Multiscale Simulations of SARS-CoV-2 3CL Protease Inhibition with Aldehyde Derivatives. Role of Protein and Inhibitor Conformational Dynamics in the Reaction Mechanism.

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 here investigate the mechanism of SARS-CoV-2 3CL protease inhibition by one of the most promising families of inhibitors, those containing an aldehyde group as warhead. These compounds are covalent inhibitors that inactivate the protease forming a stable hemithioacetal complex. Inhibitor 11a is a potent inhibitor that has been already tested in vitro and in animals. Using a combination of classical and QM/MM simulations we determined the binding mode of the inhibitor into the active site and the preferred rotameric state of the catalytic histidine. In the noncovalent complex the aldehyde group is accommodated into the oxyanion hole formed by the NH main chain groups of residues 143 to 145. In this pose, P1-P3 groups of the inhibitor mimic the interactions established by the natural peptide substrate. The reaction is initiated with the formation of the catalytic dyad ion pair after a proton transfer from Cys145 to His41. From this activated state, covalent inhibition proceeds with the nucleophilic attack of the deprotonated $S\gamma$ atom of Cys145 to the aldehyde carbon atom and a water mediated proton transfer from the $N\varepsilon$ atom of His41 to the aldehyde oxygen atom. Our proposed reaction transition state structure is validated by comparison with x-ray data of recently reported inhibitors, while the activation free energy obtained from our simulations agrees with the experimentally derived value, supporting the validity of our findings. Our study stresses the interplay between the conformational dynamics of the inhibitor and the protein with the inhibition mechanism and the importance of including conformational diversity for accurate predictions about the inhibition of the main protease of SARS-CoV-2. The conclusions derived from our work can also be used to rationalize the behavior of other recently proposed inhibitor compounds, including aldehydes and ketones with high inhibitory potency.

Introduction

The impact and rapid expansion of COVID-19, caused by the coronavirus SARS-CoV-2, has urged the research to find therapeutic remedies.[1] There are, in principle, two main strategies to be used against this disease: the development of vaccines and antiviral drugs. While vaccines have the advantage of preventing the disease, antivirals could be beneficial not only to fight against SARS-CoV-2 but also for other related coronaviruses, including those that could infect human beings in a near future. While some antiviral drugs developed to fight against other viruses, such as remdesivir, lopinavir or favipiravir, have been already tested to treat COVID-19, there are no clear evidences of their efficiency.[2,3] Therefore, there is an urgent need to develop new antiviral drugs that are both effective and selective for SARS-CoV-2.

One of the strategies for the development of antiviral drugs is to target the inhibition of one of those enzymes that are essential for the vital cycle of the virus. When SARS-CoV-2 infects a cell utilizes the transcription machinery to translate the viral genome into two long polyproteines. These long chains must be cleaved at specific sites to produce the non-structural proteins that the virus needs for replication and transcription of its genome. This key function is performed by two proteases: the 3C-like (3CL) or main protease and the Papain-like (PL) protease. The former cleaves the polyprotein at 11 positions targeting for Gln-(Ser/Ala/Gly) peptide bond, a sequence preference that is not used by human proteases.[4] This characteristic makes of the 3CL protease and excellent candidate as a drug target, because those compounds designed to bind and block the active site of this protein have less chances to interact with proteases of the host.[5] The 3CL protease is a cysteine protease and its active site is formed by a catalytic dyad, His41 and Cys145, in charge of the hydrolysis of peptide bonds. This process takes place in two main steps:[6,7] acylation and de-acylation. During the acylation step a peptide fragment is released while the other forms an acyl-enzyme complex by means of a covalent bond with the $S\gamma$ atom of the catalytic cysteine. During de-acylation the second fragment is released by the action of a water molecule, recovering the enzyme for a new catalytic cycle.

Up to know, several families of inhibitors have been proposed and tested in vitro against the 3CL protease of SARS-CoV-2, including Michael acceptors,[8] α-ketoamides,[9] aldehyde derivatives[10,11] and ketones.[12] These compounds first bind into the active site of the protease forming a noncovalent complex (EI) and then react with the thiol group of the catalytic cysteine to form a stable covalent acyl-enzyme complex (E-I), see Figure 1a. The design and improvement of these compounds is usually guided by the information provided

by the x-ray structure of the covalent complex. However, as we will stress in this work, this information can be incomplete if the inhibitors and or the enzyme present conformational diversity. This diversity could play a critical role in a proper understanding of the general inhibition process.

The aldehyde group seems a promising warhead for the development of antiviral drugs to fight COVID-19, as it can react with the catalytic cysteine to form hemithioacetal complexes (the E-I complex).[13] Several aldehyde derivatives have been shown to have large inhibitory properties against the 3CL protease of SARS-CoV-2 during in vitro essays.[14] Among these compounds, two of them are potent nanomolar inhibitors that have already been tested in animals: **GC373** (or its prodrug **GC376**) and **11a** (see Figure 1a). The former was developed for the treatment of feline infectious peritonitis, a disease caused by a coronavirus and it has been shown to be also an inhibitor of SARS-CoV-2 main protease.[11] The structurally similar **11a** was initially designed analyzing the substrate binding pocket of the ortholog main protease of SARS-CoV.[10] The structures of their hemithioacetal products have been deposited in the Protein Data Bank with codes 6WTK and 6WTT (for **GC373**)[14] and 6LZE (for **11a**).[10] In both cases the inhibitor adopts a similar pose in the active site of the main protease. The distance between the Sγ atom of the catalytic cysteine and the aldehyde carbon atom is of about 1.8 Å, corresponding to the formation of a S-C covalent bond. The 6LZE and 6WTK PDB[15] structures shown in Figure 1b correspond to the (S) configuration of the hemitioacetal, where the hydroxyl oxygen is pointing to the oxyanion hole formed by main chain NH groups of Cys145, Ser144 and Gly143.[10] The main difference between the two structures showed in Figure 1b is the rotameric state of the catalytic histidine. In 6WTK the Nε atom of His41 is pointing towards the binding site while in 6LZE the Nδ atom is the one pointing to the inhibitor. These two configurations will be hereafter denoted as ε- and δ-rotamers, respectively. The rotameric state of His41 can be relevant for the reaction mechanism because the Nε atom in the ε-rotamer is better oriented to serve as a proton donor to the aldehydic oxygen atom of the inhibitor
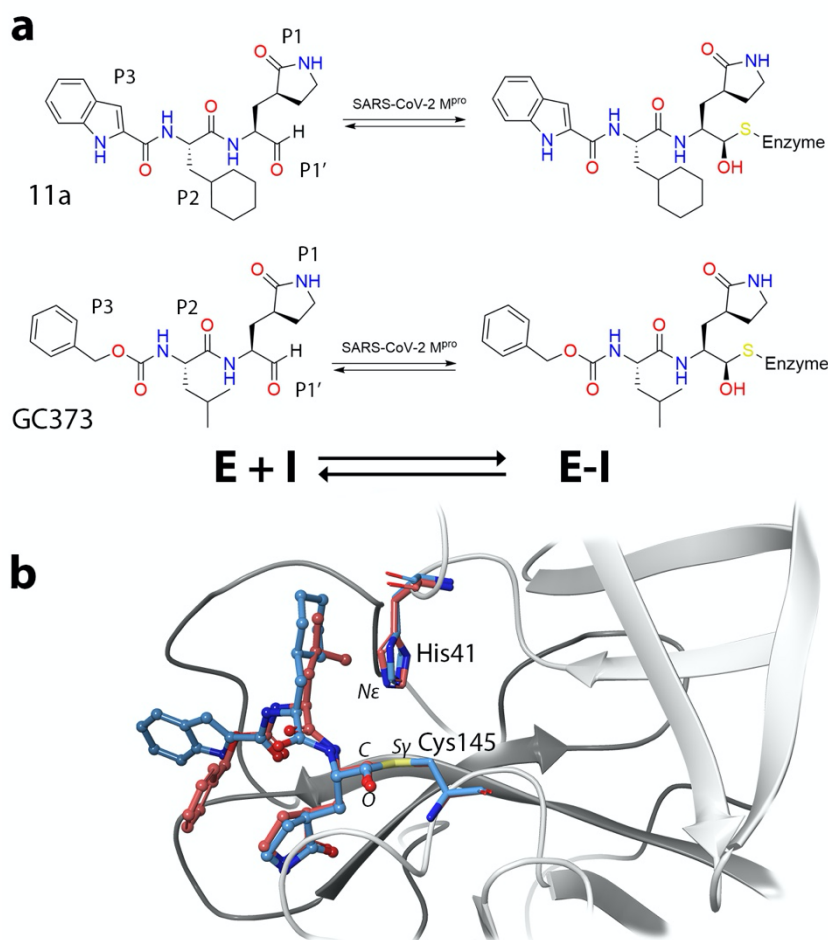
Figure 1. Inhibitors of the SARS-CoV-2 3CL protease presenting an aldehyde group as warhead. (**1a**) Aldehyde derivatives bind to the active site forming the noncovalent EI complex and then react with the thiol group of the catalytic cysteine to yield a hemithioacetal (the E-I covalent complex). (**1b**) Overlap of the x-ray structures of the (S)-hemithioacetal complexes formed between the protease and two aldehyde derivatives, **GC373**, where carbon atoms are shown in brown color, and **11a**, with carbons atoms in light blue. The PDB codes are 6WTT and 6LZE, respectively.

In this work we use classical and hybrid QM/MM molecular dynamics simulations to explore the inactivation mechanism of SARS-CoV-2 3CL protease by an aldehyde derivative, **11a**. We provide atomistic details of the inhibition process, including the formation of the noncovalent complex (EI) and the covalent product (the hemithioacetal complex or E-I). Our simulations show that the ε-rotamer is the most stable configuration both in the apo form and in the noncovalent complex with the inhibitor. From this complex the reaction proceeds via proton transfer from Cys145 to His41 that gives rise to a catalytic dyad ion pair (IP). This is the first step needed to initiate the reactions catalyzed by the 3CL protease, as observed both in experimental[7] and computational studies.[16,17] The process continues with the

nucleophilic attack of the Sγ atoms on the aldehydic carbon atom and the proton transfer from the catalytic histidine to the aldehyde oxygen atoms through a water molecule. The mechanism is confirmed by comparison with the structures of recently proposed inhibitor PF-00835231[12] that mimic the transition state of protease inhibition by aldehydes. Our simulations are also useful to rationalize the behavior of other inhibitors considering the possible orientations of the inhibitor warhead in the active site and how the reaction mechanism can differ depending on the conformational diversity of the inhibitor. All in all, the simulations presented here reveal the detailed reaction mechanism for the inhibition of SARS-CoV-2 3CL protease with a promising family of inhibitors, the aldehyde derivates and also with related compounds, such as ketones.

Methodology

Classical Molecular Dynamics simulations
To build the Michaelis Complex (MC) the PDB accession code 6LZE structure was used. This is the crystal structure of COVID-19 main protease forming a covalent complex with the **11a** inhibitor (with a resolution of 1.5 angstrom).[10] **11a** was parameterized following the non-standard residue parameterization procedure implemented in Amber using the Antechamber program[18] from the AmberTools18[19] package. Atomic charges were obtained using the Restrained Electrostatic Potential (RESP) method[20] at the HF/6-31G* level.

Tleap tool from abertools[19] was used to build the system, with ff14SB forcefield[21] to describe the canonical amino acids. PROPKA3.0[22] was used to determine the most likely protonation state of every residue at pH 7.4. Na$^+$ ions were added to neutralize the charge of the protein-inhibitor complex. The protein-inhibitor complex was solvated using a box of TIP3P water molecules in such a way that any protein-inhibitor atom is not closer than 12 Å to the limits of the simulation box. A minimization was made using 500 steps of steepest descent method followed by the conjugate gradient method until the root mean square of the gradient was below 10^{-3} kcal·mol^{-1}Å^{-1}. Heating was performed using a linear heating ramp, rising the temperature from 0 to 300 K along 120 ps followed by 20 ps simulation at 300 K. During this process the positions of the heavy atoms of the protein backbone were restrained using a harmonic potential with a force constant of 20 kcal·mol^{-1}·Å^{-2}. Then the system was equilibrated in the NPT ensemble (300K and 1 bar). For this equilibration, the force constants for the positional restraint were reduced from 15 to 0 kcal·mol^{-1}·Å^{-2}. So that, every 1.25 ns the force constant was decreased by 3 units. Then, after 6.25 ns the positional restraint was completely removed. Finally, the system was run restraint-free for another 1.25 ns. In order to get enough sampling 4 replicas of 1 μs of the noncovalent enzyme inhibitor complex with the catalytic dyad residues (Cys145 and His41) in neutral state and with these residues charged forming an ion pair. In these simulations the timestep was 2 fs

using SHAKE[23] to constraint bonds involving hydrogen atoms. Long-range electrostatic interactions were described using the Particle Mesh Ewald Method,[24,25] while the cut-off radius to evaluate the short-range interactions was 10 Å. Pressure and temperature were controlled using Berendsen barostat and Langevin thermostat, respectively. For all classical molecular dynamic simulations AMBER19 GPU version of PMEMD[26,27] was employed.

In order to study the free energy profiles associated to conformational changes of the catalytic histidine and of the aldehyde inhibitor, classical potential of mean forces were performed using umbrella sampling.[28] For the change in the rotameric state of His41, the dihedral angle formed by the $C\alpha-C\beta-C\gamma-N\delta$ atoms of the residue was used as distinguished coordinate. The dihedral angle was evolved using increments of 5 degrees with a total of 41 windows. For the rotation of the aldehyde group of the inhibitor, (the dihedral angle formed by the $N-C-C_{aldh}-O_{aldh}$ was selected as distinguished coordinate. In this case, the dihedral angle was evolved using increments of 10 degrees during 21 simulation windows. In both cases, each simulation window was minimized under the harmonic restraint with a force constant of 100 kcal·mol^{-1}·rad^{-2} using 2000 steps of steepest descent method followed by the conjugate gradient method until the root mean square of the gradient was below 10^{-3} kcal·mol^{-1}Å^{-1}. Afterwards, a total of 52 ns of classical MD was performed for each window, the first 2 ns were run for relaxation followed by 50 ns of production. Therefore, the total simulation time for the production stage was longer than 2 μs for each free energy profile corresponding to His41 rotation and longer than 1 μs for the rotation of the aldehyde group of the inhibitor. The free energy profiles were integrated using the Weighted Histogram Analysis Method (WHAM).[29]

QM/MM calculations

The Adaptative String Method (ASM)[30] developed in our group was used to explore the free energy landscape associated to the chemical reaction. This methodology has the advantage to avoid the oversimplified picture of what a reaction mechanism is, in a real system a reaction mechanism involves many reaction coordinates, and the free energy pathway depends on all of them. Using the ASM we can find the minimal free energy pathway (MFEP) in a multidimensional free energy surface not just by evaluating the change of energy related to the variation of a couple of distances, angles, or dihedrals (are made in most of computational studies) but including all the needed degrees of freedom without implying and additional computational expense.

The mechanistic proposals were explored using 96 replicas of the system (string nodes) to connect the reactant and product structures along the minimal free energy path in a space of arbitrary dimensionality defined by the collective variables (CVs) shown in Scheme 1.

Using QM/MM MD simulations, nodes are moved according to their free energy gradient and redistributed equidistantly along the string, avoiding them to fall to the global minima (reactants and products). This procedure is continued until the string converges to the Minimum free Energy Path (MFEP) displaying a RMSD below 0.1 $amu^{1/2} \cdot Å$ for at least 2 ps. Replica exchange between nodes was attempted every 50 steps to ensure convergence. After convergence, a path-CV (denoted as s) measuring the advance of the system along the MFEP from reactants to products is defined and employed to trace the reaction free energy profile. 10 ps of QM/MM simulations were run for every node and WHAM[29] was selected as the integration method. To ensure a probability density distribution of the reaction coordinate as homogeneous as possible, the values of the force constants employed to bias the ASM simulations were determined on-the-fly.[30] The QM region was described using a B3LYP functional[31,32] with a 6-31+G* basis set and D3 dispersion corrections.[33] This is an good choice to describe both the acylation of a peptide substrate[16] and an inhibitor[34] by the SARS-CoV-2 protease with activation energy results in excellent agreement with experiment. A systematic study on the cysteine-histidine proton transfer also pointed out to the B3LYP functional as the most adequate.[35] All QM/MM calculations were performed using a modified version of Amber18[19,36] coupled to Gaussian16[37] for Density Functional Theory calculations. The cutoff-radii used for all the QM-MM interactions was of 15 Å. For mechanistic determinations, the QM region included the side chains of the catalytic dyad (His41 and Cys145), the water molecule involved in the reaction mechanism, the backbone atoms of residues P1 and P1' in the **11a** inhibitor. Any other atom was described a MM level as explained in the classical molecular dynamic section. The integration time step in QM/MM simulations was of 1 fs.
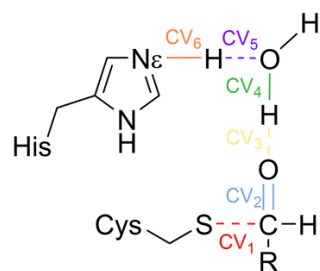
Scheme 1. Representation of the Collective Variables (CVs) used in the exploration of the reaction mechanism.

For the study of the proton transfer within the catalytic dyad, i. e. from Cys145 to His41, the antisymmetric combination of the distances of the proton to the donor and acceptor atoms (d(Sγ-H)-d(Nε-H)), was employed to trace the free energy profile using Umbrella Sampling.[28] The integration was carried out using WHAM method. In this case only the side chains of

the two involved residues were included in the QM region (described at the B3LYPD3/6-31+G* level of theory with D3 corrections). 40 windows evenly spaced each 0.06 Å were used to cover the whole range of the reaction coordinate. A force constant of 600 kcal·mol^{-1}·Å^{-2} was employed to drive the reaction coordinate. All the rest of details of the simulations were as described previously.

Results and Discussion

The rotameric state of the catalytic histidine
Our simulations of the noncovalent complex formed by **11a** were carried out using the x-ray coordinates of the hemithioacetal complex (PDB file 6LZE) as starting point. In agreement with the x-ray structure we simulated the inhibitor in the active sites of the two protomers (A and B) of the dimeric enzyme. The x-ray structure was modified lengthening the Sγ-C distance between the enzyme and the inhibitor. The catalytic dyad (Cys145 and His41) was modelled in the neutral state, which is the most stable form for the Michaelis and EI complexes.[16,34,35]

As explained before, in the 6LZE x-ray structure the catalytic histidine is found in the δ-rotameric state. With the help of Pocketome[38] we explored 36 different x-ray structures of the 3CL main protease in the apo form and containing several covalent and noncovalent inhibitors (see Table S1). In 80% of the active sites, including all those corresponding to apo enzyme and most of the inhibited enzymes, His41 was found in the ε-rotameric state; while less than 20% presented the δ-rotameric state. We modelled the noncovalent complex of the 3CL protease with **11a** with the two His41 rotamers (see Figures 2a and 2b) and ran 1 μs long MD simulations. In principle, both rotameric states could be productive as far as in the two states His41 and Cys145 form a strong hydrogen bond, which can eventually lead to a proton transfer and the formation of the catalytic dyad IP. Figure 2c shows that the probability distribution of Nε(His41)-Sγ(Cys145) distances obtained from MD simulations of the **11a** noncovalent complex with the two rotameric states are almost indistinguishable and the most probable distance (3.4 Å) correspond to a hydrogen bonded dyad. In order to determine the relative stability of the two rotameric states we traced the free energy profile associated to the rotation of His41 around the Cβ-Cγ bond in the apo form and in the noncovalent complex (EI) with **11a**. The free energy profiles obtained after 2 μs of classical MD simulations are shown in Figure 2d. According to these results, both in the apo form and in the presence of the **11a** inhibitor the ε-rotamer is more stable than the δ-rotamer by 2.0 and 3.0 kcal·mol^{-1}, respectively. This result agrees with the observation that the ε-rotamer predominates in the x-ray structures of the 3CL protease, as explained above. However, the free energy difference between the two rotameric states is not too high,

which could explain that the δ-rotamer is found in 20% of the active sites reported in Table S1 of the SARS-CoV-2 3CL protease.
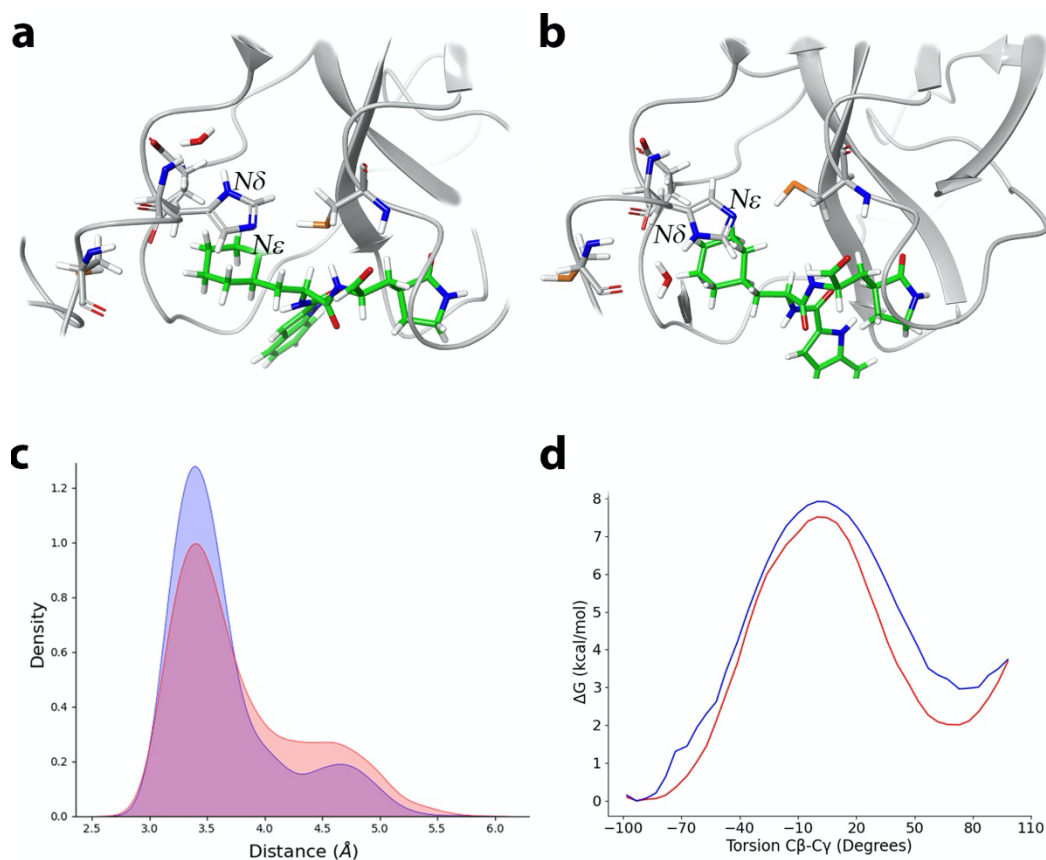

Figure 2. Rotameric state of the catalytic His41. (**2a**) Representation of the noncovalent complex of **11a** (carbon atoms in green) with SARS-CoV-2 3CL protease with His41 in the ε-rotameric state (**2b**) Representation of the complex presenting the δ-rotameric state of His41 (**2c**) Probability densities of the distances from the Cys145-Sγ atom to the Nε atom of His41 for the ε- (blue) and δ-rotameric state (red). **(2d)** Free energy profile associated to the rotation around the Cγ-Cβ bond of His41, from the ε-rotamer (left) to the δ-rotamer (right) in the apo enzyme (red) and in the noncovalent complex with **11a** (blue).

The noncovalent Enzyme-Inhibitor complex

According to the results of the previous section we selected the ε-rotameric state for all subsequent simulations of the complex formed between the 3CL protease of SARS-CoV-2 and **11a**. The resulting MD simulations of the noncovalent EI complex (4 replicas of 1 μs each) were stable in all cases (see RMSD time evolutions in Figure S1), showing a binding pose consistent with the x-ray structure (see Figure 3a). In this pose, P1-P3 sites of the inhibitor present an interaction pattern similar to that of a peptide substrate with sequence

-Val-Leu-Gln|Ser- (the vertical line indicates the scissile bond).[16] Figure 3b compares the fraction of hydrogen bonds between inhibitor/peptide sites and enzymatic residues during the MD simulations of the inhibitor and of the peptide substrate. The γ-lactam ring at the P1 position is frequently found in inhibitors of SARS-CoV-2 and SARS-CoV main proteases, because this moiety is expected to mimic Gln at P1 position of the peptide substrate, which is a requirement of these enzymes. This γ-lactam ring can form hydrogen bonds with His163, Glu166 and Phe140, displaying an interaction pattern similar to that of Gln-P1. The cyclohexyl group at the P2 position of the inhibitor stacks with the imidazole ring of His41 and also presents interactions with other nearby residues, such as Met165. Inhibitors of the SARS-CoV-2 main protease present hydrophobic groups at the P2 position, similarly to the side chain of Leu-P2 in the peptide substrate. Finally, the indole group at P3 is exposed to the solvent and stabilized by hydrogen bond interactions with main chain atoms of Glu166.

The chemical step in SARS-CoV-2 3CL protease requires a proton transfer from Cys145 to His41 to form a catalytic dyad IP.[16,35] As presented above, in the case of the non-covalent EI complex these two residues are kept at a hydrogen bond distance a significant fraction of the simulation (see Figure 2c for one replica; results for other replicas are presented in Figure S2). After this proton transfer, the covalent inhibition of the enzyme should proceed with the formation of the corresponding hemithioacetal (the E-I complex) that requires the nucleophilic attack of the Sγ atom to the aldehyde carbon atom and the protonation of the oxygen atom. Figure 3c shows the distribution of distances between the Cys145 Sγ atom and the aldehyde carbon atom. The distribution shows a bimodal shape, with two peaks centered at 3.3 and 5.1 Å that correspond to the *trans* and *gauche* conformations of Cys145 side chain, respectively. Thus, in this pose the aldehyde carbon atom of the inhibitor can be found already at short enough distances to suffer the nucleophilic attack by the Sγ atom of Cys145. The position of the aldehyde oxygen atom is stabilized by means of hydrogen bond interactions with the main chain NH groups of Cys145 (2.4 $\pm$ 0.3 Å), Ser144 (2.8 $\pm$ 0.4 Å) and Gly143 (2.3 $\pm$ 0.3 Å), as seen in Figure 3a. These three residues form the oxyanion hole, placed in a U-turn of the loop connecting β10 and β11 and that closes one of the sides of the active site. These interactions are also observed in the x-ray structure of the (S)-hemithioacetal product (see Figure 1b).[10] In this pose the aldehyde oxygen atom is placed far from the Nε atom of His41, as confirmed by the probability distribution of Nε-O distances, peaked at 6.3 Å (see Figure 3c). This separation precludes a direct proton transfer between these two atoms after formation of the IP. However, as discussed below, once the IP is formed a water molecule can be accommodated in between the catalytic histidine and the aldehyde group of the inhibitor, facilitating a water-mediated proton transfer from His41 to the aldehyde oxygen atom to form the hemithioacetal.
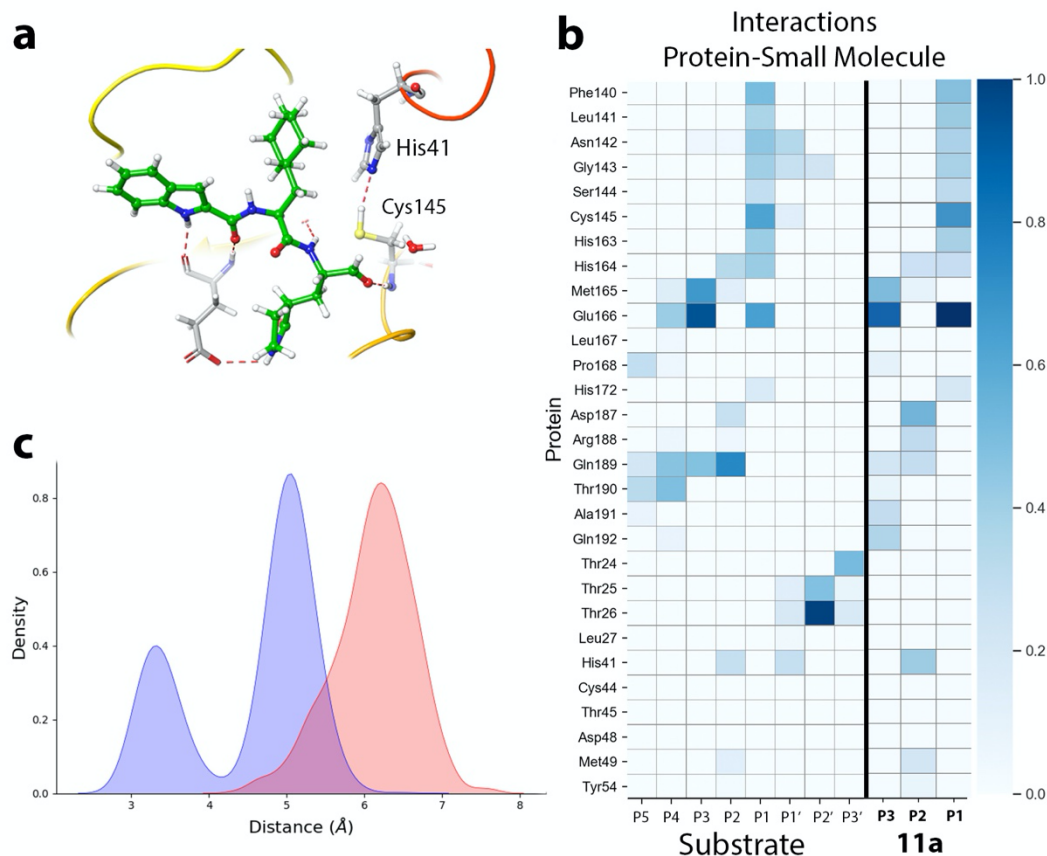
Figure 3. Noncovalent complex formed between the aldehyde derivative **11a** and the 3CL protease of SARS-CoV-2. (**3a**) Binding pose of the inhibitor in the active site of the protease, showing the location of the catalytic dyad and the oxyanion hole. Note that the aldehyde oxygen points to the oxyanion hole. (**3b**) Fraction of hydrogen bond contacts between residues of **11a** and a peptide substrate[16] 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°. (**3c**) Probability densities of the distances from the Cys145-Sγ atom to the C carbon atom of the substrate, in blue and from the Nε atom of His41 to the aldehyde oxygen atom, in red.

The proposed mechanism, with a water-mediated proton transfer, could also explain the reactivity of other aldehyde inhibitors to form the (S)-hemithioacetal, which is the enantiomer more frequently observed in the x-ray structures (see Table S1). X-ray structures of other (S)-hemithioacetal complexes show a similar pose for all of them, with the aldehyde oxygen atom pointing to the oxyanion hole and thus far enough from the catalytic histidine for a direct proton transfer. However, there is one case where the (R) enantiomer of the product has been observed: the 6WTT x-ray structure contains a (R)-hemithioacetal in one of the three protomers of the asymmetrical unit and the (S) configuration in the other two.[14] The (R) enantiomeric form can be obtained if the aldehyde

group of the inhibitor rotates around the C-C(O) bond before the nucleophilic attack by Cys145. In that case the aldehyde oxygen atom would point towards the catalytic His41 instead of towards the oxyanion hole and the distance to the Nε atom would become small enough to allow a direct proton transfer between them (see Figure 4a). While this binding pose of the inhibitor simplifies the reaction mechanism it is not the most stable for **11a**, as demonstrated by the free energy profile associated to the rotation around the C-C(O) (see Figure 4b). The pro-(S) conformation in the noncovalent EI complex, where the aldehyde oxygen atom is accommodated in the oxyanion hole, is 2.3 kcal·mol^{-1} more stable than the pro-(R) conformer, where the aldehyde oxygen atom points towards His41. This small free energy difference agrees with the fact that the (R)-hemithioacetal can also be formed, as discussed above. In addition, this suggests that the preference for one or another orientation of the oxygen atom can be modulated through changes in the chemical structure of the inhibitor, for example substituting the aldehyde hydrogen atom by other groups able to interact more favorably with the oxyanion hole. This seems to be the case of Boceprevir, a ketone inhibitor where the aldehyde hydrogen is substituted by an acetamide group. In this case, the x-ray structure (PDB code 7BRP[39]) shows that the acetamide group is accommodated in the oxyanion hole while the ketone oxygen atom points to the catalytic His41, facilitating a reaction mechanism with a direct proton transfer from the catalytic histidine to the carbonyl oxygen atom.[39]
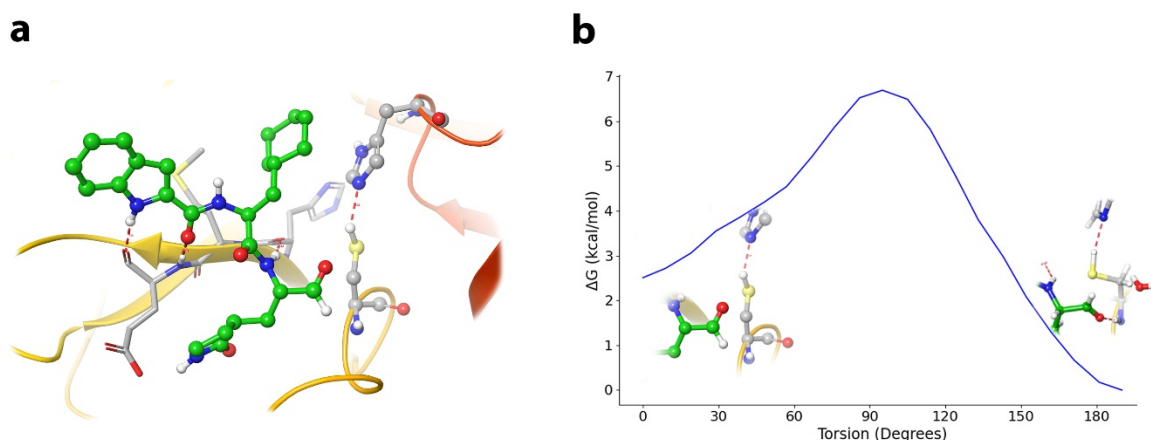

Figure 4. Orientation of the aldehyde group in the noncovalent complex. **(4a)** Representation of the EI complex with inhibitor **11a** in the pro-(R) conformation **(4b)** Classical free energy profile for the rotation of the aldehyde group from the pro-(S) conformer (right) to the pro-(R) conformer (left).

Formation of the (S)-hemithioacetal complex

Chemical transformations in the active site of 3CL protease are initiated with the proton transfer from Cys145 to His41 to form an ionized catalytic dyad (the IP) from which the reaction proceed. We traced the QM/MM free energy profile for the proton transfer from Cys145 to His41 using as distinguished coordinate the antisymmetric combination of proton distances to the donor and acceptor atoms (d(Sγ-H)-d(Nε-H), see Figure 5a). Starting from the EI complex, the free energy cost of forming the IP is 9.3 kcal·mol^{-1}, a value slightly larger than that obtained when a peptide substrate is present in the active site (4.8 kcal·mol^{-1}).[16] The value obtained for **11a** is similar to those found for the formation of IP under the presence of other inhibitors: 7.3 kcal·mol^{-1} with an α-ketoamide inhibitor[40] and 10.3 with a Michael acceptor.[34] These value are systematically larger than the free energy cost evaluated for the apo enzyme, which was found to be 2.9 kcal·mol^{-1} in two different works,[16,40]indicating that desolvation of the active site upon ligand binding can destabilize the IP form. Diminution of this energy penalty through ligand design could be a promising strategy to improve inhibition binding and kinetics.

After formation of the IP, the inhibition reaction must proceed with the attack of the activated nucleophile (the Sγ atom of Cys145) on the aldehyde carbon atom and the proton transfer from the Nε atom of His41 to the aldehyde oxygen atom. As explained before the distance between these two atoms in the IP is too large for a direct proton transfer. However, MD simulations show that a water molecule can be placed in between His41 and the aldehyde group, attracted by the large dipole moment associated to the IP (see Figure 5b). This water molecule can act as a proton relay, accepting a proton from His41 and donating a proton to the aldehyde oxygen atom to form the hydroxyl group of the hemithioacetal.
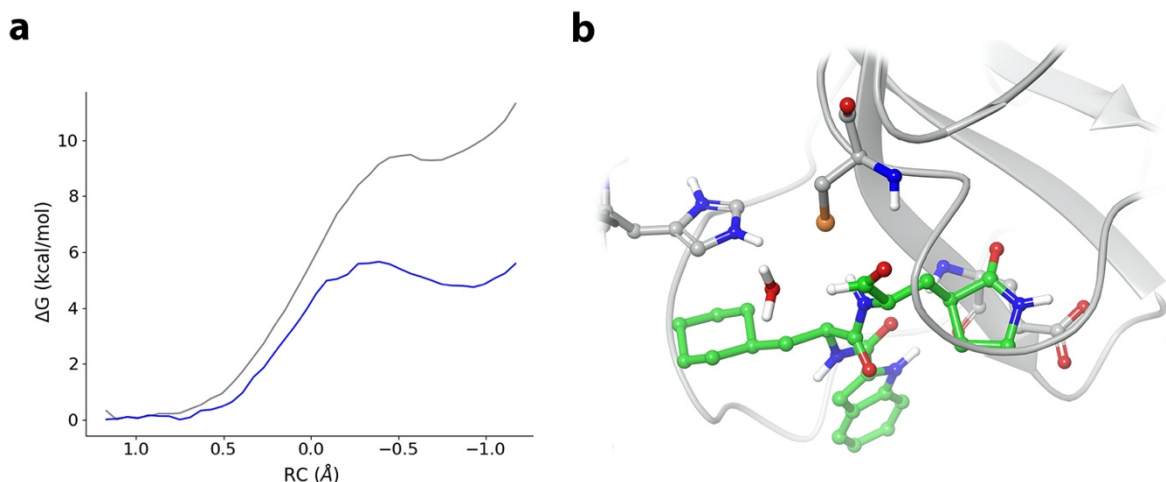
Figure 5. Proton transfer from Cys145 to His41 in SARS-CoV-2 3CL protease. **(5a)** Free energy profile for the transformation from the neutral catalytic dyad (left) to the ion pair (IP, right) in the EI complex with **11a** (grey line) and in the complex with the peptide substrate[16] (blue line). The reaction coordinate (RC) is the antisymmetric combination d(Sγ-H)-d(Nε-H). **(5b)** IP structure with **11a** showing the presence of a water molecule in between the protonated histidine and the inhibitor aldehyde group.

Starting from the structure presented in Figure 5b we obtained the corresponding QM/MM MFEP for the formation of (S)-hemithioacetal product, the enantiomer experimentally observed for inhibitor **11a**. Figure 6a shows the B3LYPD3/6-31+G*/MM free energy profile associated to the reaction from IP to the covalent E-I complex, while Figure 6b shows the evolution of key distances along this MFEP (see also Scheme 1). The reaction begins with the approach of the Sγ atom to the aldehyde carbon atom, reducing the Sγ-C distance from 2.95 to 2.18 Å at the first transition state (TS1, shown in Figure S3). Then, the water molecule is accommodated in between the proton donor His41 and the proton acceptor aldehyde oxygen. The rate-limiting TS is TS2 (see Figure 5c) where the formation of the S-C bond (presenting a distance of 1.94 Å) is accompanied by a lengthening of the aldehyde double bond (from 1.22 to 1.38 Å), due to the charge transfer from Cys145 to the aldehyde oxygen atom to form an oxyanion. This charge transfer triggers a concerted proton transfer from the water molecule to the oxyanion and from the protonated His41 to the water molecule. These coupled proton transfer events are not completely synchronous because at the TS the distance of the transferred proton to the aldehyde oxygen is 1.20 Å, while the distance from the proton transferred from Hus41 to the oxygen water molecule is slightly longer, 1.38 Å. The process continues downhill up to the products (see Figure 6d), completing the formation of the Sγ-C bond (1.85 Å), the lengthening of the aldehyde bond distance up to a typical value for a single C-O bond (1.42 Å) and the proton transfers from the water molecule to the hemithioacetal and from His41 to the water molecule (see Figure

6d). The small barrier observed at the end of the reaction path corresponds to the rotation of the hemithioacetal hydroxyl group to fit into the oxyanion hole.

The plausibility of the proposed water-mediated mechanism can be confirmed comparing the structure obtained for the reaction TS with the x-ray structure of the protein with the PF-00835231 inhibitor (PDB code 6XHM).[12] This recently proposed inhibitor of the SARS-CoV-2 main protease presents a ketone group as a warhead and the aldehyde hydrogen atom has been substituted by a hydroxymethyl group. As observed in Figure 6e the hydroxyl moiety of the PF-00835231 inhibitor in the 6XHM structure occupies exactly the same position that the bridging water molecule in our TS. Then, the PF-00835231 inhibitor would be a perfect Transition State Analogue of protease inhibition by aldehydes. On one hand this structural agreement confirms the role of the water molecule in the inhibition process by aldehydes and, on the other hand, suggests that in the case of the hydroxymethylketone inhibitor the proton relay role of the water molecule could be played by the hydroxyl group. Finally, as a further confirmation of our mechanism, the structure obtained for the final product nicely overlaps with that corresponding to the (S)-hemithioacetal complex in the x-ray structure 6LZE (see Figure 6f).
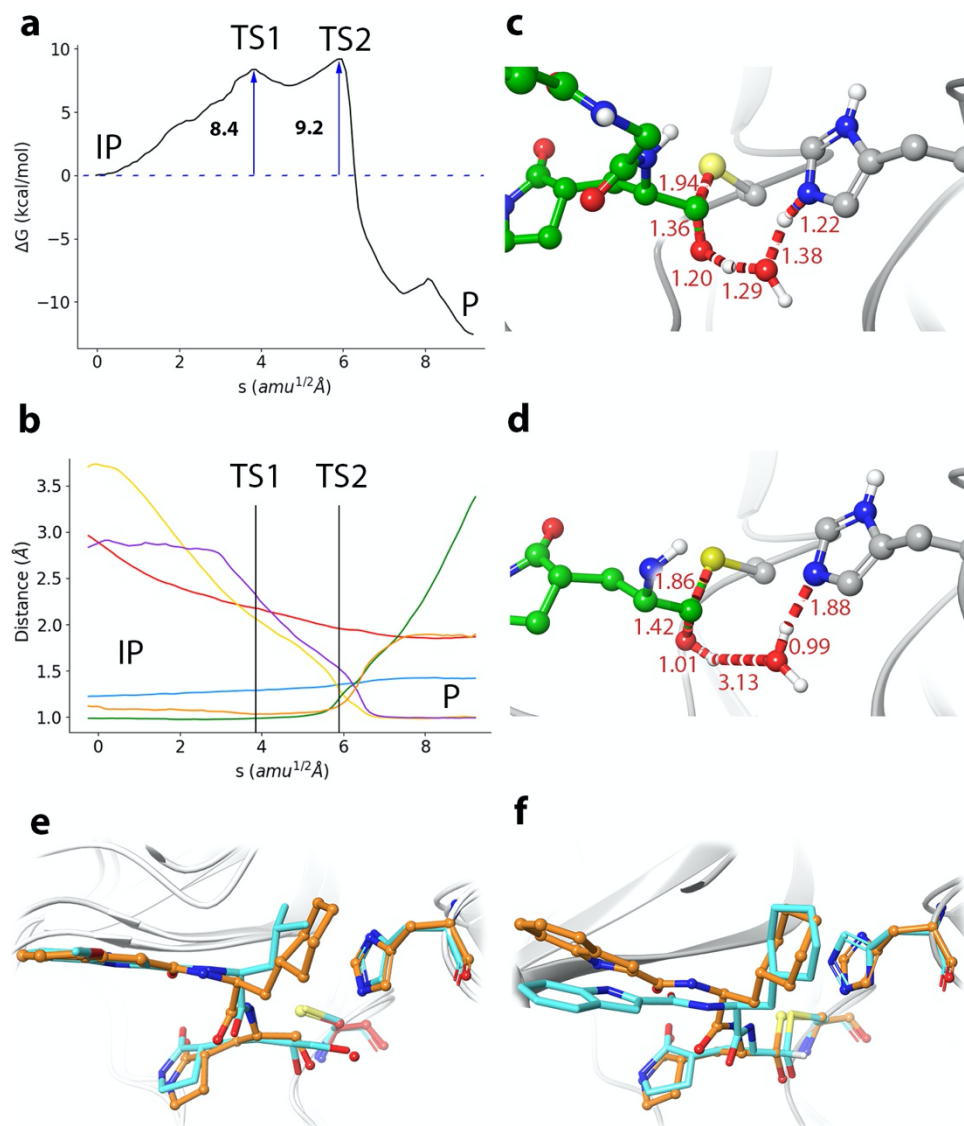
Figure 6. Formation of the (S)-hemithioacetal product from the ion pair. **(6a)** B3LYPD3/6-31+G*/MM free energy profile along the path-CV for the formation of the covalent E-I complex from the IP. **(6b)** Evolution of the selected CVs along the MFEP. The color code corresponds to Scheme 1. **(6c)** Representation of TS. The values of the distances correspond (in Å) to the coordinates of the MFEP where TS1 is located. **(6d)** Representation of the (S)-hemithioacetal complex. **(6e)** Overlap of the TS 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). Note that the hydroxyl group of the inhibitor is placed at a position equivalent to the water molecule in the TS of the inhibition by **11a**. (**6f**). Overlap of the product structure obtained from our simulations (balls & sticks with carbon atoms in orange) with the x-ray structure 6LZE (licorice with carbon atoms in light blue).

Combining the free energy profiles presented in Figures 5 and 6, we can obtain a full picture of the transformation from the noncovalent EI complex formed between the SARS-CoV-2 protease and the inhibitor **11a** to the covalent E-I complex (the (S)-hemithioacetal). The total reaction free energy can be obtained combining the free energy cost of forming the IP and the free energy change from the IP to the reaction product, E-I. The first term was estimated to be 9.3 kcal·mol^{-1} (see Figure 5a), while the free energy difference between E-I and IP (Figure 6a) is –12.1 kcal·mol^{-1}. Then, our simulations predict that the free energy of the covalently bonded E-I complex relative to the noncovalent EI complex is -2.8 kcal·mol^{-1}. This reaction free energy is significantly smaller, in absolute value, than the value obtained for the **N3** inhibitor (-15.4 kcal·mol^{-1}),[34] an irreversible inhibitor of the 3CL protease.[8] Thus, according to our simulations, inhibition of SARS-CoV-2 3CL protease with **11a** would be significantly more reversible than with **N3**. In fact, aldehyde derivative inhibitors have been proposed to act via a reversible formation of the hemithioacetal,[11] which is now confirmed by our simulations. It must be noticed that reversibility can be a desired feature of cysteine protease inhibitors, in particular when extended therapy periods are required.[41]

Regarding the kinetic of the process, our free energy simulations provides the activation free energy as the sum of two contributions: the free energy cost of forming the reactive IP from EI (9.3 kcal·mol^{-1}, Figure 5a) plus the free energy of the TS relative to IP (9.2 kcal·mol^{-1}, Figure 6a). This gives in a total activation free energy of 18.5 kcal·mol^{-1}. The inhibition rate constant of SRAS-CoV-2 protease by **11a** has not been experimentally determined. However, the rate constant for the inhibition of the protease with **GC376** (the prodrug of the aldehyde inhibitor **GC373**) has been measured to be 2.45·10^{-3} s^{-1} at 30ºC.[14] According to Transition State Theory this is equivalent to an activation free energy of 21.1 kcal·mol^{-1}. In spite of the differences between both inhibitors, **11a** and **GC373**, they present the same warhead and the same group at the P1 position (see Figure 1a), suggesting that the activation free energy derived for **GC373** could be a reasonable reference for **11a** and thus for our simulations. The good agreement between the experimental observations made for **GC373** (21.1 kcal·mol^{-1}) and our theoretical predictions for **11a** (18.5 kcal·mol^{-1}), together with the structural evidences discussed above, strongly supports our mechanistic proposal for the inhibition of SARS-CoV-2 3CL protease, not only by **11a** but also by other aldehyde and ketone derivatives.

Conclusions

We have used a combination of classical MM and hybrid QM/MM Molecular Dynamics simulations to explore the inhibition mechanism of the SARS-CoV-2 3CL protease by an aldehyde derivative, **11a**, selected as an example of this promising family of inhibitors. Starting from the x-ray structure of the hemithioacetal complex, we have explored the binding mode of the **11a** inhibitor in the noncovalent EI complex with the protease. In agreement with x-ray observations, our MD simulations show that there are two possible rotamers for the catalytic histidine, depending on the internal rotation around the Cβ-Cγ bond. Both for the apo form and for the noncovalent complex formed between the 3CL protease and the inhibitor we found that the rotamer where the Nε atom of His41 lies closer to the inhibitor is more stable. Regarding the inhibitor, the orientation of the aldehyde group is determined by the interaction of the oxygen atom with the oxyanion hole, favoring the formation of the (S)-enantiomer of the product. This configuration of the protein and the inhibitor facilitates a reaction mechanism where, after formation of the catalytic dyad ion pair, a proton is transferred from His41 to the aldehyde oxygen atom through a water molecule. Other reaction mechanisms, involving different proton transfer routes and/or leading to the (R) form of the product, can be feasible if other conformations are populated. This conclusion is supported by the analysis of the x-ray structures available for the inhibited 3CL protease.

Our simulations of the noncovalent complex with **11a** show that P1-P3 groups of the inhibitor establish well-defined interactions in the active site that closely mimic those of the peptide substrate. The aldehyde group is placed close to Cys145, facilitating the nucleophilic attack of the Sγ atom on the aldehyde carbon atom while the oxygen atom is stabilized by interactions with NH main chain groups that constitute the oxyanion hole. The reactivity of the SARS-CoV-2 main protease is triggered after the proton transfer from Cys145 to His41, which are kept at hydrogen bond distances in the EI complex. After ion pair formation the inhibition reaction proceeds by means of the nucleophilic attack of the activated nucleophile, the unprotonated Cys145, and a water-mediated proton transfer from the protonated His41, yielding the (S)-hemithioacetal product. In our QM/MM Minimum Free Energy Profile the activation free energy is 18.5 kcal·mol^{-1}, in good agreement with the value experimentally derived for a similar aldehyde inhibitor (21.1 kcal·mol^{-1}). In addition, our reaction profile indicates that the process is moderately exergonic, in agreement with the proposed reversibility for the protease inhibition by aldehydes.

This study illustrates the importance of molecular simulations to unravel the reaction mechanisms and to guide the design of possible inhibitors. The binding pose observed in x-

ray structures can be not enough to figure out the mechanistic details of the inhibition process when conformational rearrangements in the protein, the solvent and/or the inhibitor are required for the process to take place. In addition, apparently small modifications in the chemical structure of the inhibitor can change the binding pose offering different mechanistic ways for the inactivation of the enzyme. The consideration of the mutual coupling observed between active site, solvent and inhibitor dynamics along the reaction path is an undeniable challenge for the design of selective and efficient inhibitors that can be approached by means of adequate simulation methods.

Acknowledgements

References

(1)   Coronavirus disease (COVID-19) Weekly Epidemiological Update and Weekly Operational Update https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports (accessed Jun 12, 2020).

(2)   Horby, P. W.; Mafham, M.; Bell, J. L.; Linsell, L.; Staplin, N.; Emberson, J.; Palfreeman, A.; Raw, J.; Elmahi, E.; Prudon, B.; Green, C.; Carley, S.; Chadwick, D.; Davies, M.; Wise, M. P.; Baillie, J. K.; Chappell, L. C.; Faust, S. N.; Jaki, T.; Jefferey, K.; Lim, W. S.; Montgomery, A.; Rowan, K.; Juszczak, E.; Haynes, R.; Landray, M. J. Lopinavir–Ritonavir in Patients Admitted to Hospital with COVID-19 (RECOVERY): A Randomised, Controlled, Open-Label, Platform Trial. *Lancet* **2020**, *396* (10259), 1345–1352. https://doi.org/10.1016/S0140-6736(20)32013-4.

(3)   Wu, R.; Wang, L.; Kuo, H.-C. D.; Shannar, A.; Peter, R.; Chou, P. J.; Li, S.; Hudlikar, R.; Liu, X.; Liu, Z.; Poiani, G. J.; Amorosa, L.; Brunetti, L.; Kong, A.-N. An Update on Current Therapeutic Drugs Treating COVID-19. *Curr. Pharmacol. Reports* **2020**, *6* (3), 56–70. https://doi.org/10.1007/s40495-020-00216-7.

(4)   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.

(5)   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* (14), 6595–6628. https://doi.org/10.1021/acs.jmedchem.5b01461.

(6)   Brocklehurst, K.; Willenbrock, F.; Sauh, E. Cysteine Proteinases. In *Hydrolytic Enzymes*; Elsevier: New York, 1987; pp 39–158. https://doi.org/10.1016/S0167-7306(09)60016-8.

(7)   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.

(8)   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. 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.

(9)   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.

(10)   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.

(11) 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.

(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.; Steppan, 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) Mole, J. E.; Horton, H. R. Kinetics of Papain-Catalyzed Hydrolysis of α-N-Benzoyl-L-Arginine-p-Nitroanilide. *Biochemistry* **1973**, *12* (5), 816–822. https://doi.org/10.1021/bi00729a005.

(14) 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.

(15) Berman, H.; Henrick, K.; Nakamura, H. Announcing the Worldwide Protein Data Bank. *Nat. Struct. Mol. Biol.* **2003**, *10* (12), 980–980. https://doi.org/10.1038/nsb1203-980.

(16) Ramos-Guzmán, C. A.; Ruiz-Pernía, J. J.; Tuñón, I. Unraveling the SARS-CoV-2 Main Protease Mechanism Using Multiscale Methods. *ACS Catalysis* **2020**, *10* (21), 12544–12554. https://doi.org/10.1021/acscatal.0c03420.

(17) Świderek, K.; Moliner, V. Revealing the Molecular Mechanisms of Proteolysis of SARS-CoV-2 Mpro from QM/MM Computational Methods. *Chem. Sci.* **2020**, No. 1. https://doi.org/10.1039/D0SC02823A.

(18) 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.

(19) 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*.

(20) 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.

(21)   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.

(22)   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.

(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.

(24)   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* (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/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)   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 Inhibitors Design. *ChemRxiv* **2020**, 2–6. https://doi.org/https://doi.org/10.26434/chemrxiv.12895064.v1.

(35)   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* (3), 1765–1777. https://doi.org/10.1021/ct301082y.

(36)   Zinovjev, K. String-Amber https://github.com/kzinovjev/string-amber (accessed Jun 24, 2020).

(37)   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.

(38)   Kufareva, I.; Ilatovskiy, A. V.; Abagyan, R. Pocketome: An Encyclopedia of Small-Molecule Binding Sites in 4D. *Nucleic Acids Res.* **2012**, *40* (D1), 535–540. https://doi.org/10.1093/nar/gkr825.

(39)   Fu, L.; Ye, F.; Feng, Y.; Yu, F.; Wang, Q.; Wu, Y.; Zhao, C.; Sun, H.; Huang, B.; Niu, P.; Song, H.; Shi, Y.; Li, X.; Tan, W.; Qi, J.; Gao, G. F. Both Boceprevir and GC376 Efficaciously Inhibit SARS-CoV-2 by Targeting Its Main Protease. *Nat. Commun.* **2020**, *11* (1), 4417. https://doi.org/10.1038/s41467-020-18233-x.

(40)   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**, acs.biochem.0c00782. https://doi.org/10.1021/acs.biochem.0c00782.

(41)   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.