Electronic Supplementary Material (ESI) for ChemComm.

Supplementary Information

Computational Simulations on the Binding and Reactivity of a Nitrile Inhibitor of SARS-CoV-2 Main Protease

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:

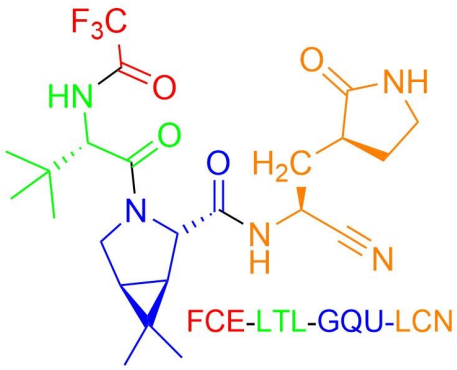
ignacio.tunon@uv.es

j.javier.ruiz@uv.es

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[1] from the AmberTools18[2] 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)[3] 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.[1] The only modification to those parameters was the value reported for Cabaleiro and Rios[4] 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[5] 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[6] and the protonation states for the amino acid side chains in the enzyme were obtained with PROPKA3.0[7] from the aforementioned tool. For neutral histidine residues, the δ/ε 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^+ ions were added to neutralize the charge of the system.

The solvated system was minimized using a 500 steps pf steepest descent algorithm followed by conjugate gradient minimization steps the RMS gradient was below 10^{-3} kcal·mol^{-1}Å^{-1}. The temperature of the system was raised from 0 to 300 K using a heating ramp of 2.5 K·ps^{-1} 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^{-1}Å^{-2}. 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[8] 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[9,10] 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.[11,12]

Alchemical Transformations

The Amber thermodynamic integration protocol reported by Xibing et al[13] was used here to evaluate the changes in the binding free energy related to ligands modifications ($\Delta\Delta G_{bind}$). The $\Delta\Delta G_{bind}$ 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 λ 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 λ 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 λ 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) λ values in the series. This consecutive scheme of simulations was repeated until the first and last λ 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 λ.

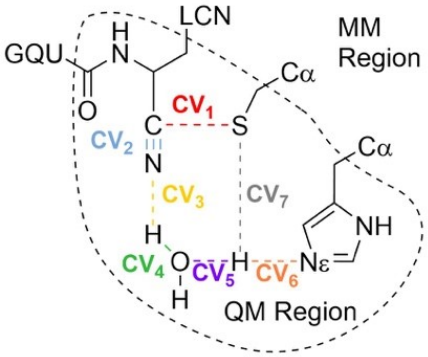
QM/MM calculations

To obtain the minimum free energy pathway (MFEP) along the chemical reaction the Adaptative String Method (ASM)[14] 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 amu$^{1/2}$·Å 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 composed of the side chains of Cys145 and His41, a water molecule and the warhead of the inhibitor (see Scheme S2) and it was described using the B3LYP functional with a 6-31+G* basis set and D3 dispersion corrections. Our previous studies on the same enzyme with different substrates (including a peptide and different inhibitors)[15–17] show that this choice for the size of the QM region and level provides free energy profiles in good agreement with the experimental data. This computational set-up is thus a reasonable compromise between accuracy and cost. A previous study on cysteine-histidine proton transfer found that the B3LYP functional was the best choice to obtain an electronic description in agreement with higher level methods.[18] Some concerns have been reported regarding the ability of the B3LYP functional to describe correctly enolate and carbanion intermediates.[19,20] However, this functional does a good work when the nucleophilic attack takes place accompanied by a proton transfer to the substrate (water mediated in our mechanism). We showed that the results obtained with the B3LYP and M06-2X functionals were very similar for the inhibition with a Michael acceptor inhibitor that presents a reaction mechanism very similar to that of the present inhibitor.[15] Regarding the basis set we also showed that diffuse functions provide a free energy barrier in better agreement with experimental estimates for the hydrolysis of a peptide bond.[16] The computational cost needed to converge a string and obtain the associated free energy profile with the current computational set-up is roughly 1 million hours of CPU on Xeon Platinum processors, which makes almost unaffordable the use of larger QM regions or larger basis sets. Instead, dispersion corrections introduce a negligible computational cost, while they can improve the description obtained with the functional for the nucleophlic attack.[19]

To run the string a modified version of AMBER18[21] with Gaussian16[22] for DFT calculations was used. The collective variables used, and the atoms included in the QM region are shown in 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^{-1}) were estimated using Thermodynamic Integration and each average value is given with the corresponding standard deviation.

V0 to V1		$\Delta\Delta G_{bind}$			0.95 ± 0.38
System	Replica	ΔG	System	Replica	ΔG
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		$\Delta\Delta G_{bind}$			6.25 ± 1.21
System	Replica	ΔG	System	Replica	ΔG
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		$\Delta\Delta G_{bind}$			10.57 ± 0.85
System	Replica	ΔG	System	Replica	ΔG
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		$\Delta\Delta G_{bind}$			1.20 ± 0.39
System	Replica	ΔG	System	Replica	ΔG
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

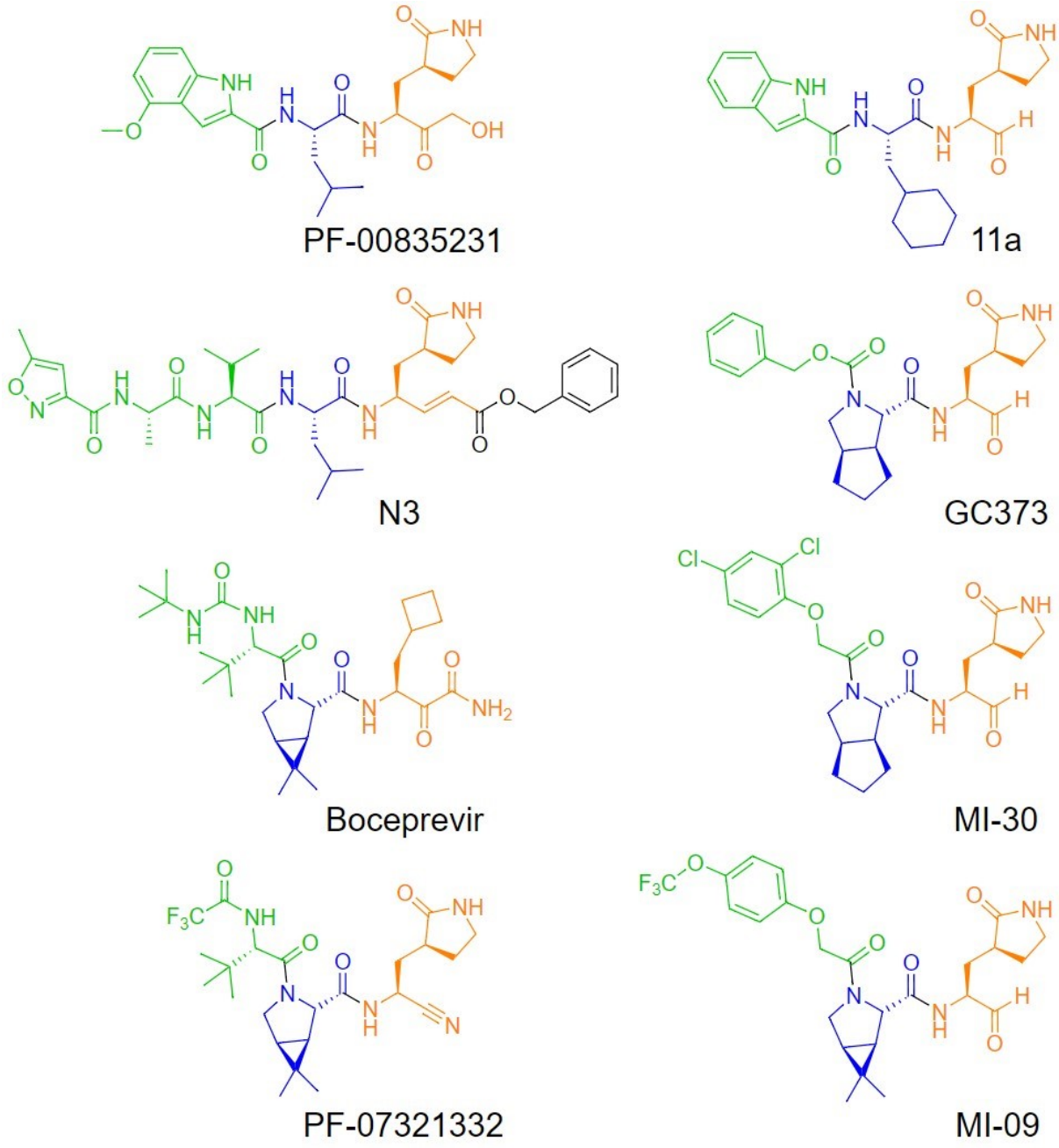


Figure S1. Some of the known inhibitors of the main protease of SARS-CoV-2. P1 group is shown in orange, P2 in blue and P3/P4 in green.

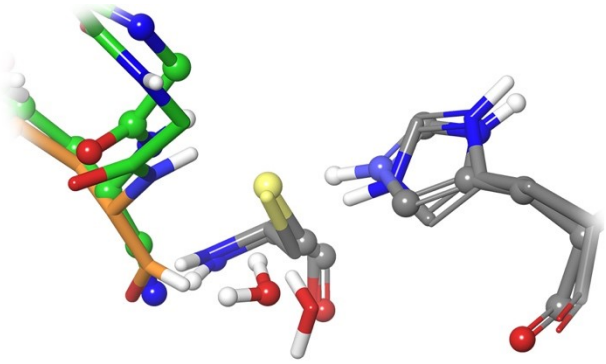


Figure S2. Overlap of the Ion Pair structures of the SARS-CoV-2 3CLpro active site obtained with PF-07321332 inhibitor (green, CPK) and 11a (orange, licorice). Note that the presence of the hydrogen atom bonded to the carbonyl electrophilic carbon in 11a hinders the approach of the water molecule.

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 and I. Tuñón, *J. Phys. Chem. A*, 2017, **121**, 9764–9772.

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 A. Paasche, T. Schirmeister and B. ENgels, *J. Chem. Theory Comput.* **2013**, *9*, 1765–1777.

19 E. Awoonor-Williams, W. C. Isley III, S. G. Dale, E. R. Johnson, H. Yu, A. D. Becke, B. Roux, C. N. Rowley, *J. Comput. Chem.* **2020**, 41, 427–438.

20 J. M. Smith, Y. J. Alahmadi and C. N. Rowley, *J. Chem. Theory Comput.* **2013**, *9*, 4860–4865.

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

22 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.