

# **Binding and Reactivity of a Nitrile Oral Inhibitor of SARS-CoV-2 Main Protease Revealed by Computational 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](mailto:ignacio.tunon@uv.es)

[j.javier.ruiz@uv.es](mailto:j.javier.ruiz@uv.es)

**Abstract**

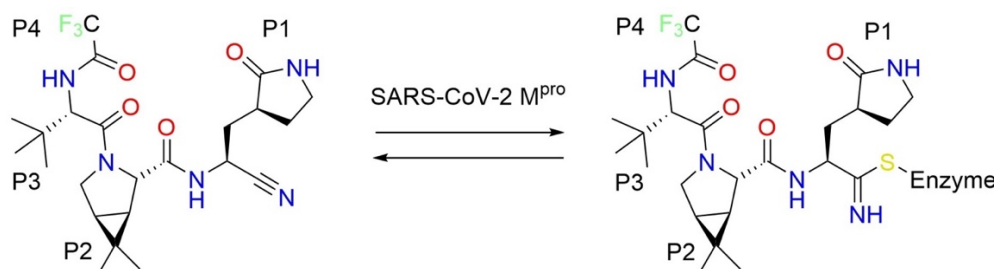
We present a detailed analysis of the binding mode and reactivity of the novel oral inhibitor PF-07321332 developed against SARS-CoV-2 3CL protease. Classical and QM/MM Molecular Dynamics simulations are used to quantify the contributions to the binding free energy and the reaction mechanism for covalent inhibition. The small size of the nitrile warhead confers additional advantages to this inhibitor.

*Keywords:* 3CL protease, SARS-CoV-2, PF-07321332, Minimum Free Energy Path, QM/MM, Binding Free Energy, nitrile-based inhibitor

## Introduction

One of the possible strategies to fight against COVID-19 is the development of compounds able to inhibit the activity of one of the essential enzymes of the SARS-CoV-2 virus. In this sense, one of the most attractive targets is the main protease of the virus, also known as 3CLpro. This enzyme is a cysteine protease in charge of the proteolysis of the polyproteins produced after translation of its genome in the infected cells.<sup>1</sup> Inhibition of this essential enzyme results in the inability of the virus to produce its nonstructural proteins, interrupting its replication cycle. As an additional advantage, this protease uses a recognition sequence not known in any of the human proteases, decreasing the risk of side effects after the administration of these compound.<sup>2</sup> Some of the compounds reported to have inhibitory capacity against SARS-CoV-2 3CLpro are covalent inhibitors that contain an electrophilic site able to react with the catalytic cysteine, Cys145.<sup>3-7</sup>

Last March, Pfizer announced that one of the antiviral compounds developed by this company against SARS-CoV-2, PF-07321332, went under clinical trial.<sup>8</sup> PF-07321332 (see Figure 1) is a peptidomimetic covalent inhibitor that contains a nitrile group as a warhead. Peptide nitriles act as reversible active site directed inhibitors, being able to react with the catalytic cysteine to form a thioimidate.<sup>9</sup> The main advantage of PF-07321332 with respect to other inhibitors of the SARS-CoV-2 3CLpro is the possibility of oral administration, a feature that could dramatically facilitate the treatment of COVID-19. This compound presents a  $\gamma$ -lactam ring as a substituent in the peptidic group just before the warhead in the peptide sequence (the P1 position). This is a common feature in most of the covalent inhibitor developed against SARS-CoV-2 3CLpro, taking advantage of the selectivity of this enzyme by a glutamine residue in the equivalent position of the cleaved polyprotein. A bicycloproline moiety, also present in boceprevir and other 3CLpro inhibitors<sup>10</sup> is present at the P2 position, which takes into account the preference of the enzyme by a hydrophobic leucine residue at P2 in the natural substrate. Larger variations are observed among different inhibitors regarding the groups present at P3 and P4 positions. At the P3 position PF-07321332 presents a tert-leucine residue, while at P4 the substituent is a trifluoromethyl group.



**Figure 1.** Chemical structure of the PF-07321332 inhibitor and the covalent complex formed after reaction with the catalytic residue of a cysteine protease. P<sub>i</sub> sites are shown.

We have used classical and hybrid QM/MM simulations to investigate the binding mode of the PF-07321332 inhibitor and its reactivity with the catalytic cysteine to yield a covalent thioimide complex:



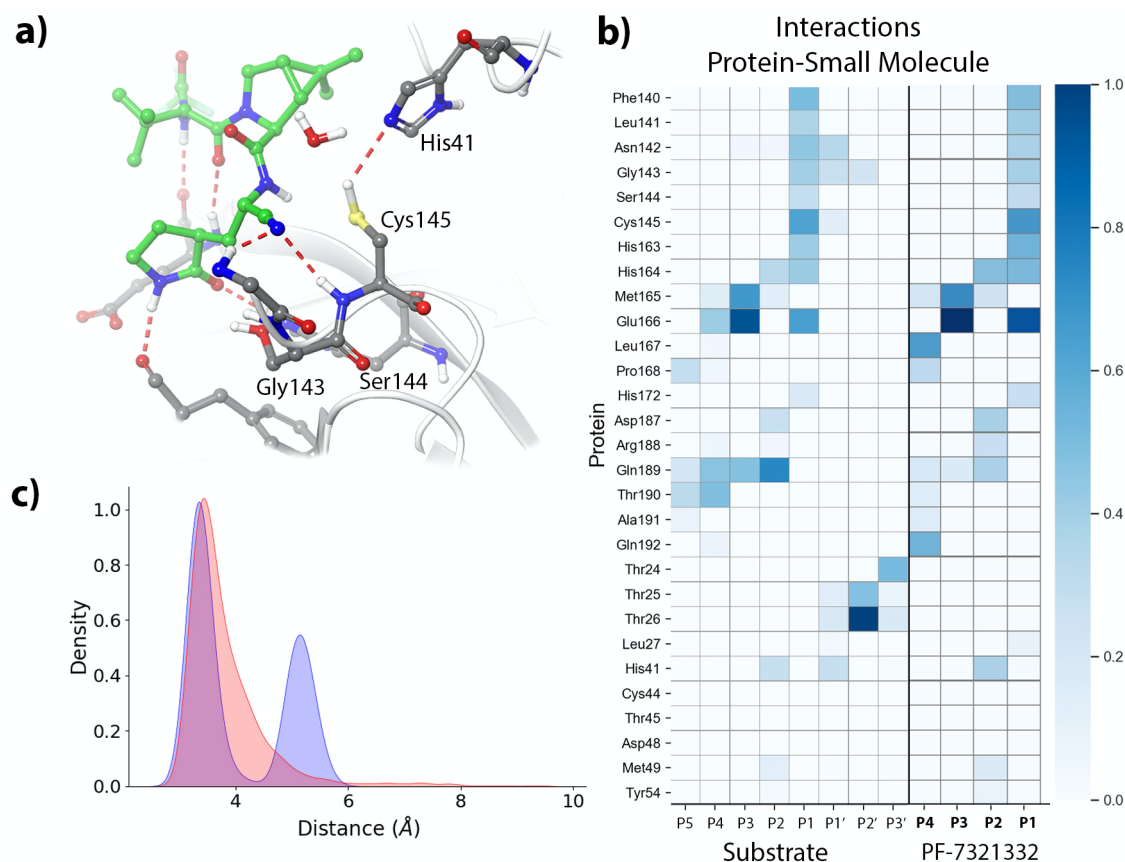
We explored the binding mode of the noncovalent complex (EI) formed by the PF-07321332 inhibitor in the active site of the SARS-CoV-2 3CL<sub>pro</sub> by means of  $\mu$ s Molecular Dynamics (MD) simulations. Alchemical transformations were then used to evaluate the contribution of the different P<sub>i</sub> groups of the inhibitor to the binding free energy of the noncovalent complex, with the goal of providing a free energy map for to guide the design of future generations of the inhibitor. Then, we explored the free energy profile associated to the reaction of the inhibitor with Cys145 to form the covalent complex, E-I in (1), characterizing the reaction mechanism and the associated activation free energy.

## Results

MD simulations (see ESI for methodological details) of the EI noncovalent complex were started from a binding pose derived from the x-ray structure of the related inhibitor PF-00835231 (PDB code: 6XHM).<sup>7</sup> 3 MD replicas of 1  $\mu$ s each show that the binding pose of the PF-07321332 inhibitor is very similar to that observed for related aldehydes or hydroxymethylketones inhibitors.<sup>11,12</sup> The nitrile group is placed in the so-called oxyanion hole, formed by interactions with the main chain NH groups of Cys145 ( $2.79 \pm 0.57$  Å), Ser144 ( $3.46 \pm 0.46$  Å) and Gly143 ( $2.54 \pm 0.34$  Å), as seen in Figure 2a. Figure 2b compares the fraction of contacts established between the inhibitor and the enzyme with those obtained from a MD simulations of a peptide substrate with sequence Ac-Ser-Ala-Val-Leu-Gln|Ser-Gly-Phe-NMe (the vertical line corresponds to the cleavage site).<sup>13</sup> The P1  $\gamma$ -lactam ring of the inhibitor almost perfectly mimics the interactions established by the P1 glutamine residue of the peptide. As said, the presence of glutamine just before the cleavage site is a

requirement in SARS-CoV and SARS-CoV-2 3CLpro<sup>14</sup> and then the presence of the  $\gamma$ -lactam ring at this position is common feature of most inhibitors of these enzymes. The S2 site of the enzyme accommodates a Leu residue in the case of the peptide substrate or a hydrophobic bicycloproline moiety in the case of the inhibitor. Again, this group reproduces the interactions made by the peptidic residue. The bicycloproline group stacks with the imidazole ring of the catalytic histidine, His41, while main chain atoms interact with Gln189. The t-butyl group at the P3 position of the inhibitor also keeps very similar interactions to those established by a valine residue in the case of the peptide. While the side chain of these P3 groups are partially solvent-exposed, main chain N and O atoms are hydrogen bonded to main chain atoms of Glu166. Larger differences are observed in the interactions made by the P4 group of the inhibitor with respect to the peptide. In the case of the inhibitor hydrogen bonds are observed between the NH main chain group of Gln192 and the fluorine atoms, while in the peptide the P4 serine residue prefers contacts with Glu166, Gln189 and Thr190.

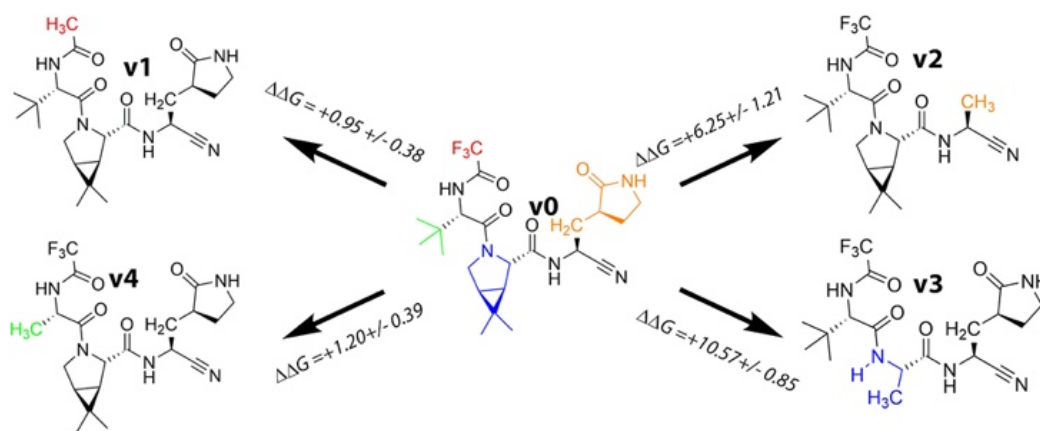
Regarding the catalytic dyad, the distance between the Cys145-S $\gamma$  atom and the electrophilic carbon atom of the inhibitor is  $4.05 \pm 0.88 \text{ \AA}$ . The pair distribution functions (see Figure 2c, blue) shows two peaks corresponding to the trans and gauche conformers of the Cys145 side chain. This residue remains most of the time hydrogen bonded to His41, as reflected by the pair distribution distance between Cys145-S $\gamma$  and His41-N $\epsilon$  atoms (Figure 2c, red) and the average distance ( $3.82 \pm 0.76 \text{ \AA}$ ). Thus, the binding pose of the inhibitor is compatible with an inhibition mechanism triggered by the proton transfer from Cys145 to His41 and the subsequent attack of the nucleophilic sulfur atom on the carbon atom of the nitrile group.<sup>11,15</sup>



**Figure 2.** Noncovalent complex formed between PF-07321332 and the 3CL protease of SARS-CoV-2. **(2a)** 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. **(2b)** Fraction of contacts between residues of PF-07321332 and a peptide substrate<sup>13</sup> and those of the protease. A contact is counted when the donor-acceptor distance is  $< 3.8 \text{ \AA}$  and the hydrogen bond angle is  $> 120^{\circ}$ . **(2c)** Pair distribution function of the distances between the  $S_{\gamma}$  atom of Cys145 and the electrophilic atom of the inhibitor (blue) and the  $N_{\epsilon}$  atom of His41 (red).

Previous description of the contacts established between the PF-07321332 inhibitor and enzymatic residues in the active site of SARS-CoV-2 3CLpro stresses the role of P1 and P2 group in the binding process. A more quantitative evaluation of the contribution of the different  $P_i$  groups of the inhibitor to the binding free energy can be obtained from *in silico* alchemical transformations. As shown in Figure 3 we evaluated the binding free energy change corresponding to the transformation of each of the  $P_i$  groups into a methyl group. Using Thermodynamic Integration, the contribution of each group to the binding free energy was obtained from the difference between the transformation carried out in the enzymatic active site and in aqueous solution, using five independent replicas, as explained in the Methodology section (see ESI). These contributions are provided in Figure 3 with their standard deviation, while the values obtained for each replica are shown in Table S1. As can be observed, the free energy changes obtained from our simulations indicate that all the  $P_i$

groups of the inhibitor make favorable contributions to the binding when compared to the methyl counterparts. Not surprisingly, the largest effects on the binding free energy appear after the substitutions of the P1 and P2 groups, 6.3 and 10.6 kcal·mol<sup>-1</sup>, respectively. These large values clearly indicate the preference of the enzyme by these two groups and the difficulties for a drastic redesign at these positions of the inhibitor. Instead, our free energy calculations suggest that there is still some room to improve the binding of the PF-07321332 inhibitor modifying the P3 and/or P4 groups. The isobutyl (P3) and trifluoromethyl (P4) moieties make favorable but small contributions to the binding free energy when compared to methyl groups, 1.2 and 0.9 kcal·mol<sup>-1</sup>, respectively.



**Figure 3.** Binding free energy changes associated to the alchemical transformation of the different  $P_i$  groups of the PF-07321332 inhibitor into methyl groups.

After formation of the noncovalent complex, the inhibition process of the 3CLpro by the PF-07321332 inhibitor continues with the formation of a covalent bond between the  $S_\gamma$  atom of the catalytic cysteine and the electrophilic carbon atom of the nitrile group. This process requires of the activation of the nucleophile by means of a proton transfer from the thiol group to the neighbor catalytic histidine, His41. To complete the reaction, after the nucleophilic attack a proton must be transferred from His41 to the nitrogen atom of the inhibitor's nitrile group. In our previous research on 3CLpro inhibition by N3, aldehyde or ketone derivatives we found that this proton transfer must be mediated by a water molecule<sup>11,15</sup> or a hydroxyl group.<sup>12</sup> In the case of the present inhibitor, we also found a water molecule that can act as a proton shuttle between His41 and the inhibitor. We determined the Minimum Free Energy Path (MFEP) for the transformation from the noncovalent inhibitor (EI) into the covalent product (E-I) using the Adaptive String Method (ASM) and a QM/MM description. The QM subsystem, composed of the side chains of Cys145 and His41, a water molecule and the warhead of the inhibitor, was described at the B3LYPD3/6-31+G\* level. The MFEP was determined exploring a free energy landscape defined by 7 Collective Variables (CVs) that include

the distances of all the bonds being broken, formed or whose formal order is changed during the process, such as the C-N bond of the nitrile group (see Figure 4). The Potential of Mean Force (PMF) of the reaction was finally determined as a function of a path CV (denoted as  $s$ ) that measures the advance of the system along the MFEP (see ESI for details). The resulting free energy profile and the evolution of the CVs along the reaction is shown in Figures 4a and 4b. The inhibition mechanism begins with the proton transfer from Cys145 to His41 to yield the catalytic dyad ion pair (IP1). The process continues with the approach of the negatively charged Cys145 cysteine to the electrophilic center of the inhibitor (the  $S_{\gamma}$ -C distance decreases from 3.0 to 2.1 Å) and its separation from the positively charged His41 (the distance between the  $S_{\gamma}$  atom and the transferred proton increases from 1.8 to 2.9 Å). The free energy profile shows a consistent increase associated to the separation of the catalytic dyad until a plateau region is reached (IP2, in Figure 4a). From this plateau the reaction continues up to the rate-determining Transition State (TS), where the nucleophilic attack is almost completed (the  $S_{\gamma}$ -C distance is reduced to 1.9 Å, see Figure 4c). At the TS there are two protons in flight: from His41 to the water molecule and from this to the nitrogen atom of the inhibitor's nitrile group (Figure 4e). Finally, at the product state the water-mediated proton transfer is completed and the thioimidate product is formed, presenting a  $S_{\gamma}$ -C distance of 1.8 Å (Figure 4f).

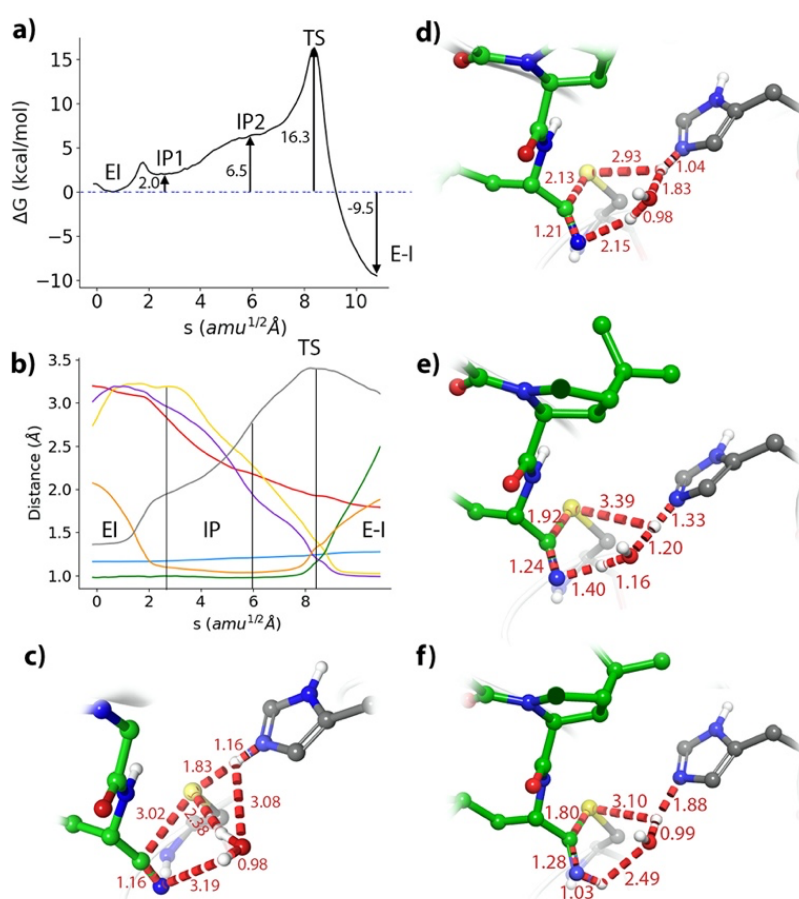
The catalytic ion pair IP1 in Figure 4a is only 2.0 kcal·mol<sup>-1</sup> above the EI complex. Significantly larger values were found for the IP obtained in the presence of other inhibitors: 9.3 kcal·mol<sup>-1</sup> for the aldehyde 11a<sup>11</sup> and 10.3 kcal·mol<sup>-1</sup> for the Michael acceptor N3.<sup>15</sup> The value found for the PF-07321332 inhibitor is close to that calculated when the IP is formed in the apo form of the enzyme, 2.9 kcal·mol<sup>-1</sup>, where the charge-separated state can be stabilized by water molecules.<sup>13</sup> The larger stabilization observed for the IP in the case of the PF-07321332 inhibitor when compared to 11a or N3 is related to smaller size of the warhead and the lack of additional substituents on the electrophilic carbon atom. This allows a closer positioning of the water molecule that simultaneously interacts with the charges on the catalytic dyad and the nitrogen atom of the nitrile group (see Figure 4c). In the case of the aldehyde derivative 11a, the presence of an additional hydrogen atom bonded to the electrophilic carbon atom impedes a closer approach of the water molecules (see Figure S1).

The activation free energy obtained for the inactivation of SARS-CoV-2 3CLpro with PF-07321332 is 16.3 kcal·mol<sup>-1</sup>. Although there are not experimental determinations of the inactivation rate constant with this inhibitor, this value can be compared to those obtained for similar inhibitors. So, the activation free energy derived from the experimental rate constant for the inhibition with the aldehyde derivative GC373 at 30° C is 21.1 kcal·mol<sup>-1</sup>,<sup>16</sup> while the value we determined theoretically for another aldehyde, 11a, was 18.5 kcal·mol<sup>-1</sup>.<sup>11</sup> In the case of the Michael acceptor N3, we used a very similar computational setup to determine a free energy barrier of 21.3 kcal·mol<sup>-1</sup>.<sup>15</sup> The larger reactivity of the PF-07321332 inhibitor correlates with the larger stability of the IP formed after the proton transfer from Cys145 to His41. As said before, the smaller size of the warhead in the case of the PF-07321332 inhibitor facilitates a closer positioning of the water molecule in charge of the



proton transfer from His41 to the nitrogen atom of the nitrile group, reducing the free energy cost associated to this proton transfer (see Figure 4e).

Regarding the reaction free energy, Figure 4a shows that the covalent complex (E-I) is more stable than the noncovalent complex (EI), being the free energy difference of  $-9.5 \text{ kcal}\cdot\text{mol}^{-1}$ . The reaction free energy obtained for the covalent inhibition of 3CLpro by PF-07321332 is larger than for inhibitors 11a ( $-2.8 \text{ kcal}\cdot\text{mol}^{-1}$ )<sup>11</sup> and PF-00835231 ( $-6.2 \text{ kcal}\cdot\text{mol}^{-1}$ ),<sup>12</sup> but smaller than for the N3 inhibitor ( $-15.4 \text{ kcal}\cdot\text{mol}^{-1}$ ).<sup>15</sup> This means that inhibition of SARS-CoV-2 3CLpro with PF-07321332 is more irreversible than with compounds that have a carbonyl group as a warhead but less than using Michael acceptors.



**Figure 4.** Formation of the thioimidate covalent product (**4a**) B3LYPD3/6-31+G\*/MM free energy profile along the path-CV for the formation of the covalent E-I complex from the noncovalent (EI) one. (**4b**) Evolution of the selected CVs along the MFEP. Blue: C-N, orange: H-N $\epsilon$ , green: Hw-Ow, black: S $\gamma$ -H, yellow: N-Hw, red: S $\gamma$ -C, purple: Ow-H. (**4c**) Representation of IP1. The values of the distances correspond (in  $\text{\AA}$ ) to the coordinates of the MFEP. (**4d**) representation of IP2. (**4e**) Representation of TS. (**4f**) Representation of the thioimidate complex.

## Conclusions

Our simulations have thus revealed some important characteristics of the novel SARS-CoV-2 3CLpro inhibitor PF-07321332. This inhibitor acts in two stages: i) formation of the noncovalent complex with the enzyme and ii) reaction with the catalytic cysteine to form a thioimidate. Regarding the binding process, our MD simulations and free energy calculations demonstrate that most of the affinity between the inhibitor and the enzyme is due to the P1 and P2 groups, which are the same (the  $\gamma$ -lactam ring at P1 position) or similar (the hydrophobic group at P2) to those present in most of the known inhibitors of this protease. Instead, P3 and P4 also contribute to the binding process, but with a significantly smaller free energy contribution, suggesting that it could still be possible to improve the affinity for 3CLpro. Regarding the nitrile warhead, its small size seems to be an advantage for the covalent inhibition. The closer positioning of a water molecule in the active site stabilizes the ion pair formed during the process, lowering also the activation free energy and increasing then the rate of the process. The thioimidate obtained after inhibition is also significantly more stable than the initial noncovalent complex, suggesting that this inhibition is more irreversible than with aldehyde or 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. We also acknowledge the use of the Tirant supercomputer at the Universitat de València and the support of Alejandro Soriano from Servei d'Informàtica from the Universitat de València.

## References

- 1 C. R. M. Bangham, *J. Gen. Virol.*, 2003, **84**, 3177–3189.
- 2 T. Pillaiyar, M. Manickam, V. Namasivayam, Y. Hayashi and S.-H. Jung, *J. Med. Chem.*, 2016, **59**, 6595–6628.
- 3 Z. Jin, X. Du, Y. Xu, Y. Deng, M. Liu, Y. Zhao, B. Zhang, X. Li, L. Zhang, C. Peng, Y. Duan, J. Yu, L. Wang, K. Yang, F. Liu, R. Jiang, X. Yang, T. You, X. Liu, X. Yang, F. Bai, H. Liu, X. Liu, L. W. Guddat, W. Xu, G. Xiao, C. Qin, Z. Shi, H. Jiang, Z. Rao and H. Yang, *Nature*, 2020, **582**, 289–293.
- 4 L. Zhang, D. Lin, X. Sun, U. Curth, C. Drosten, L. Sauerhering, S. Becker, K. Rox and R. Hilgenfeld, *Science.*, 2020, **368**, 409–412.
- 5 W. Dai, B. Zhang, X.-M. Jiang, H. Su, J. Li, Y. Zhao, X. Xie, Z. Jin, J. Peng, F. Liu, C. Li, Y. Li, F. Bai, H. Wang, X. Cheng, X. Cen, S. Hu, X. Yang, J. Wang, X. Liu, G. Xiao, H. Jiang, Z. Rao, L.-K. Zhang, Y. Xu, H. Yang and H. Liu, *Science.*, 2020, **368**, 1331–1335.
- 6 W. Vuong, M. B. Khan, C. Fischer, E. Arutyunova, T. Lamer, J. Shields, H. A. Saffran, R. T. McKay, M. J. van Belkum, M. A. Joyce, H. S. Young, D. L. Tyrrell, J. C. Vederas and M. J. Lemieux, *Nat. Commun.*, 2020, **11**, 4282.
- 7 R. L. Hoffman, R. S. Kania, M. A. Brothers, J. F. Davies, R. A. Ferre, K. S. Gajiwala, M. He, R. J. Hogan, K. Kozminski, L. Y. Li, J. W. Lockner, J. Lou, M. T. Marra, L. J. Mitchell, B. W. Murray, J. A. Nieman, S. Noell, S. P. Planken, T. Rowe, K. Ryan, G. J. Smith, J. E. Solowiej, C. M. Steppan and B. Taggart, *J. Med. Chem.*, 2020, **63**, 12725–12747.
- 8 <https://www.pfizer.com/news/press-release/press-release-detail/pfizer-initiates-phase-1-study-novel-oral-antiviral>.
- 9 R. Löser, K. Schilling, E. Dimmig and M. Gütschow, *J. Med. Chem.*, 2005, **48**, 7688–7707.
- 10 J. Qiao, Y.-S. Li, R. Zeng, F.-L. Liu, R.-H. Luo, C. Huang, Y.-F. Wang, J. Zhang, B. Quan, C. Shen, X. Mao, X. Liu, W. Sun, W. Yang, X. Ni, K. Wang, L. Xu, Z.-L. Duan, Q.-C. Zou, H.-L. Zhang, W. Qu, Y.-H.-P. Long, M.-H. Li, R.-C. Yang, X. Liu, J. You, Y. Zhou, R. Yao, W.-P. Li, J.-M. Liu, P. Chen, Y. Liu, G.-F. Lin, X. Yang, J. Zou, L. Li, Y. Hu, G.-W. Lu, W.-M. Li, Y.-Q. Wei, Y.-T. Zheng, J. Lei and S. Yang, *Science.*, 2021, 1374–1378.
- 11 C. A. Ramos-Guzmán, J. J. Ruiz-Pernía and I. Tuñón, *ACS Catal.*, 2021, **11**, 4157–4168.
- 12 C. A. Ramos-Guzmán, J. J. Ruiz-Pernía and I. Tuñón, *ChemRxiv*. DOI:10.26434/chemrxiv.13340939.v1
- 13 C. A. Ramos-Guzmán, J. J. Ruiz-Pernía and I. Tuñón, *ACS Catal.*, 2020, **10**, 12544–12554.
- 14 R. Hilgenfeld, *FEBS J.*, 2014, **281**, 4085–4096.
- 15 C. A. Ramos-Guzmán, J. J. Ruiz-Pernía and I. Tuñón, *Chem. Sci.*, 2021, **12**, 3489–3496.
- 16 C. Ma, M. D. Sacco, B. Hurst, J. A. Townsend, Y. Hu, T. Szeto, X. Zhang, B. Tarbet, M. T. Marty, Y. Chen and J. Wang, *Cell Res.*, 2020, **30**, 678–692.

# Supporting Information

## **Inhibition Mechanism of SARS-CoV-2 Main Protease with Ketone-Based Inhibitors Unveiled by Multiscale Simulations. Insights for Improved Designs**

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](mailto:ignacio.tunon@uv.es)

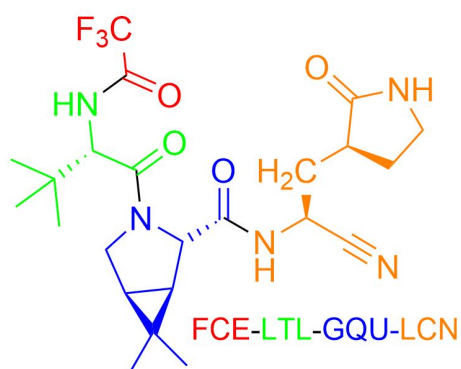
[j.javier.ruiz@uv.es](mailto:j.javier.ruiz@uv.es)

Methodological details	S2
Table S1. Results for alchemical transformations	S5
Figure S2. Superposition of Ion Pair structures	S6
References	S7

## Methodological Details

### *PF-07321332 inhibitor parameterization*

The inhibitor was divided into four different residues (Scheme S1) and each one of them was parameterized following the non-standard residue parameterization procedure implemented in Amber with the Antechamber program<sup>1</sup> from the AmberTools18<sup>2</sup> package. For this procedure the residues were capped using the corresponding ACE and NME groups.



**Scheme S1.** Partition of the inhibitor used for its parametrization.

The geometry optimization of every residue was carried out in vacuo using B3LYP/6-31G\*. The electrostatic potential for this optimized geometry was calculated using the Restrained Electrostatic Potential (RESP)<sup>3</sup> method at the HF/6-31G\* level. Atom types and parameters for bonds, angles, dihedrals, improper torsions, and non-bonded interactions were obtained using antechamber.<sup>1</sup> The only modification to those parameters was the value reported for Cabaleiro and Rios<sup>4</sup> for the linear angle formed by the nitrile group.

### *Classical Molecular Dynamics Simulations*

The Enzyme-Inhibitor (EI) system was built using the structure with PDB code 6XHM<sup>5</sup> as template. It contains the structure of the dimeric form of SARS-CoV-2 3CLpro covalently bonded to the inhibitor PF-00835231. The H-bond assignment of the protein was made using the protein preparation wizard tool from Maestro<sup>6</sup> and the protonation states for the amino acid side chains in the enzyme were obtained with PROPKA3.0<sup>7</sup> from the aforementioned tool. For neutral histidine residues, the  $\delta/\epsilon$  protonation state was confirmed after visual inspection of the x-ray structure. Standard amino acids were described using the ff14SB forcefield. The inhibitor PF-07321332 was built in the two active sites aligning its backbone atoms the corresponding atoms in the PF-00835231 inhibitor in the structure 6XHM. A water box around the resulting complex was added in such a way that protein and inhibitor atoms were found at least 12 Å away from the limits of the simulation box. Na<sup>+</sup> ions were added to neutralize the charge of the system.

The solvated system was minimized using a 500 steps of steepest descent algorithm followed by conjugate gradient minimization steps the RMS gradient was below  $10^{-3}$  kcal·mol<sup>-1</sup>Å<sup>-1</sup>. The temperature of the system was raised from 0 to 300 K using a heating ramp of 2.5 K·ps<sup>-1</sup> in the NPT ensemble. After reaching the target temperature of 300 K the system run along 20 ps. During the heating process the protein backbone atoms were restrained using a harmonic restraint with a force constant of 20 kcal·mol<sup>-1</sup>Å<sup>-2</sup>. The system continued the equilibration process in the NPT ensemble during 6.25 ns where the pressure of the system was controlled using the Berendsen barostat and the temperature was controlled by the Langevin thermostat. During this simulation time, the force constant of the harmonic potential was decreased by 3 units every 1.25 ns, from 15 during the first 1.25 ns until a restraint free run in the last 1.25 ns of simulation. 3 replicas 1μs each in the NVT ensemble were run in order to guarantee enough sampling. Shake<sup>8</sup> was used to freeze the bonds involving hydrogen atoms, making possible to use a time step of 2 fs during the simulations. Particle Mesh Ewald<sup>9,10</sup> was used to treat electrostatic interactions and the short-range interactions were calculated between particles using a cutoff radius of 10 Å. All Classical molecular dynamics simulations were simulated in the GPU version of PMEMD in AMBER.<sup>11,12</sup>

### *Alchemical Transformations*

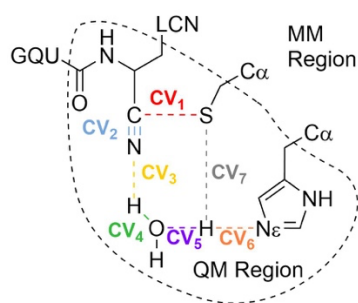
The Amber thermodynamic integration protocol reported by Xibing et al<sup>13</sup> was used here to evaluate the changes in the binding free energy related to ligands modifications ( $\Delta\Delta G_{\text{bind}}$ ). The  $\Delta\Delta G_{\text{bind}}$  values were calculated as the difference in the average values for the transformations of the ligands in aqueous solution and in the protein. Every modification was equilibrated along 200 ps at a  $\lambda$  value of 0.5 using the CPU version of AMBER pmemd. From the last 100 ps a frame of the trajectory was extracted each 20 ps of simulation. From each of those frames a replica was performed (five in total). In each replica, 9  $\lambda$  values corresponding to the gaussian quadrature schedule were selected (0.01592, 0.08198, 0.19331, 0.33787, 0.5, 0.66213, 0.80669, 0.91802 and 0.98408). After this equilibration for every  $\lambda$  value 5 ns of simulation were made. The structure at 3.0 ns of lambda 0.5 was used as the starting point for the previous (0.33787) and the next (0.66213)  $\lambda$  values in the series. This consecutive scheme of simulations was repeated until the first and last  $\lambda$  values, 0.01592 and 0.98408, were reached. In order to compute the average value of  $dU/d\lambda$  the first ns of simulation was considered as an equilibration run and Thermodynamic Integration was made using the last 4 ns of simulation of each  $\lambda$ .

### *QM/MM calculations*

To obtain the minimum free energy pathway (MFEP) along the chemical reaction the Adaptive String Method (ASM)<sup>14</sup> was used. With this method it is possible to obtain the MFEPs on multidimensional free energy surfaces (FESs) of large dimensionality. Scheme S2 shows the 7 collective variables (CVs) used to describe the chemical transformation under study. We included as CVs the distances of all the bonds whose formal order is changed during the chemical reaction. In order to obtain the MFEP on the multidimensional FES the string is discretized in nodes (96 were used string nodes were used in this case) and every node is displaced according to their free energy gradient but keeping them equidistant along the path. Along this convergence process, every 50 steps of simulation replica exchange is performed

among the nodes, increasing the speed of the convergence towards the MFEP. The string was considered to be converged to the MFEP when displays a RMSD below  $0.2 \text{ amu}^{1/2} \cdot \text{\AA}$  for at least 2 ps. With the string converged a path-CV is defined as the reaction coordinate ( $s$ ) to measure the advance of the system along the MFEP. This path-CV is used to trace the corresponding reaction free energy profile. 10 ps of simulations QM/MM were done for every node and the Weighted Histogram Analysis Method was used as method of integration. In the ASM the force constant values used to bias the string are obtained on-the-fly to ensure a homogeneous probability density distribution of the reaction coordinate. The time step used during the simulations was 1 fs and the mass of the protons involved in the chemical reaction were changed to 2 amu.

The QM region was described using the B3LYP functional with a 6-31+G\* basis set and D3 dispersion corrections. Previous studies<sup>15-17</sup> show this is a good choice to evaluate the activation energies for this kind of systems with results in good agreement with the experimental data. To run the string a modified version of AMBER18<sup>2,18</sup> with Gaussian16<sup>19</sup> for DFT calculations was used.. The collective variables used, and the atoms included in the QM region are shown in the Scheme S2.

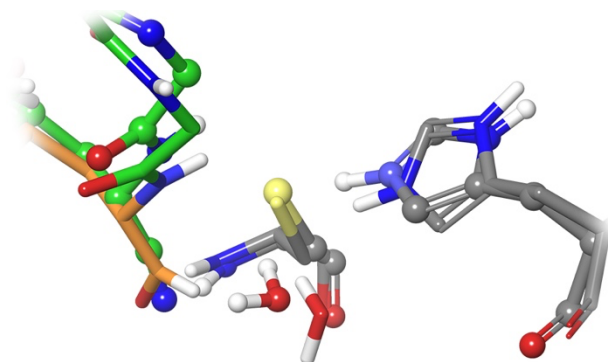


**Scheme S2.** Definition of the QM subsystem and Collective variables used in the string calculations,

**Table S1.** Free energy changes associated to alchemical transformations performed in aqueous and protein environments (see Figure 3 in main text). Free energy values (in kcal·mol<sup>-1</sup>) were estimated using Thermodynamic Integration and each average value is given with the corresponding standard deviation.

V0 to V1		$\Delta\Delta G_{\text{bind}}$				0.95	$\pm 0.38$
System	Replica	$\Delta G$	System	Replica	$\Delta G$		
Aqueous	1	-53.67	Protein	1	-52.79		
Aqueous	2	-53.82	Protein	2	-52.17		
Aqueous	3	-53.72	Protein	3	-52.78		
Aqueous	4	-53.79	Protein	4	-53.36		
Aqueous	5	-53.78	Protein	5	-52.92		
Mean		-53.76	Mean		-52.8		
std		0.06	std		0.38		
V0 to V2		$\Delta\Delta G_{\text{bind}}$				6.25	$\pm 1.21$
System	Replica	$\Delta G$	System	Replica	$\Delta G$		
Aqueous	1	0.32	Protein	1	8.05		
Aqueous	2	0.80	Protein	2	5.70		
Aqueous	3	1.00	Protein	3	8.65		
Aqueous	4	1.06	Protein	4	6.10		
Aqueous	5	1.62	Protein	5	7.57		
Mean		0.96	Mean		7.21		
std		0.42	std		1.13		
V0 to V3		$\Delta\Delta G_{\text{bind}}$				10.57	$\pm 0.85$
System	Replica	$\Delta G$	System	Replica	$\Delta G$		
Aqueous	1	-26.70	Protein	1	-18.20		
Aqueous	2	-28.00	Protein	2	-16.64		
Aqueous	3	-27.87	Protein	3	-16.78		
Aqueous	4	-27.29	Protein	4	-16.90		
Aqueous	5	-27.56	Protein	5	-16.04		
Mean		-27.48	Mean		-16.91		
std		0.46	std		0.71		
V0 to V4		$\Delta\Delta G_{\text{bind}}$				1.20	$\pm 0.39$
System	Replica	$\Delta G$	System	Replica	$\Delta G$		
Aqueous	1	1.38	Protein	1	3.24		
Aqueous	2	1.85	Protein	2	2.66		
Aqueous	3	1.48	Protein	3	2.27		
Aqueous	4	1.61	Protein	4	3.07		
Aqueous	5	1.47	Protein	5	2.56		
Mean		1.56	Mean		2.76		
std		0.16	std		0.35		





**Figure S1.** Overlap of the Ion Pair structures of the SARS-CoV-2 3CLpro active site obtained with PF-07321332 inhibitor (green, CPK) and 11a (orange, licorice). Note that the presence of the hydrogen atom bonded to the carbonyl electrophilic carbon in 11a hinders the approach of the water molecule.

## References

- 1 J. Wang, W. Wang, P. A. Kollman and D. A. Case, *J. Mol. Graph. Model.*, 2006, **25**, 247–260.
- 2 D. A. Case, D. S. Cerutti, T. E. I. Cheatham, T. A. Darden, R. E. Duke, T. J. Giese, H. Gohlke, A. W. Goetz, D. Greene, N. Homeyer, S. Izadi, A. Kovalenko, T. S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. Mermelstein, K. M. Merz, G. Monard, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D. R. Roe, A. Roitberg, C. Sagui, C. L. Simmerling, W. M. Botello-Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu, L. Xiao, D. M. York and P. A. Kollman, *Univ. California, San Fr.*
- 3 C. I. Bayly, P. Cieplak, W. Cornell and P. A. Kollman, *J. Phys. Chem.*, 1993, **97**, 10269–10280.
- 4 E. M. Cabaleiro-Lago and M. A. Ríos, *J. Phys. Chem. A*, 1997, **101**, 8327–8334.
- 5 R. L. Hoffman, R. S. Kania, M. A. Brothers, J. F. Davies, R. A. Ferre, K. S. Gajiwala, M. He, R. J. Hogan, K. Kozminski, L. Y. Li, J. W. Lockner, J. Lou, M. T. Marra, L. J. Mitchell, B. W. Murray, J. A. Nieman, S. Noell, S. P. Planken, T. Rowe, K. Ryan, G. J. Smith, J. E. Solowiej, C. M. Steppan and B. Taggart, *J. Med. Chem.*, 2020, **63**, 12725–12747.
- 6 Schrödinger Release 2021-2: Maestro, Schrödinger, LLC, New York, NY, 2021.
- 7 M. H. M. Olsson, C. R. Søndergaard, M. Rostkowski and J. H. Jensen, *J. Chem. Theory Comput.*, 2011, **7**, 525–537.
- 8 J.-P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, *J. Comput. Phys.*, 1977, **23**, 327–341.
- 9 T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089–10092.
- 10 U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577–8593.
- 11 S. Le Grand, A. W. Götz and R. C. Walker, *Comput. Phys. Commun.*, 2013, **184**, 374–380.
- 12 R. Salomon-Ferrer, A. W. Götz, D. Poole, S. Le Grand and R. C. Walker, *J. Chem. Theory Comput.*, 2013, **9**, 3878–3888.
- 13 X. He, S. Liu, T. Lee, B. Ji, V. H. Man, D. M. York and J. Wang, *ACS Omega*, 2020, **5**, 4611–4619.
- 14 K. Zinovjev, 2017, 1–3.
- 15 C. A. Ramos-Guzmán, J. J. Ruiz-Pernía and I. Tuñón, *Chem. Sci.*, 2021, **12**, 3489–3496.
- 16 C. A. Ramos-Guzmán, J. J. Ruiz-Pernía and I. Tuñón, *ACS Catal.*, 2020, **10**, 12544–12554.
- 17 C. A. Ramos-Guzmán, J. J. Ruiz-Pernía and I. Tuñón, *ACS Catal.*, 2021, **11**, 4157–4168.
- 18 K. Zinovjev, String-Amber, <https://github.com/kzinovjev/string-amber>, (accessed 24 June 2020).
- 19 Gaussian 16, Revision C.01, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. Gaussian, Inc., Wallingford CT, 2016.