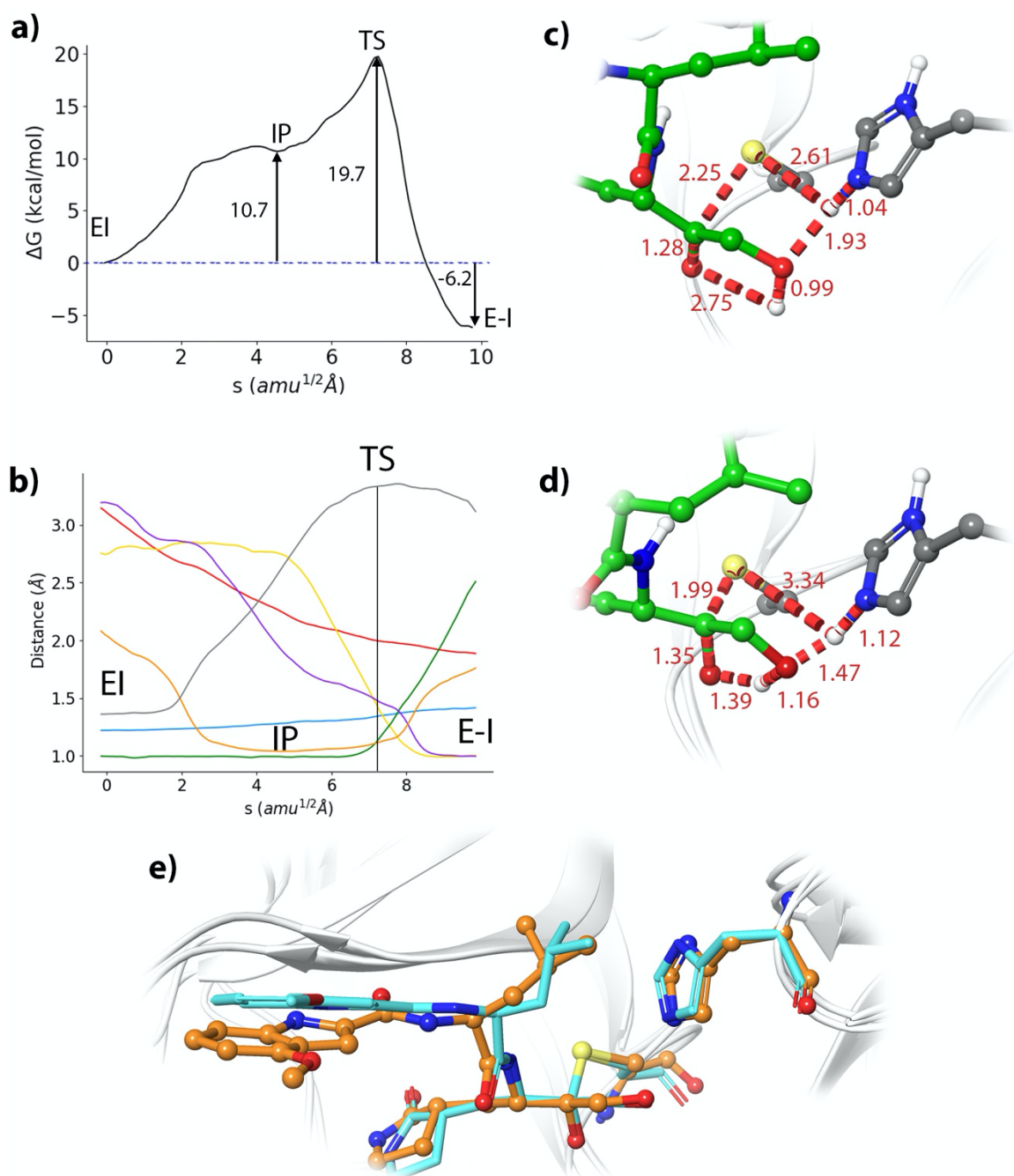


Figure 5. Formation of the (S)-hemithioacetal product. **(5a)** B3LYPD3/6-31+G*/MM free energy profile along the path-CV for the formation of the covalent E-I complex from the EI one. **(5b)** Evolution of the selected CVs along the MFEP. The color code corresponds to Scheme 1. **(5c)** Representation of the IP state. The values of the distances correspond (in \AA) to the coordinates of the MFEP where the IP is located. **(5d)** Representation of the rate limiting TS. **(5e)** Overlap of the product structure (balls & sticks with carbon atoms in orange) with the x-ray structure 6XHM containing a hydroxymethylketone inhibitor PF-00835231 (licorice with carbon atoms in light blue).

Conclusions

We have used a combination of classical and hybrid QM/MM Molecular Dynamics simulations to explore the covalent inhibition mechanism of the SARS-CoV-2 3CL protease by a ketone-based inhibitor that is currently under clinical trial, PF-00835231. We first explored the binding mode and interactions established by the inhibitor starting from the x-ray structure of the hemithioacetal complex. We determined the preferred rotameric state of the catalytic His41, because the x-ray structure shows a different conformation for this residue in each of the active sites of the dimer. The conformation determined as the most stable is that presenting the N ϵ atom of His41 close to the inhibitor warhead, in agreement with the preferred conformation observed in previous studies.

Our simulations emphasize the role played by the hydroxymethyl group at the P1' position of this inhibitor. First, this group is able to establish hydrogen bond interactions with the active site residues, partly recovering some of the interactions established by the residue serine at the P1' position of the peptide substrate. Second, this group plays also an active role during the formation of the hemithioacetal complex, mediating the proton transfer from the catalytic histidine to the carbonyl oxygen atom. As in preceding cases, the formation of the covalent complex is initiated by means of a proton transfer from Cys145 to His41 to form an ion pair. Once the catalytic dyad is activated, the process continues with the nucleophilic attack of the cysteine sulphur atom on the electrophilic carbonyl carbon atom of the inhibitor and the proton transfer from His41 to the carbonyl oxygen atom. The rate limiting TS shows a short carbon-sulphur bond distance, while the proton transfers from the catalytic histidine to the hydroxyl group of the inhibitor and from this group to the carbonyl oxygen atom are found at an earlier stage. The activation free energy associated to this TS is similar to that found for aldehyde inhibitors, while formation of the hemithioacetal complex is more exergonic for the ketone-based inhibitor. Our results can be useful in order to achieve new inhibitors of 3CL enzyme of SARS-CoV-2 based in ketone derivatives.

Acknowledgements

The authors acknowledge financial support from Consellería de Innovación, Universidades, Ciencia y Sociedad Digital, Generalitat Valenciana (GVCOV19/Decreto180/2020). We want to acknowledge Barcelona Supercomputing Center (BSC) for awarding us access to MareNostrum and the staff from BSC for the technical support (Project QSB-2021-1-0006). We also acknowledge the use of the Tirant supercomputer at the Universitat de València (financed by the FEDER funds for Scientific Infrastructures;

IDIFEDER-2018-063) and the support of Alejandro Soriano from Servei d'Informàtica from the Universitat de València.

References

- (1) Bangham, C. R. M. The Immune Control and Cell-to-Cell Spread of Human T-Lymphotropic Virus Type 1. *J. Gen. Virol.* **2003**, *84* (12), 3177–3189. <https://doi.org/https://doi.org/10.1099/vir.0.19334-0>.
- (2) 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. <https://doi.org/10.1021/bi702107v>.
- (3) Ramos-Guzmán, C. A.; Ruiz-Pernía, J. J.; Tuñón, I. Unraveling the SARS-CoV-2 Main Protease Mechanism Using Multiscale Methods. *ACS Catal.* **2020**, *10* (21), 12544–12554. <https://doi.org/10.1021/acscatal.0c03420>.
- (4) Świderek, K.; Moliner, V. Revealing the Molecular Mechanisms of Proteolysis of SARS-CoV-2 M pro by QM/MM Computational Methods. *Chem. Sci.* **2020**, *11* (39), 10626–10630. <https://doi.org/10.1039/D0SC02823A>.
- (5) Hilgenfeld, R. From SARS to MERS: Crystallographic Studies on Coronaviral Proteases Enable Antiviral Drug Design. *FEBS J.* **2014**, *281*, 4085–4096. <https://doi.org/10.1111/febs.12936>.
- (6) Steuten, K.; Kim, H.; Widen, J. C.; Babin, B. M.; Onguka, O.; Lovell, S.; Bolgi, O.; Cerikan, B.; Neufeldt, C. J.; Cortese, M.; Muir, R. K.; Bennett, J. M.; Geiss-Friedlander, R.; Peters, C.; Bartenschlager, R.; Bogyo, M. Challenges for Targeting SARS-CoV-2 Proteases as a Therapeutic Strategy for COVID-19. *ACS Infect. Dis.* **2021**, *acsinfecdis.0c00815*. <https://doi.org/10.1021/acsinfectdis.0c00815>.
- (7) 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. X.; You, T.; Liu, X. X. X.; Yang, X. X. X.; Bai, F.; Liu, H.; Liu, X. X. X. X.; Guddat, L. W.; Xu, W.; Xiao, G.; Qin, C.; Shi, Z.; Jiang, H.; Rao, Z.; Yang, H. Structure of Mpro from SARS-CoV-2 and Discovery of Its Inhibitors. *Nature* **2020**, *582* (7811), 289–293. <https://doi.org/10.1038/s41586-020-2223-y>.
- (8) 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 (80-.)*. **2020**, *368* (March), 409–412. <https://doi.org/10.1126/science.abb3405>.
- (9) 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 (80-.)*. **2020**, *368* (6497), 1331 LP – 1335. <https://doi.org/10.1126/science.abb4489>.
- (10) Vuong, W.; Khan, M. B.; Fischer, C.; Arutyunova, E.; Lamer, T.; Shields, J.; Saffran, H. A.; McKay, R. T.; van Belkum, M. J.; Joyce, M. A.; Young, H. S.; Tyrrell, D. L.; Vederas,

- J. C.; Lemieux, M. J. Feline Coronavirus Drug Inhibits the Main Protease of SARS-CoV-2 and Blocks Virus Replication. *Nat. Commun.* **2020**, *11* (1), 4282. <https://doi.org/10.1038/s41467-020-18096-2>.
- (11) Qiao, J.; Li, Y.-S.; Zeng, R.; Liu, F.-L.; Luo, R.-H.; Huang, C.; Wang, Y.-F.; Zhang, J.; Quan, B.; Shen, C.; Mao, X.; Liu, X.; Sun, W.; Yang, W.; Ni, X.; Wang, K.; Xu, L.; Duan, Z.-L.; Zou, Q.-C.; Zhang, H.-L.; Qu, W.; Long, Y.-H.-P.; Li, M.-H.; Yang, R.-C.; Liu, X.; You, J.; Zhou, Y.; Yao, R.; Li, W.-P.; Liu, J.-M.; Chen, P.; Liu, Y.; Lin, G.-F.; Yang, X.; Zou, J.; Li, L.; Hu, Y.; Lu, G.-W.; Li, W.-M.; Wei, Y.-Q.; Zheng, Y.-T.; Lei, J.; Yang, S. SARS-CoV-2 M pro Inhibitors with Antiviral Activity in a Transgenic Mouse Model. *Science* (80-.). **2021**, eabf1611. <https://doi.org/10.1126/science.abf1611>.
- (12) Hoffman, R. L.; 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.; Murray, B. W.; Nieman, J. A.; Noell, S.; Planken, S. P.; Rowe, T.; Ryan, K.; Smith, G. J.; Solowiej, J. E.; Stepan, C. M.; Taggart, B. Discovery of Ketone-Based Covalent Inhibitors of Coronavirus 3CL Proteases for the Potential Therapeutic Treatment of COVID-19. *J. Med. Chem.* **2020**, *63* (21), 12725–12747. <https://doi.org/10.1021/acs.jmedchem.0c01063>.
- (13) Menéndez, C. A.; Byléhn, F.; Perez-Lemus, G. R.; Alvarado, W.; de Pablo, J. J. Molecular Characterization of Ebselen Binding Activity to SARS-CoV-2 Main Protease. *Sci. Adv.* **2020**, *6* (37), eabd0345. <https://doi.org/10.1126/sciadv.abd0345>.
- (14) Powers, J. C.; Asgian, J. L.; Ekici, Ö. D.; James, K. E. Irreversible Inhibitors of Serine, Cysteine, and Threonine Proteases. *Chem. Rev.* **2002**, *102* (12), 4639–4750. <https://doi.org/10.1021/cr010182v>.
- (15) Drag, M.; Salvesen, G. S. Emerging Principles in Protease-Based Drug Discovery. *Nat. Rev. Drug Discov.* **2010**, *9* (9), 690–701. <https://doi.org/10.1038/nrd3053>.
- (16) (a) Ramos-Guzmán, C. A.; Ruiz-Pernía, J. J.; Tuñón, I. Multiscale Simulations of SARS-CoV 2 3CL Protease Inhibition with Aldehyde Derivatives. Role of Protein and Inhibitor Conformational Changes in the Reaction Mechanism. *ACS Catal.* **2021**. <https://doi.org/10.1021/acscatal.0c05522>. (b) Ramos-Guzmán, C. A.; Ruiz-Pernía, J. J. Tuñón, I. Multiscale Simulations of SARS-CoV-2 3CL Protease Inhibition with Aldehyde Derivatives. Role of Protein and Inhibitor Conformational Dynamics in the Reaction Mechanism. *ChemRxiv. Preprint.* **2020**. <https://doi.org/10.26434/chemrxiv.13340942.v1>
- (17) Schrödinger Release 2021-1: Maestro, Schrödinger, LLC, New York, NY. 2021.
- (18) 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* (2), 525–537. <https://doi.org/10.1021/ct100578z>.
- (19) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graph. Model.* **2006**, *25* (2), 247–260. <https://doi.org/10.1016/j.jmgm.2005.12.005>.
- (20) 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.; Botello-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*.
- (21) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. *J. Phys. Chem.* **1993**, *97* (40), 10269–10280. <https://doi.org/10.1021/j100142a004>.
- (22) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB. *J. Chem. Theory Comput.* **2015**, *11* (8), 3696–3713. <https://doi.org/10.1021/acs.jctc.5b00255>.
- (23) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. . Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* **1977**, *23* (3), 327–341. [https://doi.org/10.1016/0021-9991\(77\)90098-5](https://doi.org/10.1016/0021-9991(77)90098-5).
- (24) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An $N \cdot \log(N)$ Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98* (12), 10089–10092. <https://doi.org/10.1063/1.464397>.
- (25) 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* (19), 8577–8593. <https://doi.org/10.1063/1.470117>.
- (26) 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* (2), 374–380. <https://doi.org/10.1016/j.cpc.2012.09.022>.
- (27) 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* (9), 3878–3888. <https://doi.org/10.1021/ct400314y>.
- (28) Torrie, G. M.; Valleau, J. P. Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *J. Comput. Phys.* **1977**, *23*, 187–199. [https://doi.org/10.1016/0021-9991\(77\)90121-8](https://doi.org/10.1016/0021-9991(77)90121-8).
- (29) 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* (8), 1011–1021. <https://doi.org/10.1002/jcc.540130812>.
- (30) Zinovjev, K.; Tuñón, I. Adaptive Finite Temperature String Method in Collective Variables. *J. Phys. Chem. A* **2017**, *121* (51), 9764–9772. <https://doi.org/10.1021/acs.jpca.7b10842>.
- (31) Becke, A. D. Density-functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652. <https://doi.org/10.1063/1.464913>.
- (32) 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* **1988**, *37*, 785–789. <https://doi.org/10.1103/PhysRevB.37.785>.

- (33) 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. <https://doi.org/10.1063/1.3382344>.
- (34) Zinovjev, K. String-Amber <https://github.com/kzinovjev/string-amber> (accessed Jun 24, 2020).
- (35) 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 Jr., J. A.; 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. 2016.
- (36) Ramos-Guzmán, C. A.; Ruiz-Pernía, J. J.; Tuñón, I. A Microscopic Description of SARS-CoV-2 Main Protease Inhibition with Michael Acceptors. Strategies for Improving Inhibitor Design. *Chem. Sci.* **2021**. <https://doi.org/10.1039/D0SC04978F>.
- (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* (23), 13190–13195. <https://doi.org/10.1073/pnas.1835675100>.
- (38) Mondal, D.; Warshel, A. Exploring the Mechanism of Covalent Inhibition: Simulating the Binding Free Energy of α -Ketoamide Inhibitors of the Main Protease of SARS-CoV-2. *Biochemistry* **2020**, *59* (48), 4601–4608. <https://doi.org/10.1021/acs.biochem.0c00782>.
- (39) Arafet, K.; Serrano-Aparicio, N.; Lodola, A.; Mulholland, A. J.; González, F. V.; Świderek, K.; Moliner, V. Mechanism of Inhibition of SARS-CoV-2 M pro by N3 Peptidyl Michael Acceptor Explained by QM/MM Simulations and Design of New Derivatives with Tunable Chemical Reactivity. *Chem. Sci.* **2021**, *12* (4), 1433–1444. <https://doi.org/10.1039/D0SC06195F>.
- (40) de Vries, M.; Mohamed, A. S.; Prescott, R. A.; Valero-Jimenez, A. M.; Desvignes, L.; O'Connor, R.; Stepan, C.; Devlin, J. C.; Ivanova, E.; Herrera, A.; Schinlever, A.; Loose, P.; Ruggles, K.; Koralov, S. B.; Anderson, A. S.; Binder, J.; Dittmann, M. A Comparative Analysis of SARS-CoV-2 Antivirals Characterizes 3CL pro Inhibitor PF-00835231 as a Potential New Treatment for COVID-19. *J. Virol.* **2021**. <https://doi.org/10.1128/JVI.01819-20>.
- (41) Ma, C.; Sacco, M. D.; Hurst, B.; Townsend, J. A.; Hu, Y.; Szeto, T.; Zhang, X.; Tarbet, B.; Marty, M. T.; Chen, Y.; Wang, J. Boceprevir, GC-376, and Calpain Inhibitors II, XII

Inhibit SARS-CoV-2 Viral Replication by Targeting the Viral Main Protease. *Cell Res.* **2020**, *30* (8), 678–692. <https://doi.org/10.1038/s41422-020-0356-z>.