

Departamento de Química Física
Instituto de Ciencia Molecular

Doctorado en Química Teórica y Modelización Computacional



**Estudio del mecanismo de reacción e inhibición de cisteína
proteasas mediante métodos multiescala**

Carlos Alberto Ramos Guzmán

Directores

Dr. D. Ignacio Nilo Tuñón García de Vicuña

Dr. D. Jose Javier Ruiz Pernía

Dr. D. Kirill Zinovjev

Enero 2022

Ignacio Nilo Tuñón García de Vicuña, Catedrático de Química Física del Departamento de Química Física de la Universitat de València; Jose Javier Ruiz Pernía, Profesor Titular de Química Física del Departamento de Química Física de la Universitat de València y Kirill Zinovjev, investigador asociado de la escuela de bioquímica de la Universidad de Bristol.

CERTIFICAN:

Que el trabajo con título: 'Estudio del mecanismo de reacción e inhibición de cisteína proteasas mediante métodos multiescala' ha sido realizado por Carlos Alberto Ramos Guzmán bajo nuestra dirección, para optar al grado de Doctor en Química.

Así, autorizamos la presentación de este trabajo a efectos de seguir los trámites correspondientes de la Universitat de València.

Valencia, 22 de Noviembre de 2021

DIRECTORES



Dr. D. Ignacio Nilo Tuñón
García de Vicuña



Dr. D. Jose Javier Ruiz Pernía



Dr. D. Kirill Zinovjev

Esta tesis se presenta como compendio de publicaciones. De acuerdo con la normativa vigente de la Universidad de Valencia este documento incluye un resumen de la temática tratada en los artículos, los resultados y las conclusiones obtenidas. Posteriormente se incluye un anexo que contiene una copia completa de los trabajos mencionados a continuación:

- Ramos-Guzmán, C. A.; Zinovjev, K.; Tuñón, I. Modeling Caspase-1 Inhibition: Implications for Catalytic Mechanism and Drug Design. *Eur. J. Med. Chem.* 2019, 169, 159–167.
- Ramos-Guzmán, C. A.; Ruiz-Pernía, J. J.; Tuñón, I. Unraveling the SARS-CoV-2 Main Protease Mechanism Using Multiscale Methods. *ACS Catal.* 2020, 10 (21), 12544–12554.
- Ramos-Guzmán, C. A.; Ruiz-Pernía, J. J.; Tuñón, I. A Microscopic Description of SARS-CoV-2 Main Protease Inhibition with Michael Acceptors. Strategies for Improving Inhibitor Design. *Chem. Sci.* 2021, 12 (10), 3489–3496.
- Ramos-Guzmán, C. A.; Ruiz-Pernía, J. J.; Tuñón, I. Multiscale Simulations of SARS-CoV-2 3CL Protease Inhibition with Aldehyde Derivatives. Role of Protein and Inhibitor Conformational Changes in the Reaction Mechanism. *ACS Catal.* 2021, 11 (7), 4157–4168.
- Ramos-Guzmán, C. A.; Ruiz-Pernía, J. J.; Tuñón, I. Computational Simulations on the Binding and Reactivity of a Nitrile Inhibitor of the SARS-CoV-2 Main Protease. *Chem. Commun.* 2021, 2–5.
- Ramos-Guzmán, C. A.; Ruiz-Pernía, J. J.; Tuñón, I. Inhibition Mechanism of SARS-CoV-2 Main Protease with Ketone-Based Inhibitors Unveiled by Multiscale Simulations: Insights for Improved Designs. *Angew. Chemie Int. Ed.* 2021, 2–11.

El autor agradece el soporte financiero proveniente de:

- **Conselleria d'Educació, Investigació, Cultura i Esport.** Ajudes per a grups consolidables. Contrato de Técnico Superior de soporte a la investigación (AICO/2018/238). Duración 22/11/2018-31/07/2019
- **La Agencia Estatal de Investigación -AEI- y el Fondo Europeo de Desarrollo Regional – FEDER** Contrato de Técnico Superior de soporte a la investigación (PGC2018-094852-B-C22). Duración 22/11/2019-11/03/2021
- **Generalitat Valenciana** Contrato de Técnico Superior de soporte a la investigación (CGCOV19/Decreto180/2020) Duración 12/03/2021-31/10/2021
- **La Agencia Estatal de Investigación -AEI- y el Fondo Europeo de Desarrollo Regional – FEDER** Contrato de Técnico Superior de soporte a la investigación (PGC2018-094852-B-C22). Duración 01/11/2021-30/11/2021

Agradecimientos

A mi mamá, Berenice. Quiero agradecerle la infinita paciencia que siempre me ha tenido. Por sus enseñanzas; las que me ha dado a lo largo de toda su vida por medio de sus acciones. Porque somos lo que hacemos, no lo que decimos y mi mamá habla poco, pero hace mucho. Me enseñó que sólo hace falta tener paciencia, constancia y un motivo claro que lo empuje a uno cada día a seguir adelante.

A mi papá, Albeiro, por haberme inculcado el amor por la lectura y el estudio. Le agradezco el haberme dado el mejor consejo que recibí cuando era un niño: dedíquese a estudiar, porque si empieza a trabajar le va a coger gusto a la plata y no va a querer volver al colegio. El consejo caló bastante hondo, porque gracias a esto he vivido siempre aprendiendo algo.

A mi hermana, Diana, le agradezco, el hacerme sentir siempre que está ahí a mi lado y de mi parte. Aun en la distancia, siempre he sabido que cuento con ella y que no estoy solo. Le agradezco el haber traído al mundo a dos maravillosas niñas que me recuerdan lo bonita que es la vida y hacen que me esfuerce cada día para ser una mejor persona, una de la que se sientan orgullosas. Le agradezco a Isa, mi sobrina mayor, por ayudarme a mantener viva mi curiosidad y por forzarme siempre a pensar como un niño. A Sarita, mi sobrina pequeña—con sus pocos meses de vida—le agradezco el recordarme que la sola existencia de una persona es motivo suficiente para quererle.

A Iñaki, quién más que un director o un mentor científico fue un amigo, un hermano y un padre cuando hizo falta. De no haber sido por él, nada de esto habría sido posible. A Javier Ruiz por su invaluable ayuda y colaboración. A Kirill Zinovjev por las enriquecedoras discusiones y lecciones científicas, y por las no tan científicas también.

Al profe Gian Pietro Miscione, por haberme transmitido todo su amor por la ciencia y la investigación. Por haber creído en mí y por haberme ayudado a encontrar mi lugar en la ciencia.

A mis amigos de toda la vida Giselle Iregui, Myriam Jaramillo, Érica Osorio, Alex Silva y Mónica Grimaldos mis más profundos y sinceros agradecimientos por el apoyo y los consejos dados durante estos últimos años que me ayudaron a ver las cosas con claridad y entender que todo tiene una razón de ser. Sin ellos mi estancia en España no habría sido tan enriquecedora.

A mi novia, Azais, por su paciencia, comprensión y apoyo incondicional en este proceso. Por entender, pero especialmente por compartir, mi manera de ver el mundo y la vida.

"And all this
science I don't
understand, it's
just my job
5 days a week,
[ROCKET MAN, - ELTON
JOHN]

Table of Contents

Summary of the thesis	3
1. Introduction	3
1.1 What is a Disease?.....	3
1.1.1. Molecular Pathogenesis of Alzheimer’s Disease.....	4
1.1.2. COVID-19 molecular pathogenesis	5
1.2 Proteolytic Enzymes	6
1.2.1. Cysteine proteases	7
1.2.1.1. Caspase-1	8
1.2.1.2. 3CL ^{pro} of SARS-CoV-2	11
1.3 Enzymatic Inhibition.....	12
2. Computational methods in drug design	13
2.1 Classical Molecular Dynamics	14
2.1.1. Problem Solution from Newton Mechanics.....	15
2.1.2. Force Fields.....	16
2.2 QM/MM methodologies	16
2.3 Free energy calculations	18
2.3.1. Free Energy Differences	19
2.3.1.1. Thermodynamic integration.....	19
2.3.1.2. Molecular Mechanics Generalized Born Surface Area (MM/GBSA)	20
2.3.2. Potential of Mean Force (PMF)	21
2.3.2.1. Adaptive String Method (ASM)	22
2.3.2.2. Umbrella sampling	23
3. Results and conclusions	23
3.1 Caspase-1	23
3.2 3CL ^{pro} of SARS-CoV-2	26
3.2.1. Peptide substrate	26
3.2.2. Inhibitors	29
3.3 Final remarks.....	35
Resumen en lengua oficial.....	39
Bibliography	59
ANNEXES	69
Appendix 1.....	71
Appendix 2.....	93
Appendix 3.....	119
Appendix 4.....	145
Appendix 5.....	171
Appendix 6.....	185

Summary of the thesis

1. Introduction

The main reason that encouraged me to become a scientist was to stop Alzheimer disease (AD). As is well known, this disease affects the brain, causes memory problems, and alters the way we think and behave. But what may be unknown to many people is that in advanced stages patients with this disease lose control of their sphincters as well as their mobility. In such a way it is a condition that denigrates the moral integrity of both patients and their families. In addition, it destroys the most precious thing human beings have: our memories, which gives meaning to our existence.

AD is among the leading causes of death globally¹ and more than 50 million people around the world suffer from it, i.e. 0.6% of the population.² If no therapy is envisioned soon, AD have enough potential to significantly impact the world economy and the health system.³ The total cost for AD patients caring has been estimated about \$US 83.9 billion annually, a figure that increased by 35% from 2010 to 2015.⁴

While I was focused on AD throughout my doctoral studies, the Coronavirus disease 2019 (COVID-19) pandemic hit the world. Millions of lives were taken by this disease and many more, although they have recovered from the disease, have been left with permanent consequences that prevent them from continuing with a normal life. We are scientists, and anything we can do to alleviate human suffering we must do. For that reason, we could not turn our backs to the world. Even more, when we know that the knowledge and methodologies used so far could be also employed to contribute to the understanding of the mechanism behind the virus replication process.

As known, these two diseases are very different from each other. Nevertheless, as we will see, from the molecular point of they have something in common. But to understand this first we have to try to answer the question:

1.1 What is a Disease?

At first sight this can be seen as a trivial question. However, even though we have all been ill at some point in our life, no one can define it concisely and unequivocally. Mainly because every disease is different, with different symptoms and causes. But those are not the only reasons that hinder the task. Also, cultural background plays a significant role.

Osteoporosis, homosexuality, and attention deficit hyperactivity disorder (ADHD) will help us to understand better this point. The former was only recognized as disease in 1994 by the WHO. Previously it was seen merely as an unavoidable part of normal ageing (something similar happened to AD almost half century ago). In this way patients changed their diagnostic from normally old to ill. In the case of homosexuality, the process moved the other way around. "Patients" with this "disease" passed from receiving hormone treatments to electroshocks or neurosurgery, to be normal healthy individuals in 1973, after the American Psychiatric Association depathologized homosexuality and declared it as a normal variant in the spectrum of sexual behavior. And finally, ADHD is still something in between and there is a lot of controversy about children medication. For some, children behave badly because they have a disease that need medication. While for other this diagnosis means "children whom parents and teachers cannot control".⁵

For all those reasons and because science has discovered a way to modulate diseases with molecules we call medicines it will be better to try to understand the molecular mechanism of every disease rather than to try to give a global definition of disease. So, we must go down to the subcellular nanoworld to give an explanation that will help us understand the molecular pathways that contribute to the specific diseases involved in this study, i.e., the molecular pathogenesis of AD and COVID-19, which additionally will help us to understand how computational methods fits in the process of drug design.

1.1.1. Molecular Pathogenesis of Alzheimer's Disease

AD is a neurodegenerative disease. Nowadays there is no consensus about the exact molecular pathogenesis leading to this neuronal death, mainly because there are many processes involved along its development.⁶ Nevertheless, the amyloid cascade hypothesis proposed in 1992 by Hardy et al.⁷ is the one most commonly accepted so far (see **Figure 1**). According to this hypothesis, the dementia observed is the result of the deposition of the β -amyloid ($A\beta$) peptide into plaques in the brain. β -amyloid fibrils are produced as consequence of ineffective protein clearance in the brain.⁸ The aggregation process of these fibrils into plaques triggers the hyperphosphorylation of the tau-protein. As tau-protein modulates the stability of microtubules in the axons, its hyperphosphorylation leads to the loss of its structural properties separating it from the microtubules.⁹ Due to this separation the microtubules are disassembled leading to axonal instability. So, axons cannot perform their task of transporting information and organelles along the central nervous system (CNS) leading to cell loss, vascular damage and formation of neurofibrillary tangles of tau-protein, another characteristic hallmark of AD.¹⁰ In this way, tackling the accumulation of the $A\beta$ plaques can be seen as a logical way to treat the disease. After all, this would stop all the processes that were triggered by its accumulation. Nevertheless, it has been observed that there is no direct correlation between the amount of $A\beta$ plaques in the brain and the degree of dementia experienced in life.¹¹ Also, some humans with many $A\beta$ deposits in its brains never show symptoms of AD in life.¹²

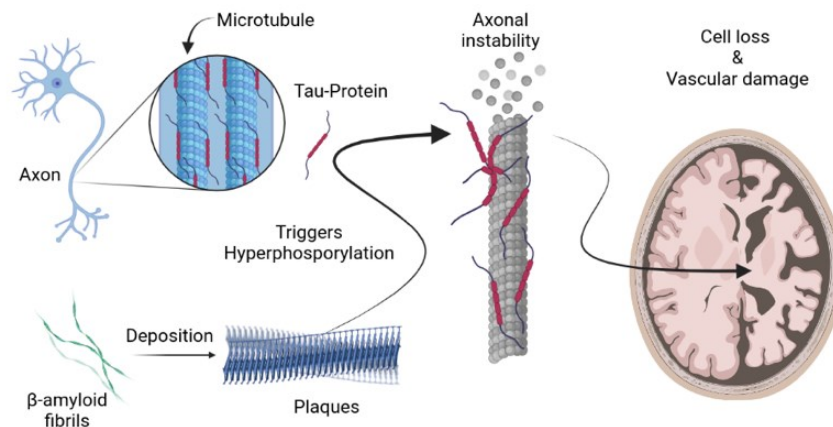


Figure 1. The Amyloid Cascade Hypothesis. This figure was created with BioRender

New hypothesis shows that there is an immune system activation during the development of the AD.^{13–15} The immune system deploys an inflammatory response to protect the affected tissue from external agents as bacteria, infections, viruses, toxins, or internal signals such as damaged cells. And this response is activated by release of proinflammatory cytokines. Through this inflammatory process, the external agent causing the signal is eliminated, necrotic cells are cleared and the tissue is repaired.¹⁶ Under normal conditions, this inflammatory process ends when the seizure has been neutralized. Nevertheless, in chronic inflammation this regulatory process is broken, and in the absence of an external offender the body attacks and destroys its own healthy cells.¹⁷ Neurons in the case of AD.

As previously mentioned, AD is a neurodegenerative disease of which an important hallmark is the formation of β -amyloid fibrils. In the inflammation pathway hypothesis, these misfolded proteins are phagocytosed by microglia¹⁸ (see **Figure 2**). As consequence, the NLRP3 inflammasome is activated, leading to the activation of the enzyme caspase-1.¹⁹ Finally, the inflammatory cytokine IL-1 β is released by the enzymatic activity of caspase-1. This IL-1 β release drives neuronal dysfunction, such as the hyperphosphorylation of the tau-protein,²⁰ leading to cellular damage and cell death. As product of this cell death, neural debris are produced, the metabolism of the Amyloid Protein Precursor (APP) is increased, more β -amyloid fibrils are formed. And all these subproducts are recognized by microglia and the process is continued once again.¹⁶ Elevated levels of active caspase-1 has been observed in brain tissues of human and mice with this disease.²¹ Additionally, this enzyme is also responsible for the characteristic inflammation observed in such diseases as atherosclerosis,²² arthritis rheumatoid,²³ acute brain injury²⁴ among others.²⁵ In this way, blocking its enzymatic activity will improve the condition of patients with other ailments.

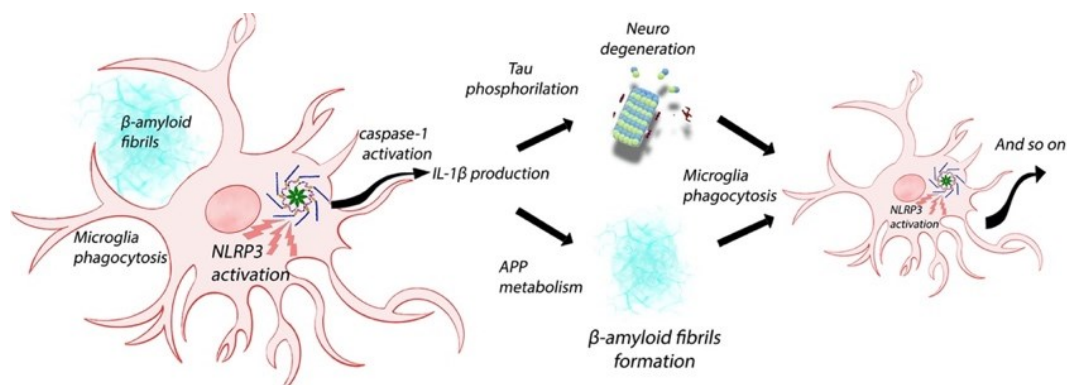


Figure 2. Role of IL-1 β Alzheimer's associated inflammation pathway.

1.1.2. COVID-19 molecular pathogenesis

Covid-19 is a viral pneumonia caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). SARS-CoV-2 is a virus single-strand with positive-sense RNA genome of 29,881 base-pair length that encode for 9860 amino acids²⁶ which form both structural and non-structural proteins required for the replication process necessary to infect the host. To start the replication cycle the RNA strand contained inside the virion (particle of the virus that causes infection) must enter into the host cells to be decoded by their translation machinery.

As show in **Figure 3**, this process is facilitated by the interaction between the receptor angiotensin-converting enzyme 2 (ACE2) in the host cell membrane and the spike protein in the surface of SARS-CoV-2.²⁷ After this recognition process the virus and host membranes pull together, and then, the RNA strand is released to the cytoplasm to be decoded. This transcoding process is the same for all the viruses in the order *Nidovirales*.²⁸ That means, one-third of the genome is transcribed to subgenomic mRNA, while the two-thirds remaining parts of the strand contain each one an open reading frame (ORF), rep1a and rep1b, which express two large coterminal polyproteins, pp1a and pp1ab. These two polyproteins encode 16 non-structural proteins (nsp1 to nsp16). Two of these non-structural proteins are the papain like protease or PLpro, (nsp3), and the 3 Chymotrypsin Like-protease or 3CL^{pro}, (nsp5). The former is in charge of cutting the nsp1/2, nsp2/3 and nsp3/4 edges, while the latter is in charge of the remaining 11 cleavages actions that will form the entire transcription-replication complex of the virus. From the transcribed subgenomic mRNA four structural proteins are transcribed: the nucleocapsid (N), the

envelope (E), the membrane (M) and the spike (S). These structural proteins are assembled to form the virion enveloped in the endoplasmic-reticulum--Golgi intermediate compartment (ERGIC) to form mature virions.^{29,30}

This process is repeated until the infected cells cannot perform their natural functions and ends with the death of the host. For that reason, the blockage of the 3CL^{pro} is an important therapeutic target against the Covid-19 disease.^{31,32} As the 3CL^{pro} active site is highly conserved among human coronaviruses the findings of this study could be extrapolated to other human CoV-like viruses.³³

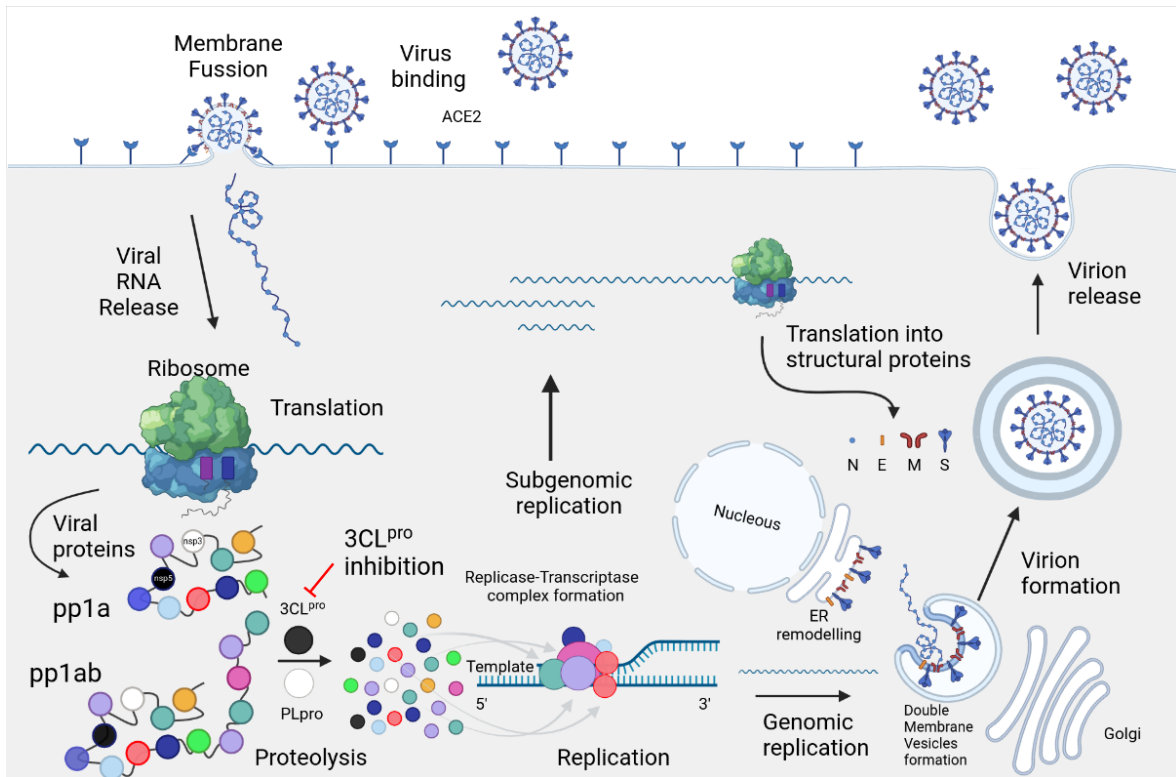


Figure 3. Replication cycle of SARS-CoV-2 in Red is shown where the cycle would finish if the 3CL^{pro} enzyme is inhibited. Nucleocapsid (N), the envelope (E), the membrane (M) and the spike (S). Created with BioRender.com

1.2 Proteolytic Enzymes

A proteolytic enzyme, or protease, is an enzyme that catalyzes the hydrolysis of peptide bonds in a protein to produce smaller proteins or single amino acids. A protease can cleavage these bonds using an amino acid side chain as nucleophile to attack the scissile peptide bond or by activating a water molecule that will consequently act as a nucleophile.³⁴ In the first group are the serine, cysteine, and threonine proteases, while in the later are the aspartic, glutamic and metalloproteases. In this second group the name of the protease is given according to the species that performs the activation of the water molecule. i.e., an aspartate, a glutamate or a metal. In metalloproteases it is a divalent metal chelated by amino acids in the enzyme that plays this role. The most common metal for metalloproteases is zinc³⁵ but there are many others as manganese, nickel, cobalt, or copper.

Proteases are a big family that contains major therapeutic targets. Well known examples of successful therapies involving proteases are the inhibition of the retropepsin³⁶ and the angiotensin-converting enzyme (ACE).³⁷ Its inhibitors are used now in treatments against acquired immunodeficiency syndrome

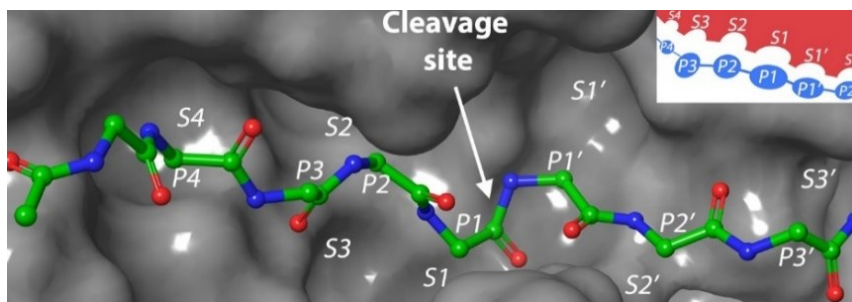
(AIDS) and hypertension, respectively. As these, there are plenty examples of proteases related to human diseases whose inhibition will improve the human health.³⁸⁻⁴³

This work was made having in mind cysteine proteases, i.e., enzymes in which the nucleophilic attack is performed for the thiol group of the side chain of a cysteine residue.

Box 1

The Schechter and Berger nomenclature

This classification was made by Schechter and Berger in 1967⁴⁴ and it is still used by its outstanding simplicity and clarity to rationalize the interaction between a substrate/inhibitor with the protease active site. Also, it allows to explain interactions that occur in distant regions from the catalytic site. In this classification the active site in the enzyme is divided into subsites S. The substrate amino acids are named with the letter P (for peptide) and numbered starting in P1 from the scissile bond up to P8 towards the N terminal side of the substrate (this part will remain bound to the enzyme after the formation of the acyl-enzyme complex). Regarding the amino acids in the carboxyl-terminal side (leaving group), they will be numbered as P1', P2', P3', etc. The subsites in the enzyme that will interact with them will be numbered in a complementary way (S and S').

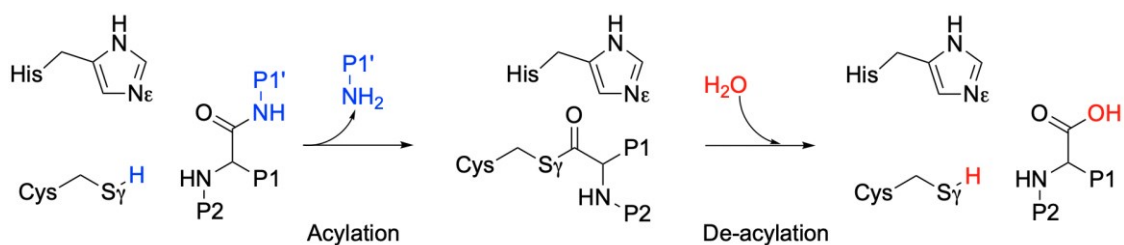


Schematic representation of the interaction sites in the active site of proteolytic enzymes, S for the sub pockets in the enzyme and P for Fragments in the substrate.

1.2.1. Cysteine proteases

Cysteine proteases are enzymes with a thiol group of a cysteine residue in charge of the proteolysis activity of the enzyme. Apart from this thiol group, cysteine proteases have an imidazolium group of a histidine forming what is known as the catalytic dyad. The putative reaction mechanism for cysteine proteases in general is described in scheme 1. This mechanism is separated in two disconnected steps: acylation, and de-acylation.⁴⁵

Cysteine proteases are associated into clans, families, subfamilies accordingly to their similar sequence motifs around the catalytic residues. For curious readers a clear explanation and classification of every one of the more than 100 cysteine peptidases found up to 2013 are summarized in the volume 2 of the Handbook of Proteolytic Enzymes.³⁴



Scheme 1. Proposed reaction mechanism for cysteine proteases.⁴⁵ being P1 the side chain of Asp in caspases and Gln in the 3CL^{pro} of SARS-CoV-2.

1.2.1.1. Caspase-1

The name caspase comes from cysteine-dependent aspartate specific protease. These enzymes are highly specific for cleaving the substrate peptide bond formed between a P1 aspartate and the P1' adjacent amino acid. The caspase family members are well differentiated from other peptidases for this strict need for an aspartate residue in this P1 position in such a way that its substitution by other residues can decrease its V_{max}/K_m (see box 2) by more than 100 times, as in the case of glutamate.⁴⁶ For the specific case of caspase-1 and their pro-IL-1 β substrate, this hydrolysis is done between the vertical lines in the sequences Asp27-|-Gly28-Pro29 and Asp116-|-Ala117-Pro118.⁴⁶ The later results in the generation of the 17.5-kDa active form of IL-1 β .

The catalytic domain of caspase-1 is a heterodimer with two subunits p10 (α subunit) and p20 (β subunit) being both derived from a common 45 kDa proenzyme.⁴⁷ A β -sheet is formed at the core of the enzyme with one antiparallel and five parallel strands. The residues forming the assumed catalytic dyad in S1 are in the p20 subunit, while the subsites S2, S3 and S4 are mainly formed by residues of the p10 subunit, see **Figure 4**. In the active form, the caspase enzymes are usually formed by a dimer of the catalytic domains in a $\alpha\beta\beta'\alpha'$ symmetry, having in this way two active sites per molecule.⁴⁸ While in caspase-7 every active site is independent,⁴⁹ experimental data suggest that the occupation of one active site in caspase-1 promotes the activity in the second one.⁵⁰

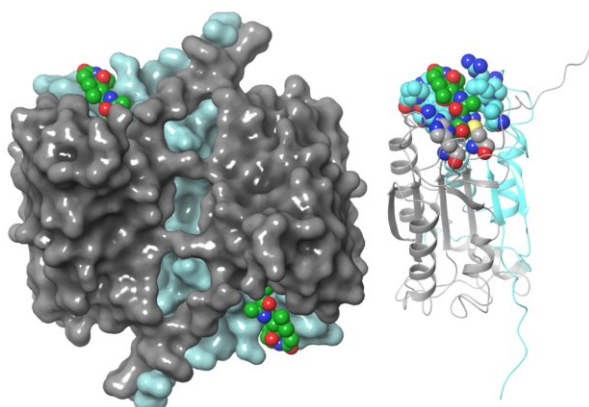


Figure 4. Left side, surface model of the $\alpha\beta\beta'\alpha'$ tetramer of caspase-1. Right side, ribbon representation of α and β subunits in cyan and gray, respectively. Residues forming S1 to S4 in CPK, gray carbons for residues interacting with S1 and cyan for residues in S2 to S4, the compound with green carbon atoms correspond to the co-crystallized inhibitor Ac-TYR-VAL-ALA-ASP-COH

Box 2

Michaelis-Menten Kinetics⁵¹

The efficiency of an enzymatic chemical reaction can be quantified accordingly to their velocity. This rate is proportional to the concentration of the substrate. In saturation conditions of substrate the reaction velocity becomes of order zero with respect to that species, and the total concentration of enzyme is assumed to be equal to the concentration of the complex enzyme-substrate.

Here we are assuming the enzyme and the substrate are in equilibrium during catalysis and product formation does not perturb this equilibrium. The reaction rate is proportional to the concentration of the enzyme-substrate (ES) complex (or Michaelis-Menten complex) So, the maximum velocity is achieved when all the enzyme is transformed into the ES complex.



$$V = k_3[ES]$$

Reaction mechanism for enzymatic reactions. Where E is the free enzyme, S the substrate, ES the complex enzyme-substrate, and P the product after the chemical reaction.

Assuming steady state conditions:

$$V_{form_{ES}} = V_{break_{ES}}$$

$$k_1[E][S] = (k_2 + k_3)[ES]$$

$$[E][S] = \frac{k_2 + k_3}{k_1}[ES]$$

$$[E][S] = K_m[ES]$$

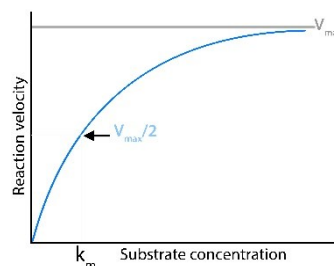
Knowing that $[E] = [E_{total}] - [ES]$ and having defined $V = k_3[ES]$

$$V = \frac{k_3[E_{total}][S]}{K_m + [S]}$$

At high concentration of substrate $[ES] = [E_{total}]$ so $V_{max} = k_3[E_{total}]$

$$V = \frac{V_{max}[S]}{K_m + [S]}$$

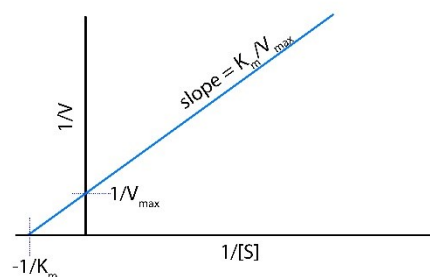
And the plot of the reaction rate versus substrate concentrations is:



Rate of an enzyme-catalyzed reaction as function for the substrate concentration

A more accurate way to determine the values of V_{max} and K_m for an enzymatic reaction is obtained plotting the data using the Lineweaver-Burk plot. This is obtained taking the inverse to the previously obtained expression to get:

$$\frac{1}{V} = \frac{K_m}{V_{max}[S]} + \frac{1}{V_{max}}$$



Scheme 1 shows a straightforward mechanism adapted from other cysteine protease to caspases. In this mechanistic proposal the cysteine is activated in the Michaelis complex after a proton abstraction carried out by the accompanying histidine in the catalytic dyad to form the thiolate-imidazolium ion pair.⁴⁵ This is likely for some cysteine proteases belonging to another clans, as cathepsin B, actinidain and papain, for example, where the donor-acceptor atoms in the dyad are close enough to react (3.7 Å in cathepsin B, 3.2 Å in actinidin, 3.9 Å in Papain), see **Figure 5a**. Nevertheless, the situation in caspases is quite different, see **Figure 5b**. As pointed out previously,^{52,53} the proton donor-acceptor atoms of the catalytic dyad are farther away than in the previously mentioned enzymes and the substrate binds between the catalytic dyad, that makes the direct hydrogen transfer unlikely (see figures 5b and 5c). For

example, in multiple crystallographic structures of caspase-1 these atoms are found at an average distance of 5.2 Å, see **Figure 5c**, making unlikely that activation occurs as proposed in scheme 1.

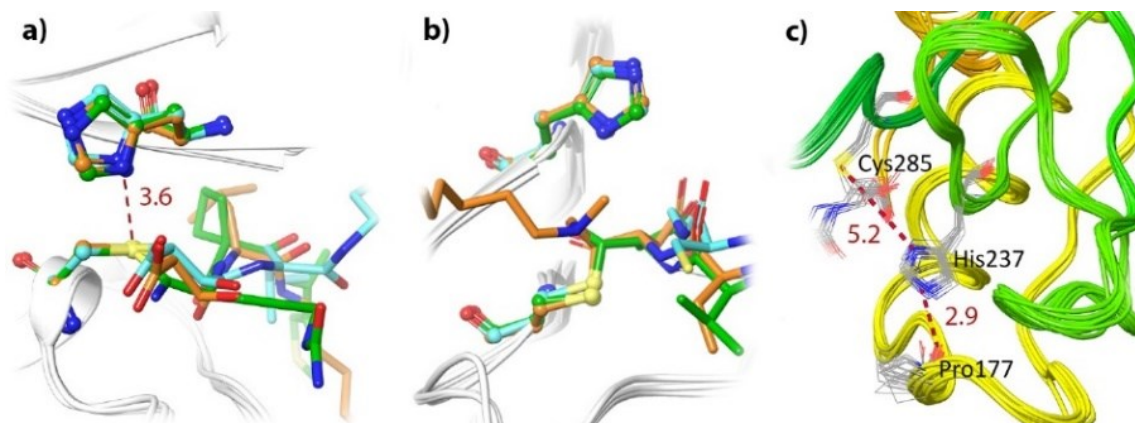


Figure 5. Catalytic dyad in cysteine proteases a) papain in cyan, actinidin in orange and cathepsin B in green (PDB codes: 1PPP, 1AEC and 6AY2 respectively) b) caspase -1 in orange, caspase-3 in cyan (1RHJ) and caspase-7 in green (1F1J) c) Catalytic center of wild type crystal structures of human caspase-1 co-crystallized with inhibitors containing an Asp in the core fragment (PDB codes: 6BZ9, 6F6R, 5MTK, 5MMV, 3NS7, 2HBQ, 1RWK, 1RWM, 1RWN, 1RWO, 1RWP, 1RWV, 1RWW, 1RWX, 1IBC and 1ICE) In red average distances from nitrogen atoms of the imidazole ring of His237 to the sulfur atom of Cys285 and to the oxygen atom of Pro177 in crystals (Å).

It is worth to mention that His237 has, in principle, three possible protonation states, two neutral (with nitrogens in epsilon or delta protonated) and one double protonated with positive charge. In the mechanism proposed by Wilson et al.⁴⁷ the protomeric state assumed for the initial step is the delta protonated histidine. Nevertheless, as show in the **Figure 5c**, the epsilon nitrogen is at H-bond distance from the carbonylic group to the residue in the adjacent loop (Pro177 in caspase-1), suggesting that the nitrogen epsilon has to be protonated. This aspect was addressed in detail in the present work⁵³ before starting the kinetic studies.

For human caspases it is unlikely that His237 deprotonates Cys285 (caspase-1 numbering) as proposed for the acylation step in the putative mechanism (Scheme 1). First, as said before, the average distance between the donor-acceptor heavy atoms in this catalytic dyad is 5.2 Å (see **Figure 5c**). And second, in the only crystal structure with an inhibitor co-crystallized with their P1' fragment (not deposited in the protein data bank), the inhibitor is placed between the two residues that form the catalytic dyad (see **Figure 5b** in orange). This structure was published as supplementary material by Galatsis et al.⁵⁴ This confirms partially that the proton from the catalytic cysteine will not be transferred to the catalytic histidine in the Michaelis complex as proposed in the putative reaction mechanism for the cysteine proteases. It is likely that the reaction mechanism found for caspase-1 can be extended to other caspases enzymes because what singularizes caspases from each other protease family are the amino acids present at the extended substrate recognition region, while the main characteristics in the catalytic center are shared among them. Additionally, the reaction mechanism proposal in the first structural study made for the caspase-1 by Wilson et al.⁴⁷ shows a cysteine being deprotonated by the epsilon nitrogen (N ϵ) in the imidazolium ring of the histidine. However, this nitrogen in the crystallographic structure is even further away from the cysteine than the N δ . Even if just a typographic error in this publication is an aspect that has to be clarified, it has been already propagated in some publications^{48,55,56} about caspase-1. A deep analysis of the protomeric states of the His237 has to be made.

A unified study evaluating the entire acylation and de-acylation process have not been performed so far on caspases. Previously, a study on caspase-7 for the acylation step was performed by Miscione et al.⁵² and another study on the de-acylation step on the caspase-3 was done by Sulpizi et al.⁵⁷ Both of them show new mechanistic possibilities for catalysis compared with the putative mechanism proposed for the proteases (scheme 1). Nevertheless, as the substrates and the enzymes used in each study are different is not possible to make a direct analysis of the overall process.

1.2.1.2. 3CL^{pro} of SARS-CoV-2

The 3CL^{pro} enzymes are highly structurally related to the chymotrypsin and picornavirus 3C cysteine proteases from where they took their name. The 3CL^{pro} encoded in the RNA of the SARS-CoV-2 share almost the same three-dimensional structure of its ortholog in the SARS-CoV,⁵⁸ what is an expected characteristic as they share 96% of their amino acid sequence. Additionally, the residues involved in catalysis, binding and dimerization are entirely conserved.⁵⁹ 3CL^{pro} is a non-structural protein (nsp) of the replicase polyproteins 1a and 1ab (pp1a and pp1ab, respectively). This enzyme is also known as the main protease (M^{pro}) as it is involved in the hydrolysis of 11 of the 14 cleavage events performed over the replicase polyproteins translated from the viral RNA.³⁰ Its cleavage site has been conserved among coronaviruses and is highly selective for Leu-Gln at the P2-P1 positions and Ser, Gly or Ala in the P1' position. From P2 towards the N terminal position in the substrate the specificity starts to be lost. The same is the case for the positions P2', P3' and so on towards the carboxyl terminal region. Studies made over the 3CL^{pro} of the Avian Infectious Bronchitis Virus (IBV) show that this enzyme is so specific for Gln at P1 that in-vitro activity is not observed when this amino acid is changed.⁶⁰

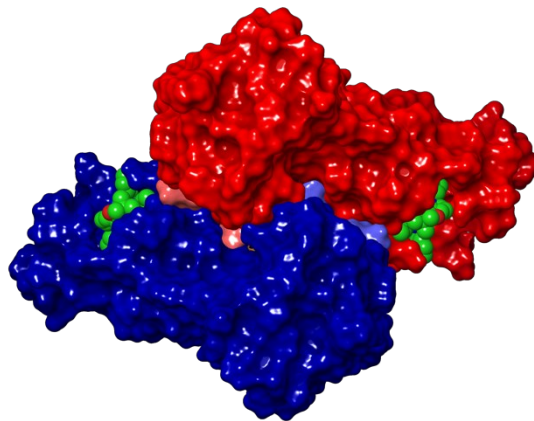


Figure 6. Molecular surface of Homodimers of 3CL^{pro} of SARS-COV-2 with their respective N-fingers colored in pale colors. In CPK the co-crystallized inhibitor O6K (PDB code: 6Y2F)⁵⁸

This enzyme is a homodimer where both monomers are intimately in contact by their residues 100-182 with the NH₂ terminal residues (1 – 10 pale colors in **Figure 6**, also known as N-finger) of the other monomer. This dimerization is needed as the N-finger of one monomer makes part of the S1 subsite of the other one. In **Figure 6** the dimer A is colored in dark blue, and the dimer B is colored in red. The figure shows how the N-finger of every monomer is close to the inhibitor present in the active site of the other monomer. In every active site there is a Cys-His catalytic dyad that hydrolyses the peptide sequence Gln-Ser-Ala-Gly at the peptide bond between Gln and Ser. As there are no human enzymes that recognize this specific sequence, this enzyme can be inhibited irreversibly by covalent inhibitors

which are expected to not interact with any enzyme of the host if the specificity recognition is used in the development of the inhibitor.³²

The proposed reaction mechanism for this enzyme is divided in two steps, acylation, and de-acylation. As shown in scheme 1, in the first step the C terminal fragment of the substrate is released after the cleavage of the peptide bond between the P1 and P1' fragments. The N terminal fragment of the substrate form a covalent complex with the enzyme after the nucleophilic attack of the S_γ atom of Cys145 on the Gln carbonyl carbon in P1. For that, this sulfur atom must be previously deprotonated and this is made by His41 in the catalytic dyad.⁶¹ In this case, contrary to caspases, the substrate is not placed in between the residues of the catalytic dyad and then a direct proton abstraction by His41 seems feasible. Additionally, it is well known this is the only chemical role played by this His in the acylation step because experimental studies show a lack of kinetic deuterium isotope effect when the acylation is carried out in presence of deuterium oxide.⁶² As the identification of intermediate structures along the proteolysis process is not an easy task to be done by experimental methods, some studies has to be performed computationally in order to try to rationalize the reaction mechanism in this type of enzymes. Nevertheless, when the pandemic started, there were no computational studies made in the family of the picornavirus 3C cysteine proteases. The closer study performed was made to describe the acylation process of a Michael acceptor inhibitor.⁶³ In order to develop effective and selective antiviral drugs to inhibit the replication process of SARS-CoV-2 a complete description of the 3CL^{pro} mechanism has to be done.

1.3 Enzymatic Inhibition

As explained, in the molecular pathogenesis of both AD and COVID-19 a proteolytic enzyme is involved in both cases, caspase-1 in AD²¹ and 3CL^{pro} in COVID-19.⁶⁴ In principle, in order to stop a disease it would be “just enough” to block the activity of one of these enzymes. This can be made with a small molecule, called inhibitor, that has to be designed in such a way that it binds the enzyme in competition with its natural substrate. If this is accomplished and this molecule is not involved in other biological processes it would be then possible to treat or cure the disease. This mechanism is known as competitive inhibition (see **Figure 7**). This name is given as the inhibitor and substrate cannot bind simultaneously in the active site, so they have to compete by the active site. Though this is an oversimplified point of view of the drug design process, it is one of the first aspects to evaluate in the development of competitive inhibitors. Noncompetitive inhibition is also another used strategy, which is out of the scope of the present work.

Enzymatic inhibition can be reversible or irreversible. Most enzyme inhibitors used today as drugs belongs to the first group.⁶⁵ Reversible inhibitors, in general, are not subjected to a chemical transformation by interacting with the enzyme. Here the enzyme-inhibitor complex is formed mainly by non-covalent interactions, such as hydrogen bonds, salt bridges, van der Waals and/or hydrophobic contacts. Conversely, on the irreversible inhibition, the enzymes are usually inhibited forming a covalent bond between the enzyme and the inhibitor, reason why they are also known as covalent inhibitors. Despite of this apparent bioactivity advantage, the use of this kind of inhibitors as drugs has been avoided because they use to be unspecific and can react with undesired targets present in the body. Because of this reason their administration as long term therapy may lead to severe side-effects.^{66,67} Nevertheless new evidence points towards the opposite direction.^{68,69} For both caspase-1 and 3CL^{pro} the covalent inhibitors designed are mainly based on the structure of the substrate^{67,70} and can be reversible or irreversible depending on the nature of the covalent bond formed between the enzyme and the

inhibitor. The strength of the covalent and non-covalent interactions can be evaluated by computational methods avoiding waste of efforts in synthesis and enzymatic assays.

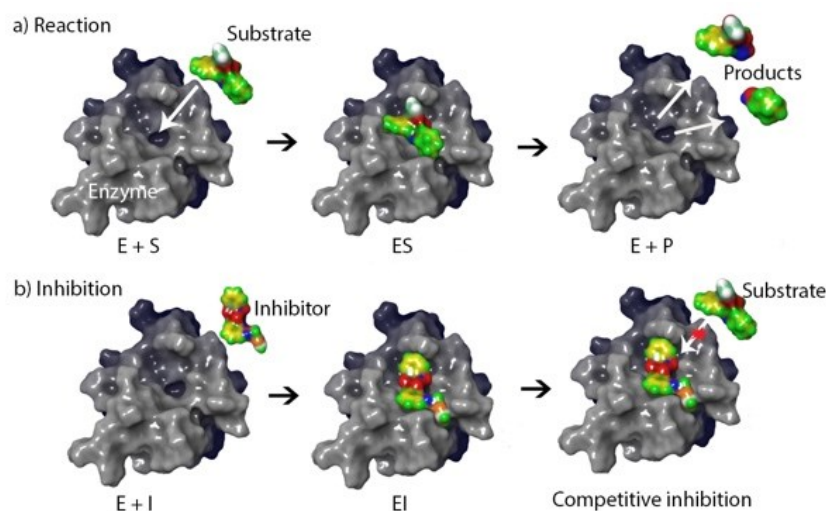


Figure 7. Competitive inhibition mechanism a) Natural catalyzed chemical reaction between the enzyme and its substrate in the active site b) Inhibition mechanism by a designed inhibitor preventing the binding of the substrate in the active site.

So, now we know two things: first, behind every disease there is a molecular deregulation; second, by fixing these deregulations the probability of curing the disease is high. But now a natural question raises: *can a cure be developed using only computational methods?* The short answer is ‘No’. Because so far, we don’t have a robust model that let us compute pharmacokinetics and pharmacodynamics parameters of drugs in humans. However computational tools are now included routinely in all stages of the drug design process,⁷¹ increasing the speed and decreasing the costs for delivering new drug candidates. Including computational methods in the drug design workflow increases the rate success and optimizes the process. Both families of enzymes, caspase-1 and 3CL^{pro}, have in common the absence of commercial drugs capable to treat the diseases in which they are involved.^{32,56} For that reason, it is important to understand the way these enzymes are interacting with their natural substrates by identifying the key enzyme-substrate interactions. And finally, it is necessary to elucidate the reaction mechanism of the enzymes with their natural substrates and putative covalent inhibitors in agreement with the experimental observations. Only having this atomistic description of the binding process and a clear understanding of the transition state (TS) structure involved in the chemical reaction it will be possible to design stable molecules able to bond tighter to the enzyme than their natural substrate.⁷²

2. Computational methods in drug design

A good understanding of the chemical properties of enzymes is a topic of great pharmacological interest, as they are involved in the molecular pathogenesis of many diseases. The understanding of enzymatic reactivity involves the description of the steps along the reaction, the identification of the transition states (TS) and the intermediates formed along the reaction, as well as the non-covalent enzyme-substrate interactions that stabilize the Michaelis-Menten complex.

Wet lab techniques can give us an idea about how this complex looks like and how the enzyme is reacting. With these techniques it is possible to measure how the reaction kinetics changes while the enzyme reacts with different kind of substances,⁴⁶ for example. Or how it changes when an atom that is presumed to be involved in the reaction mechanism is changed by its isotope.⁷³ Also, changes can be made to the enzyme, some amino acids can be mutated and their role in the catalysis evaluated by the impact of this mutation on the reaction rate.⁴⁷ And finally, crystallographic structures can be obtained to establish a connection between the results of the afore mentioned techniques and the 3-D distribution of the atoms in the protein. But with these techniques only it is not possible to get the as detailed atomistic description of the overall process as can be achieved by computational methods.

Computational methods have proven to be a valuable tool in the structure-based drug design process and in the discovery of new drugs through lead optimization.⁷⁴ Tools as molecular dynamics (MD) simulations, for example, let us to study the structural flexibility of the enzyme-substrate/inhibitor complex which can aid us to predict thermodynamic and kinetic properties comparable with experimental data. Another example is molecular docking: this methodology is used for virtual screening campaigns.⁷⁵ Here, hundreds of thousands, or even millions, of compounds are scored for potential activity in the enzyme active site, and the compounds top-ranked are then selected to be tested experimentally or for further computational optimizations. *De novo design* is a methodology that help us to design molecules from the scratch if the target binding site is known.⁷⁶ An example of this approach is the fragment-based drug design. Here, a small molecule is initially placed in the active site, then this fragment is transformed into other different fragments to improve the enzyme-substrate interactions, hydrogens bonds and/or solvation energy, for example. Once this is achieved, a new fragment is attached to the previous one and the process is repeated until obtaining a molecule with the desired properties.

Many things have changed since Pauling proposed that the molecular size and shape of enzymes were the most important aspects in the enzymatic catalysis process, and that the main role of enzymes was to be complementary in structure to the activated complexes.⁷⁷ Nowadays, with a better understanding of the structure and dynamics of enzymes, we known that enzymatic active sites are electrostatically complementary to the TS.⁷⁸ Also, we know that conformational changes may be required to facilitate interaction between an enzyme and its natural substrate and that the presence of the substrate in or close to the active site can also induce conformational changes.⁷⁹ For those reasons, it is so important to have a methodology that allows to obtain the conformational distribution of enzymes in certain states in their natural environment or as close as possible.

2.1 Classical Molecular Dynamics

'If we were to name the most powerful assumption of all, which leads one on and on in an attempt to understand life, it is that all things are made of atoms, and that everything that living things do can be understood in terms of the jiggings and wiggings of atoms.'

Richard Feynman.

As Richard Feynman pointed out, if we want to understand life, we need to understand how atoms move. The problem with this is that enzymes are molecules with tens of thousands of atomic nuclei with their corresponding electrons. So, a rigorous description of this process would involve evaluation of the

interaction between all those particles by quantum mechanics. This is an impractical approach for systems of this number of atoms. However, for many processes (like ligand binding or protein conformational changes) it can be assumed that the formation or breaking of covalent chemical bonds does not happen. In this case, the electronic structure of the atoms in the molecule can be neglected and the description of the system can be reduced to predict the classical motion of the nuclei according to an effective molecular mechanics potential.

Classical molecular dynamics simulations are the result of using the laws of classical mechanics to predict the structures and properties of molecules.⁸⁰ Here, the potential energy of the system is related to geometrical variables such as bond lengths, bond angles, torsional angles, etc. As the electronic structure is neglected, the atoms are represented as charged van der Waals spheres connected by springs. In this way classical Newtonian mechanics can be used to obtain the forces acting on the atoms in the system and predict their displacement in time. The spatial coordinates that describe the initial positions of the atoms in the molecule are obtained by experimental techniques as nuclear magnetic resonance (NMR), X-ray diffraction crystallography or by construction of unknown structures by homology modeling from known structures.⁸¹

The force (f) acting on a specific atom i of the system at a given point of time t is obtained by solving the equation of the Newton's second law:

$$f_i(t) = m_i a_i(t) = - \frac{\partial V(x(t))}{\partial x_i(t)} \quad 1$$

Being m_i and $a_i(t)$ the mass and the acceleration of the atom i . The spatial coordinates of the N interacting atoms are described by the vector $x(t)$. $V(x)$ is an empirical potential energy function.

2.1.1. Problem Solution from Newton Mechanics

Despite the reduction of computational cost by treating the system in a classical way, Newton's second law can only be solved analytically for very simple systems. Nevertheless, in a simulation of a protein solvated in water, the number of atoms can be close to tens of thousands. So, a numerical solution has to be employed to integrate the equation of motion. This integration has to be done in discrete intervals δt small enough to warrant precision, but large enough not to compromise computational cost. These time intervals are known as time steps. As the forces have to stay approximately constant during the integration step, a small time-step has to be selected also because the forces change with the position of atoms. The rule of thumb that has proved to produce reliable forces over time is to select a time step around 10 times smaller than the fastest frequency of vibration in the system. In enzymatic systems that frequency is 10 fs, corresponding to the stretching of a bond formed between a hydrogen atom with a heavy atom. So, the time-step to be used is 1 fs. This time-step can be increased to 2 fs if the bonds involving hydrogens are frozen by algorithms such as SHAKE.⁸²

In this way the position of atoms at time $t + \delta t$ are calculated from their current position, velocities, and accelerations by the expression:

$$x_i(t + \delta t) = x_i(t) + v_i(t)\delta t + \frac{1}{2} a_i(t)\delta t^2 \quad 2$$

And their velocities by:

$$v_i(t + \delta t) = v_i(t) + \frac{1}{2} [a_i(t) + a_i(t + \delta t)]\delta t \quad 3$$

For that, first we need to find the acceleration of the particles due to the forces acting over them. By taking the negative derivative of the potential energy, the force can be calculated by the equation 1

$a_i(t)$ will be calculated to solve the equations 2 and 3, the position of the particles will be updated in the trajectory and the cycle is repeated with the new set of positions and velocities. These equations describe the Velocity Verlet integrator.⁸³ This is a second order integrator that decrease the error made in each step if compared with simplest forward Euler integration method. Now a proper expression that describe the potential of the function is needed. This mathematical expression receives the name of force field.

2.1.2. Force Fields

The following expression describes one of the simplest forms for a realistic a force field.⁸⁴

$$V(r) = \sum_{bonds} k_b(r - r_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2 + \sum_{dihedrals} k_\phi[1 + \cos(n\phi - \phi_0)] + \sum_{nonbond} \left\{ 4\varepsilon_{ij} \left[\left(\frac{\sigma_{0\ ij}}{R_{ij}} \right)^{12} - \left(\frac{\sigma_{0\ ij}}{R_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\varepsilon_0\varepsilon_r R_{ij}} \right\} \quad 4$$

Where the terms named as bonds, angles and dihedrals describe the bonded interactions, while the remaining terms are describing the non-bonded interactions within the pair approximation. The first two terms describe the potential energy contribution, as a harmonic potential, due to the stretching and bending of bonds when moved from their reference values r_0 , and θ_0 . Being k_b and k_θ are their respective force constants. The third term describe the interaction between two atoms separated by three consecutive bonds forming a dihedral rotational angle around the central bond. The rotation along this bond is described as a cosine series of M terms for each dihedral angle with an angle force constant k_ϕ for this rotation and ϕ_0 is the phase angle. The next term is the 6-12 Lennard-Jones potential describing the van der Waals, or non-bonded, interactions between two atoms i and j , being R_{ij} the distance between the two atoms, σ_0 is the distance at which the interaction vanishes and $\varepsilon_{i,j}$ is the bottom of the potential well. $\varepsilon_{i,j}$ and $\sigma_{0\ ij}$ are specific for each pair of atoms i and j . And finally, the last term is a Coulomb potential that describes the electrostatic interaction between the atoms i and j with partial charges q_i and q_j at a distance R_{ij} , being ε_0 and ε_r the permittivity of vacuum and the dielectric constant respectively. The electrostatic and van der Waals interactions are not evaluated between all the possible atom pairs in the system but only for those formed inside a cutoff radius established during the calculation. The terms k_b , r_0 , k_θ , θ_0 , k_ϕ , ϕ_0 , ε_{ij} , σ_0 , q_i and q_j are parameters to be specified in the force field.

The force field used in this work was the ff14SB.⁸⁵ It contains parameters for the atoms forming a protein made of standard amino acids. This force field gives an atom type definition for every atom present in each amino acid. The water molecules were described using the flexible TIP3P model.⁸⁶

2.2 QM/MM methodologies

As mentioned previously, classical methods are only valid when one is not interested in changes in the electronic structure, including chemical reactions. On the other hand, if we want to get a description of electronic changes that take place in a limited region of the system, it is not needed to describe the entire system by quantum methods. Usually, the changes in the electronic structure during a chemical reaction involve only a reduced number of atoms. In a hybrid Quantum Mechanics/Molecular Mechanics

(QM/MM) treatment, most of the system can still be described in a classical way by a force field, while the atoms involved in the chemical reaction will be treated by quantum mechanics, i.e. a hybrid description of the system will be used.⁸⁷

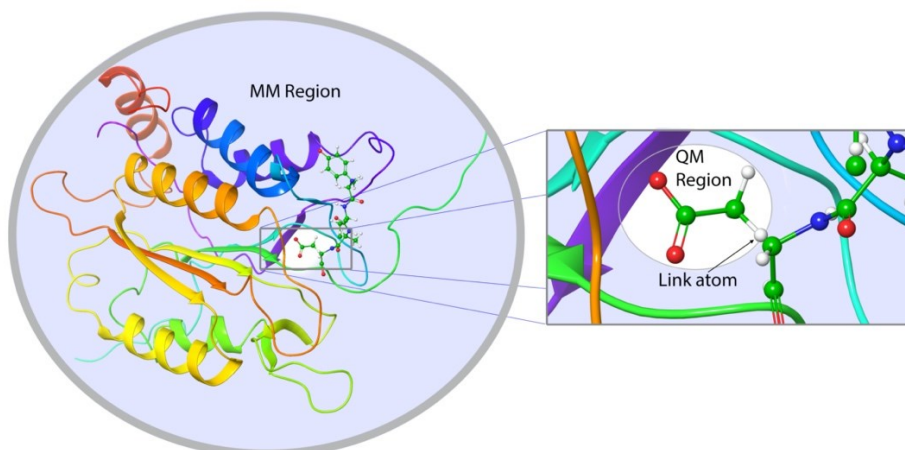


Figure 8. QM/MM regions and link atom partitioning scheme.

A strategy for partitioning the system into QM and MM regions has to be envisaged. In some cases, the chemical reaction involves only a small molecule that performs an intramolecular reaction. In these cases, the QM region includes only this molecule, while the remaining part of the system is described by classical mechanics. Nevertheless, in most enzymatic reactions, atoms from water molecules, cofactors or from the enzyme, must be included in the QM region as they are involved in the process. In these cases, the QM and MM regions will be connected by covalent bonds, and it is not possible to simply delete from the QM region the MM atoms that are not going to be included in the QM calculation, because it will cause an incorrect description of the electronic states. Typically the bond to be cut is a single carbon-carbon bond. One option to solve the problem is to include a single additional electron and an sp^3 orbital in the QM region,⁸⁸ but it is difficult to derive orbitals that can be transferred between systems. Other approach to solve the uncompleted valence is to include a link atom between the QM and MM boundaries.⁸⁹ This method has the advantage of being simple and effective. This link atom is usually a hydrogen atom included in the QM region. Its length is constrained to the equilibrium distance of a methyl C-H atom pair, 1.09 Å, and it is positioned along the vector of the bond connecting the QM and MM parts. (Figure 8) The force acting over it is distributed over the two atoms forming the bond.

In QM/MM methods the effective Hamiltonian of the system H_{eff} is made of three terms. One for the pure MM interactions, another for the QM interactions and finally one describing the interactions between the MM and QM regions. The Hamiltonian operates over the wave function $\Psi(x_e, x_{QM}, x_{MM})$ to obtain in this way the energy of the system E_{eff} :

$$H_{eff}\Psi(x_e, x_{QM}, x_{MM}) = E_{eff}(x_{QM}, x_{MM})\Psi(x_e, x_{QM}, x_{MM}) \quad 5$$

The wave function depends on the coordinates of the electrons that are explicitly described x_e , the position of the nuclei in the QM system x_{QM} , and the atoms in the MM region x_{MM} . As the energy of the MM region is independent of the position of the coordinates of the QM region the associated energy term can directly calculated using an expression similar to that of equation 4.⁸⁷

The interactions between QM and MM atoms will be evaluated considering those MM atoms surrounding the QM region and included within a cutoff distance. The polarization of the QM region by the MM region will affect the evaluation of the chemical properties. For that reason, in this work the

electronic embedding or additive scheme was used and the interaction energy between the QM and MM regions is described by:

$$E_{QM/MM} = \sum_{A \in MM} \int \rho_e \frac{Q_A}{|r - R_A|} dr + \sum_{A \in QM, B \in MM} \frac{Z_B Q_A}{R_{AB}} + \sum_{A \in QM, B \in MM} 4\epsilon_{AB} \left[\left(\frac{\sigma_{AB}}{R_{AB}} \right)^{12} - \left(\frac{\sigma_{AB}}{R_{AB}} \right)^6 \right] \quad 6$$

being the first term the electrostatic interactions between the QM electron density $\rho_e(r)$ and the MM point charges Q_A , the next term describes the interaction between charge of the nuclei in the QM region Z_B , and the MM point charges Q_A and the last one is for the van der Waals interaction between the QM and MM atoms.

2.3 Free energy calculations

So far, the attention has been centered in the prediction of the future position of atoms in a system due to the forces acting over them as well as in finding the potential energy of that specific configuration or microstate.

From here, we will need to start to understand the picture from the wider perspective coming from statistical mechanics. In the real world we do not refer to the potential energy of a specific microstate. What is experimentally measured is the average energy of a macroscopic state, which is a macroscopic description of an ensemble of microstates, each with its own specific potential energy and its own probability (or probability density, in the classical limit) to exist. The latter is a key concept to understand the whole story, because not all microstates can be populated with the same probability: the higher the energy the lower the probability. Let us consider the microstates obtained as a solution of the Hamiltonian of our system defined with a given number of particles and a fixed volume. The connection between the probability of a **microstate a** to exist and its energy at a temperature T is given by the Boltzmann factor $e^{-\frac{V(r)}{k_B T}}$. In the classical limit, the probability density $\rho(r)$ corresponding to **microstate a** is defined by:

$$\rho(\mathbf{r}_a) = \frac{e^{-\frac{V(\mathbf{r}_a)}{k_B T}}}{\int_{\Omega} e^{-\frac{V(\mathbf{r})}{k_B T}} d\mathbf{r}} \quad 7$$

The expression in the denominator is the partition function, Z from now. This function becomes a number once the potential energy function, the temperature and the volume are specified. Now, to evaluate *the probability of finding a state A we will need to integrate the probability density over all the region, A_A , encircling the whole basin in the state A:*

$$P_A = \int_A \rho(\mathbf{r}) d\mathbf{r} = Z^{-1} \int_A e^{-\frac{V(\mathbf{r})}{k_B T}} d\mathbf{r} \quad 8$$

As the probability of finding the system anywhere is 1 if the integration is made over the entire configurational space, the free energy of the state A will be:

$$G_A = -k_B T \ln P_A \quad 9$$

By replacing P_A from equation 8 into equation 9 we obtain the statistical mechanics relation:⁹⁰

$$G_A = -k_B T \ln \left(Z^{-1} \int_A e^{\frac{-V(r)}{k_B T}} dr \right) \quad 10$$

Then the free energy of any state can be obtained as in equation 10. So, as the integral term is always positive and its logarithm increases slowly, the resulting value of the free energy will decrease very slowly as the integration is made over a larger part of the phase space. i.e., when more possible values for the velocities and positions are considered. As it is not possible to evaluate this integral over the entire phase space except for very simple systems, it is also not possible to calculate the absolute free energy of molecular systems. A practical alternative is to calculate the difference of free energy between ensembles of molecular simulations of closely related states over a discretized phase space.⁹¹

2.3.1. Free Energy Differences

As mentioned above, the probability of finding a molecular system in one state (A) or another (B) is determined by the difference in free energy between those two states. Or in the other way around, knowing the relative probability of two states their free energy differences can be deduced. The most straight forward of getting these relative probabilities is the direct counting of microstates in the corresponding state. In this way the free energy difference ΔG_{AB} can be obtained as:

$$\Delta G_{BA} = G_B - G_A = -k_B T \ln \left[\frac{N_A}{N_B} \right] \quad 11$$

Here the relative probabilities were replaced by the number of times a state is found.⁹¹ Nevertheless, equation 11 can only be directly used when both states, A and B, appears in our experiment (computational or not) with a frequency high enough to provide reliable statistics. So, this methodology can only be used in process where ΔG_{BA} and its variation along the process is small enough to permit the frequent interconversion of one state into the other. When this is not possible, as in most processes we are interested in, other kind of techniques must be employed to compute ΔG_{BA} .

Before we proceed, a clarification must be made. The previous expressions are derived considering a canonical ensemble and then refer to the Helmholtz free energy F . While if the pressure is kept constant, rather than the volume, the Gibbs free energy G is a more adequate descriptor. This is more a technical clarification than a practical one. As we are working in condensed phases and not in diluted systems, the difference between one or another is almost irrelevant for the systems studied here. This clarification rises from the fact that experiments are usually made at constant pressure rather than at constant volume, but the numerical differences between them would be indistinguishable.⁹⁰

2.3.1.1. Thermodynamic integration

Thermodynamic integration (TI) is a technique that helps to find the free energy difference between two physical states A and B by introducing (un)physical states between them. This methodology was used in this project to obtain the binding free energy differences related to mutations in peptide-like inhibitors. The method is based on an alchemical thermodynamic cycle⁹² as in **Figure 9**.

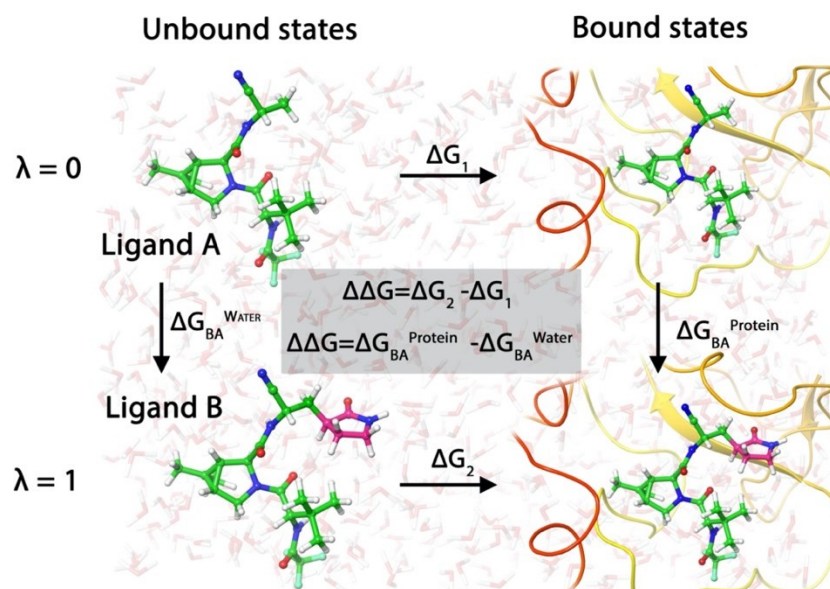


Figure 9. Thermodynamic cycle in alchemical transformation. A hydrogen in the ligand A is replaced by the γ lactam ring in purple.

To calculate directly ΔG_1 and ΔG_2 in **Figure 9** a large part of the phase space must be considered because there are many possible ways a ligand can diffuse from the bulk of the solvent to the active site in the enzyme. On the other hand, to obtain ΔG_{BA}^{Water} and $\Delta G_{BA}^{Protein}$ the energy changes will be related only to changes in the surroundings of the transformation region in the bound and unbound systems. As this is a closed thermodynamic cycle the result of evaluating $\Delta G_{BA}^{Water} - \Delta G_{BA}^{Protein}$ will be equivalent to evaluate $\Delta G_2 - \Delta G_1$ as shown in the **Figure 9**.

To connect the initial state A with the final state B, a coupling parameter λ can be introduced in the energy expression $V(r; \lambda)$. As the energy is a function of λ also F will become a function of it.

$$\frac{dG(\lambda)}{d\lambda} = \left\langle \frac{\partial V(\lambda)}{\partial \lambda} \right\rangle_{\lambda} \quad 12$$

Where $V(\lambda) = (1 - \lambda)V_A + \lambda V_B$, this λ value will change in such a way that when $\lambda = 0$ the system the potential energy function corresponds to that of the state A, $V_A(r)$, and when $\lambda = 1$ it will correspond to B, $V_B(r)$. For intermediate values the potential energy function will be a combination of both states to a greater or lesser extent. At these values of λ , a series of simulations are performed. And the difference of free energy between the states A and B is obtained by the numerical integration of:

$$G(B) - G(A) = G(\lambda_B) - G(\lambda_A) = \int_0^1 \left\langle \frac{\partial V(\lambda)}{\partial \lambda} \right\rangle_{\lambda} d\lambda \quad 13$$

In this work integration was carried out using the gaussian quadrature integration scheme with nine steps. The corresponding λ values are: 0.01592, 0.08198, 0.19331, 0.33787, 0.5, 0.66213, 0.80669, 0.91802 and 0.98408.

2.3.1.2. Molecular Mechanics Generalized Born Surface Area (MM/GBSA)

The MM/GBSA method is another computational tool used to compare the binding free energy differences between small ligands.⁹³ This methodology has the advantage of being an end point sampling method. So, only the energies of the bound and unbound states are involved in the calculation.

In this way it is possible to evaluate the difference in binding free energy for a vast number of compounds, if compared with the more computationally demanding thermodynamic integration method. Naturally this gain is obtained from severe approximations that can limit the applicability of the method.⁹⁴

The binding free energy of every small molecule is computed by obtaining the free energy value of the ensemble average of every point:

$$\Delta G_{bind} = \langle G_{complex} \rangle - \langle G_{protein} \rangle - \langle G_{ligand} \rangle \quad 14$$

Despite the binding energy is estimated from the free energy differences of the three components involved, it is worth to mention that typically a single trajectory approach is employed. In this, a trajectory is obtained only for the complex in explicit solvent and from this the ensemble, the energies of the protein, ligand and complex are obtained by deleting everything but the structures of interest—protein, ligand and complex respectively. The resulting structures are solvated within a continuum model.⁹³ For every component its free energy is obtained from the expression:

$$G = E_{MM} + G_{solv} - TS \quad 15$$

Where E_{MM} , as previously explained, is divided into three terms:

$$E_{MM} = E_{cov} + E_{ele} + E_{vdw} \quad 16$$

By using the single trajectory approach the energetic terms of the covalent contributions cancel out, as the same geometries are considered for the complex and the protein alone. So, the binding energy is now described by:

$$\Delta G_{bind} = \Delta E_{ele} + E_{vdw} + \Delta G_{solv} - T \cdot \Delta S \quad 17$$

Where the first two terms are obtained from the evaluation of the force field equation 4. In the single trajectory approach, the solvation term describes the polar and non-polar contributions to the solvation free energy computed via the Generalized Born approximation and using linear function of the Solvent-Accessible Surface Area, and the entropic contribution can be obtained after minimization and normal mode analysis. Nevertheless it is usually neglected in studies made over a series of congeneric ligands, as the conformational freedom for every small molecule is almost the same having a small impact on the ΔG_{bind} value.⁹⁵

2.3.2. Potential of Mean Force (PMF)

A PMF is the free energy whose Boltzmann factor gives the probability distribution projected along a subset of coordinates. Then, for chemical reactions, we can say a PMF is the energy landscape along the reaction coordinate (RC), which can be defined as a simple geometrical coordinate or a combination of bond distances, angles, dihedrals, etc.

The PMF of a chemical reaction is expected to have at least two minima basins corresponding to the structures of reactants and products. These minima are separated by a maximum corresponding to the TS of the chemical reaction, whose PMF difference with the reactants is related to the activation energy ΔG^\ddagger .⁹⁶ The gradient of the PMF corresponds to the mean force acting on the coordinate ξ' at a given value ξ .⁹⁰ The free energy as a function of ξ is the potential of mean force $G(\xi)$:

$$G(\xi) = -k_B T \ln Z^{-1} \int_{\Omega} e^{\frac{-V(\mathbf{r})}{k_B T}} \delta(\xi'(\mathbf{r}) - \xi) d\mathbf{r}$$

The selection of the RC can be cumbersome for chemical processes involving several elementary steps (bond breaking and forming events). In such a case one can explore multidimensional PMFs along several distances or combination of distances. In this work the RC has been defined from the exploration of the minimum free energy path (MFEP) on a multidimensional free energy surface (FES) using the Adaptive String Method.⁹⁷ This path describes most of the reactive trajectories — ones that connect the reactants and products basins. Once the RC has been defined the associated PMF can be obtained using some free energy calculation technique, such as Umbrella Sampling (US)⁹⁸ as described below.

2.3.2.1. Adaptive String Method (ASM)

The ASM⁹⁷ is a path-based method which aims to obtain the MFEP on a multidimensional free energy surface formed by an arbitrary number of collective variables (CVs).⁹⁷ Once the MFEP is obtained, it is used to define a path-CV – a RC which describes the advance of the system along the reference path.⁹⁹

To obtain the MFEP a reasonable initial guess is provided in the space of the selected CVs, see **Figure 10**. These CVs are assumed to cover all the degrees of freedom relevant for the chemical reaction under study. N evenly spaced points (string nodes) are placed along this putative path. For each point, a separate MD trajectory is launched restrained to the position of this point in the CV space. The nodes are then moved towards regions with lower free energy according to the average forces obtained from the MD trajectory. In these displacements the nodes are kept equidistant to avoid them to fall to the reactants and products minima. The process is iteratively repeated until the string converges to the MFEP. In order to properly sample the degrees of freedom orthogonal to the RC and try to avoid the trajectories to be trapped in local minimum, replica exchange¹⁰⁰ is performed between string nodes.

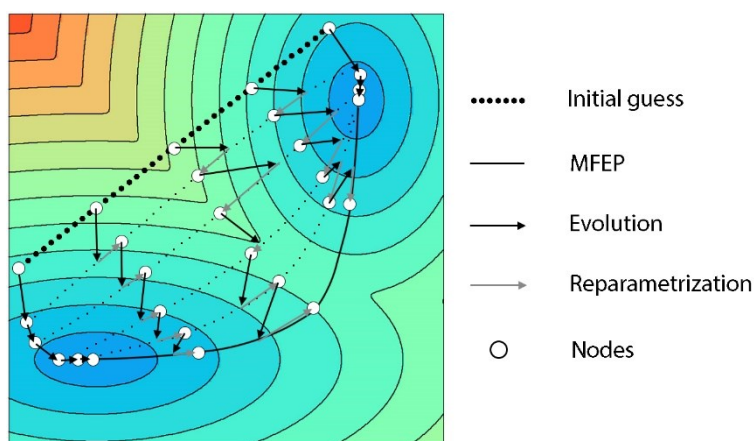


Figure 10. String evolution towards the MFEP.

With the string converged, a single path-CV (a collective variable called s) that measures the advance of the system along the MFEP is defined. This path-CV is used as the RC to trace the free energy profile associated with the chemical transformations under analysis.

2.3.2.2. Umbrella sampling

As the system in an unrestrained simulation would be trapped in the region of reactants or products, relevant regions in the phase space would remain unexplored. To avoid that, a set of biased simulations (windows) distributed along the RC must be performed. This will force the system to uniformly sample configurations along the RC. This bias usually takes the form of a parabola, centered at different values of the selected coordinate:⁹⁸

$$V_{b,i}(\xi'(\mathbf{r})) = \frac{1}{2}K_i(\xi'(\mathbf{r}) - \xi_i)^2 \quad 19$$

Where V_i is the potential bias in the window i , K_i the harmonic force constant, ξ_i the reference value of the coordinate and r are the actual values of the coordinates of the system. By including this bias potential, the biased PMF now has the form:

$$G_b(\xi) = -k_B T \ln Z_b^{-1} \int_{\Omega} e^{-\frac{(V(r)+V_b(\xi'(r)))}{k_B T}} \delta(\xi'(r) - \xi) d\mathbf{r} \quad 20$$

Where Z_b is partition function of the biased ensemble. As V_b depends only on ξ , and that is precisely where the Dirac's delta is applied, V_b can be get out of the integral and a new expression for the PMF can be obtained:

$$G_b(\xi) = G(\xi) + V_b(\xi) + k_B T \ln \frac{Z_b}{Z} \quad 21$$

As different V_b are used in each window, the last term in equation **21** will be also different, so a proper window reweighting factor $\frac{Z_b}{Z}$ has to be used, this can be obtained from an iterative process as the Weighted Histogram Analysis Method¹⁰¹ (WHAM). This method solves a series of equations for the reweighting factors and for the PMF in an iterative way. The number of windows and the value of the force constants should be selected to warrant the overlap of the histograms of adjacent windows and to accumulate enough sampling of the selected coordinate within each window.

3. Results and conclusions

3.1 Caspase-1

As previously mentioned, it was necessary to establish the proper protonation state of the catalytic dyad in the active site of the enzyme caspase-1 before performing kinetic studies. As previously explained, the direct formation of the ion pair structure by proton transfer between the catalytic dyad is not possible because of the large distance between them and the fact that the substrate is positioned in between. This can be seen more clearly in **Figure 11** where the active sites of the 3CL protease and caspase-1 are compared. As it can be seen, the substrate is placed between the Cys285 and the His327, for that reason the sulfur cannot be activated by a direct proton transfer to the catalytic histidine. This proton dyad activation is more likely to occur in the case of the 3CL^{pro} of SARS-CoV-2 as will be discussed later. So, the reaction mechanism in caspase-1 cannot proceed through the putative reaction mechanism proposal.⁴⁵

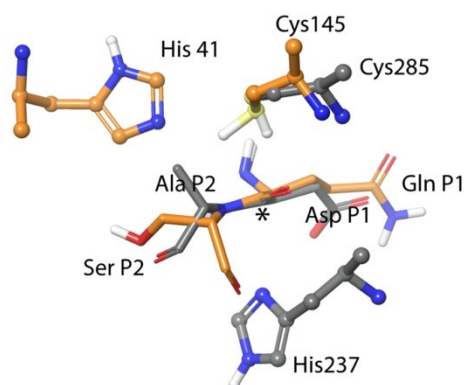


Figure 11. Comparison of the catalytic dyad in the enzyme caspase-1 in gray and the 3CL^{Pro} of SARS-CoV-2 in orange. The substrate atoms are in licorice and the enzyme atoms in CPK, the * is on the electrophilic carbon in both substrates. Both structures were extracted from molecular dynamic simulations. The PDB codes used as starting points were the 6F6R for the caspase-1 and the 6Y2F for the 3CL^{Pro} of SARS-CoV-2.

As explained in appendix 1, molecular dynamics simulations of the enzyme substrate complex with the catalytic dyad in the ionic form (Cys285⁻/His237⁺) were performed. By analyzing these simulations it was observed that the complex was not stable in this conformation and the substrate leaves the active site after a couple of nanoseconds of simulation.⁵³ This was also evaluated by Carloni et al.⁵⁷ for the enzyme caspase-3, for that case the inhibitors also leave the active site if the catalytic dyad was in its ion pair configuration.

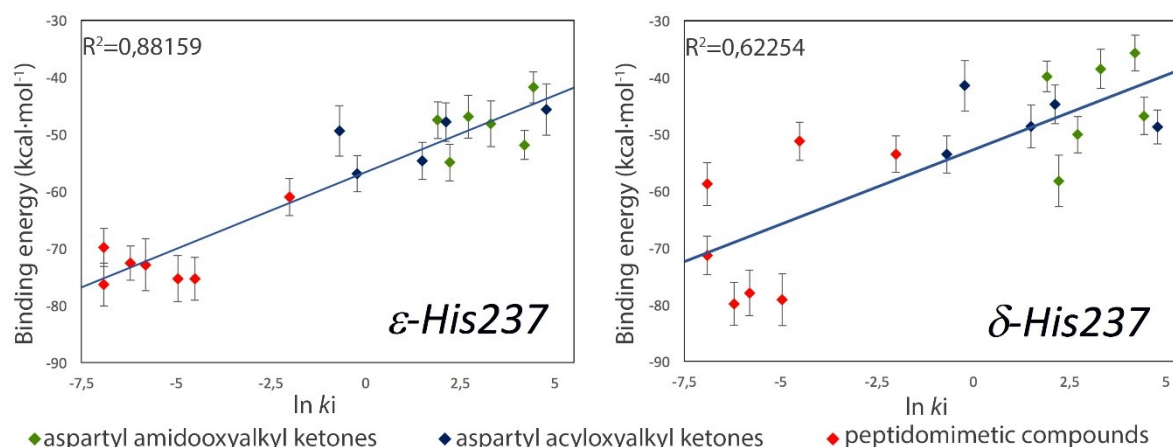


Figure 12. Correlation between the experimental K_i and the MM/GBSA binding energy calculated for the two systems in different tautomeric states for the His237. Left panel: ϵ -His237; right panel: δ -His237.

The binding free energy of a series of congeneric ligands was evaluated using MM/GBSA. Using this technique it was possible to establish the most probable protonation state for the His237. It was observed that the calculated values correlate better with experimental results when the catalytic dyad is assumed to be in the neutral state and the proton in the neutral histidine is placed in the N ϵ than in the N δ , **Figure 12**. In our study it was also possible to establish the key residues forming the key interactions in the enzyme-inhibitor complexes, Figure 8 appendix 1. It was observed that the main difference between potent inhibitors and weak inhibitors is related to the interactions they establish with residues His342, Pro343 and Arg383 in the S2 to S4 regions.⁵³ Therefore, a possible strategy for the design of new and more potent inhibitors of caspase-1 should focus in strengthening the interactions established with this region.

By having established the protonation of the catalytic dyad in the Michaelis complex we can study the chemical reaction between the caspase-1 and its substrate. So far most of the mechanistic studies made on cysteine proteases have been carried out for papain and related enzymes. To our knowledge, our study is the first one made on the caspase family that explains the role of the imidazole ring in the acylation step. A previous attempt to explain the acylation step for this family of enzymes was made by Miscione et al.⁵² using an aspartate aldehyde derivative as substrate and the caspase-7 as enzyme. In this mechanism the imidazole ring act as a general base deprotonating a series of water molecules placed between this moiety and the aldehyde group of the substrate. However, our molecular dynamic simulations revealed that there are no water molecules close to the $N\delta$ of the catalytic histidine when the leaving group or P1' group is still present in the active site, see Figure 13. Then, the proposed mechanism could be plausible only for aldehydes or inhibitors without a P1' group and not for the real substrate.

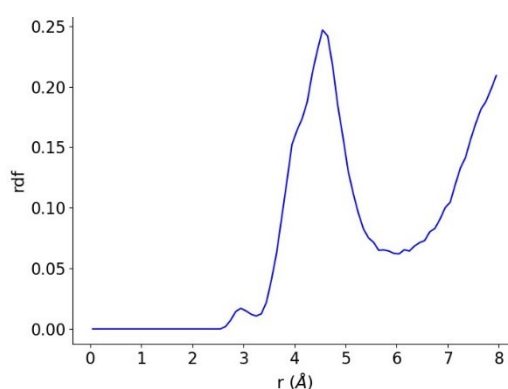


Figure 13 Radial distribution function of water molecules around the $N\delta$ atom in the imidazolium ring in the His237

The energy landscape for the formation of the acyl-enzyme complex from the neutral catalytic dyad previously found was explored at the B3LYPD3/MM level, using a computational scheme very similar to that employed for the SARS-CoV-2 3CL protease (see appendix 2 and references therein). The only difference is that in the case of caspase-1 we employed a 6-31G* basis set, without diffuse functions, because a ionic pair is not formed, as discussed above. The MM subsystem was described using the ff14SB forcefield⁸⁵ and the water molecules were described by the TIP3P model.⁸⁶ For this study the substrate used was the substrate peptide mimetic sequence ACE-Tyr-Val-Ala-Asp-Ala-Pro-Val-Arg-NME. We explored the free energy landscape corresponding to the first step of the process (see Scheme 1), corresponding to the formation of the acyl-enzyme. This is in fact the controversial part of the reaction mechanism, as it is widely accepted that the de-acylation step involves a water molecule to break the enzyme-substrate bond.

Using the ASM method (see section 2.3.2.1) we explored the multidimensional free energy landscape determined by the five CVs shown in the right panel of **Figure 14**. The left panel of **Figure 14** shows an intermediate structure formed associated to it. The mechanism comprises two steps. The first barrier corresponds to the proton transfer from the catalytic cysteine to the nitrogen atom of the target peptide bond (see CV1 and CV2) followed by the nucleophilic attack of the sulfur atom on the electrophilic carbon of the carbonyl group (as shown by the evolution of CV4). This intermediate is stabilized by a strong hydrogen bond between the protonated peptide nitrogen and the catalytic histidine (see the evolution of CV5). In the second step of the acylation mechanism the peptide bond is broken (CV3 increases) forming the E-S complex and releasing the first fragment of the original peptide. The free energy barrier obtained for the acylation step is $14.6 \text{ kcal}\cdot\text{mol}^{-1}$, which corresponds to the first step of

the overall reaction. The k_{cat} reported for the overall reaction for this enzyme with the substrate Ac-Tyr-Val-Ala-Asp-pNA is 0.78 s^{-1} .¹⁰² This value is equivalent to an activation free energy barrier of $17.6 \text{ kcal}\cdot\text{mol}^{-1}$ accordingly to the transition state theory. In this way, the result obtained by the ASM is covered under the ceiling established for the experimental result. Further studies must be performed to select the actual E-S structure, as from here at least two possible scenarios can be envisaged. In the first one, the amino leaving group is involved in the de-acylation process as was observed in the reaction mechanism of the 3CL^{pro} of SARS-CoV-2,¹⁰³ (see appendix 2). In the second scenario, the leaving group is diffused to the bulk of the solvent decreasing the energy of the system. And from here, the de-acylation reaction could occur as in the putative proposal for the de-acylation step for cysteine proteases. This is plausible as the interactions made by the amino leaving group with the histidine are not as favorable as the ones observed in the case of the leaving group in the 3CL^{pro} of SARS-CoV-2 enzyme (appendix 2).¹⁰³

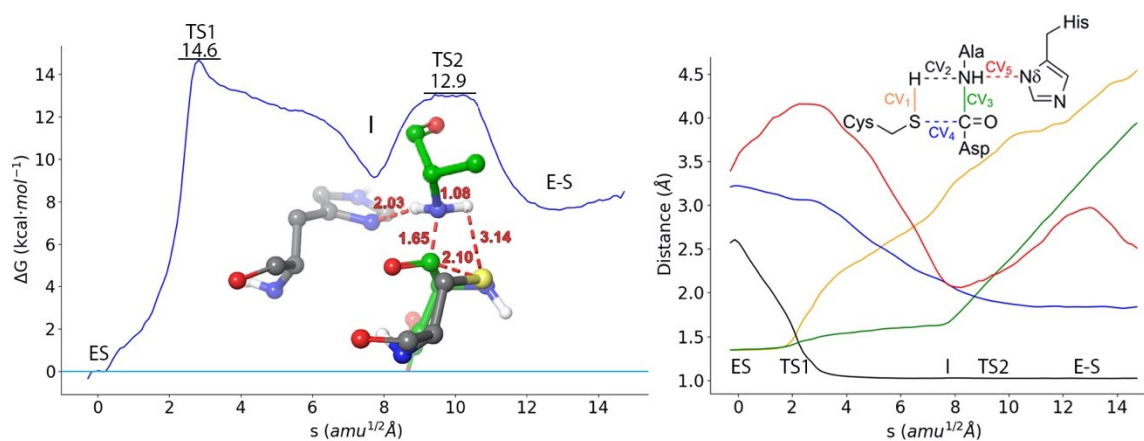


Figure 14. Caspase-1 acylation process. Left) B3LYPD3/6-31G*/MM free energy profile along the path-CV for the formation of the covalent E-S complex from the non-covalent (ES) one, the structure in balls and sticks is showing the atomic distribution of the intermediate I. Right) Evolution of the selected CVs along the MFEP.

3.2 3CL^{pro} of SARS-CoV-2

3.2.1. Peptide substrate

As described in the methodology section of appendix 2, a set of classical molecular dynamics simulations were performed for the enzyme 3CL^{pro} of SARS-CoV-2 in complex with the peptide substrate Ac-Ser-Ala-Val-Leu-Gln-Ser-Gly-Phe-NMe, see **Figure 15**. By analyzing these simulations we quantified the number of hydrogen bond contacts between the different groups of the substrate and the enzyme residues, which are depicted using a color scale in **Figure 16**. A detailed description of the interactions established between the enzyme and the substrate is more profoundly described in the results section of the appendix 2. Nevertheless, it is worth to mention that the glutamine in P1 position is strongly stabilized by hydrogen-bond interactions with the enzyme with both the side chain and the backbone atoms (see **Figure 15** and **Figure 16**), which explains the selectivity of the enzyme by this residue at the P1 position of the substrate. Considering that this active site is so well suited for glutamine, it is highly desirable to continue preserving all these interactions in the future design of any inhibitor of 3CL^{pro} of SARS-CoV-2.

The remaining interactions described between the enzyme-substrate complex along the S1-S5 subsites by our simulations agrees with the observations made on the crystallographic structures of the 3CL^{pro} of SARS-CoV-2 with inhibitors bound in the active site^{32,70} and its ortholog enzyme in SARS-CoV.¹⁰⁴ From our simulations it was possible to obtain valuable information that can help to develop new and more potent inhibitors. For example, our simulations show some important specific interaction of the P1' and P2' groups in the C-terminal fragment that are not present in the X-ray structures which are usually obtained with inhibitors that do not have these groups.

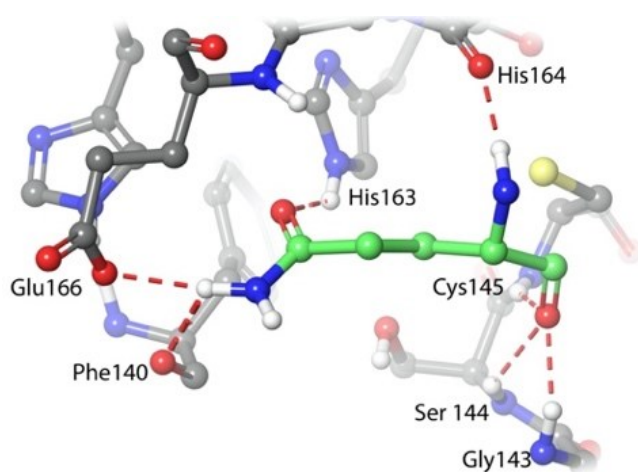


Figure 15. Stabilization of the glutamine in the substrate P1 position by hydrogen-bond interactions with the active site residues of the 3CL^{pro} of SARS-CoV-2 enzyme.

Interactions Protein-Substrate Protomer A

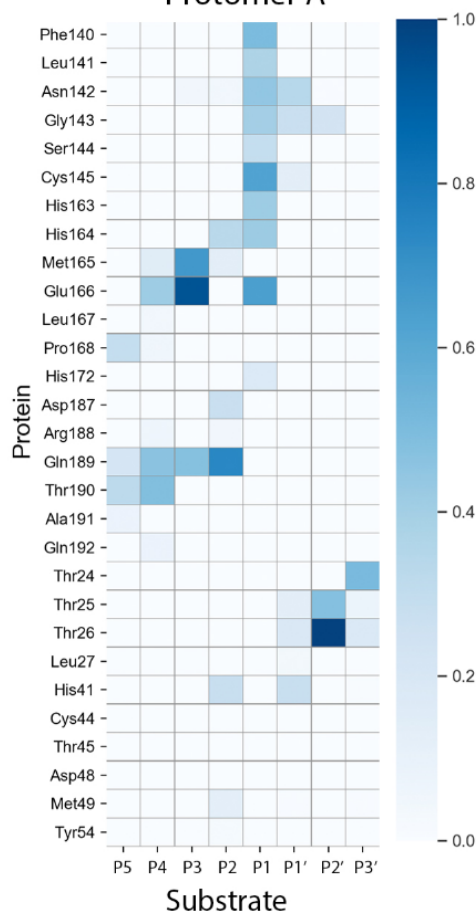


Figure 16. Substrate-protein fraction of hydrogen-bond contacts.

An analysis of the probability density of the distances between the Cys145-S γ atom to the carbonyl carbon atom in the substrate C(P1) and to the His41-N ϵ was performed over the classical molecular dynamic simulations. This analysis shows a bimodal distribution for the distances S γ -C(P1) (red curve in **Figure 17**), the peak observed at 3.4 Å correspond to the *trans* conformation of the cysteine side chain while the one at 4.7 Å is related to the *gauche* conformation. As it is show in the same figure, despite the cysteine side chain rotation, the sulfur atom remains always close to the His41-N ϵ atom, with the highest probability density found at 3.3 Å. By this and considering that there is also a water molecule (highly conserved among crystallographic structures) hydrogen bonded to the His41-N δ and to the Asp187 side chain, the basicity of the His41-N ϵ atom will be increased, so the His41-N ϵ atom acts as a base and abstracts the proton from the Cys145.

The role of the His41 during the catalysis and the possibility of finding the catalytic dyad in the ion-pair configuration (Cys⁻/HisH⁺) in the apo and holo states was also studied using a B3LYPD3/MM scheme with the 6-31+G* basis set. As it is show in the Figure 2 of the appendix 2, the neutral configuration of the catalytic dyad is more stable than the ion pair in both the apo and holo forms. However, the ion-pair configuration for the apo state is just 1.9 kcal·mol⁻¹ more stable than in the holo state. These free energy differences are related to the degree of solvation of the negatively charged sulfur atom. Finally, a low free energy barrier for the reverse proton transfer from the HisH⁺ to the Cys⁻ was observed, revealing

that the ionic form of the catalytic dyad is most probably a metastable transient species found during the reaction process.

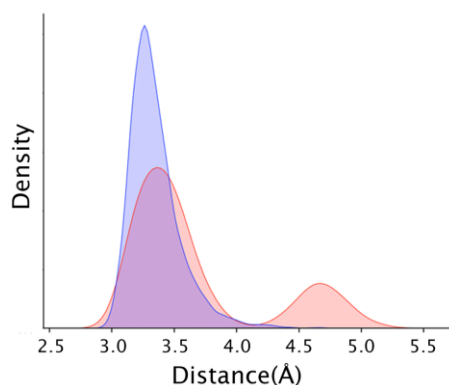


Figure 17. Probability densities of the distances between the Cys145- $S\gamma$ and the carbonyl carbon in substrate in P1 in red and to the His41-N ϵ in blue.

As previously described, the ion pair formation is a critical step, since the acylation reaction occurs from this transient configuration of the catalytic dyad. The first event involves the abstraction of the proton from the protonated His41 by the N(P1') atom of the substrate. While this proton is being transferred, the nucleophilic attack is performed and the peptide bond is broken. Even while concerted, these events occur in an asynchronous way, because the proton transfer is the process that expedites the rupture of the peptide bond. The total barrier obtained for the acylation step was of 14.6 kcal·mol⁻¹. This value includes the energy needed to form the ion-pair (red and black curves in **Figure 18**). This value strongly agrees with the reported values for peptides cleaved at the Gln-Ser bond by their ortholog enzyme, the 3CL^{Pro} of SARS-CoV. These values—derived from the steady-state rate constants measured at 25°C—range between 16.2 and 17.2 kcal·mol⁻¹.⁶¹ As these values were reported for the overall reaction and the acylation is not considered the limiting step, our results are under the upper limit value for the acylation barrier. And finally the results for this first step of the reaction help to explain the experimental observation about two protons on flight during the acylation step.⁶¹ One of them at the early stage of the reaction, when the ion pair is formed and the other associated to the TS structure, i.e. when the His41 is deprotonated by the N(P1') atom.

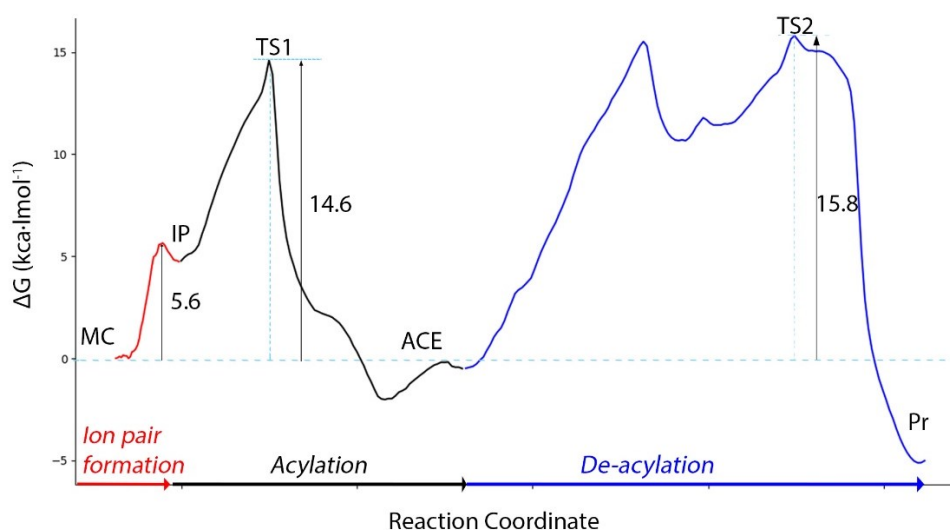


Figure 18. Free energy profile for the proteolysis mechanism in SARS-CoV-2 3CL^{Pro} at the B3LYPD3/6-31G*/MM level.

Regarding the de-acylation step, an alternative reaction mechanism to the standard one was proposed¹⁰³ from the behavior of the classical molecular dynamic simulations in the structure of the acyl-enzyme complex. In these simulations, it was observed that a water molecule is placed between the P'-NH₂ leaving group and the acyl-enzyme complex, suggesting that the water molecule can be deprotonated by this NH₂ terminal group present in the leaving group and not necessarily by His41. The terminal NH₂ group is known to be a better base than the histidine side chain.¹⁰⁵ The activation energy of 17.5 kcal·mol⁻¹ obtained for this step agrees with the values measured in the ortholog protease of SARS-CoV (from 16.2 to 17.2 kcal·mol⁻¹).⁶¹ The entire energy profile for this system can be seen in the **Figure 18** and the schematic representation of the reaction mechanism is presented in **Figure 19**. A detailed description can be found in appendix 2.

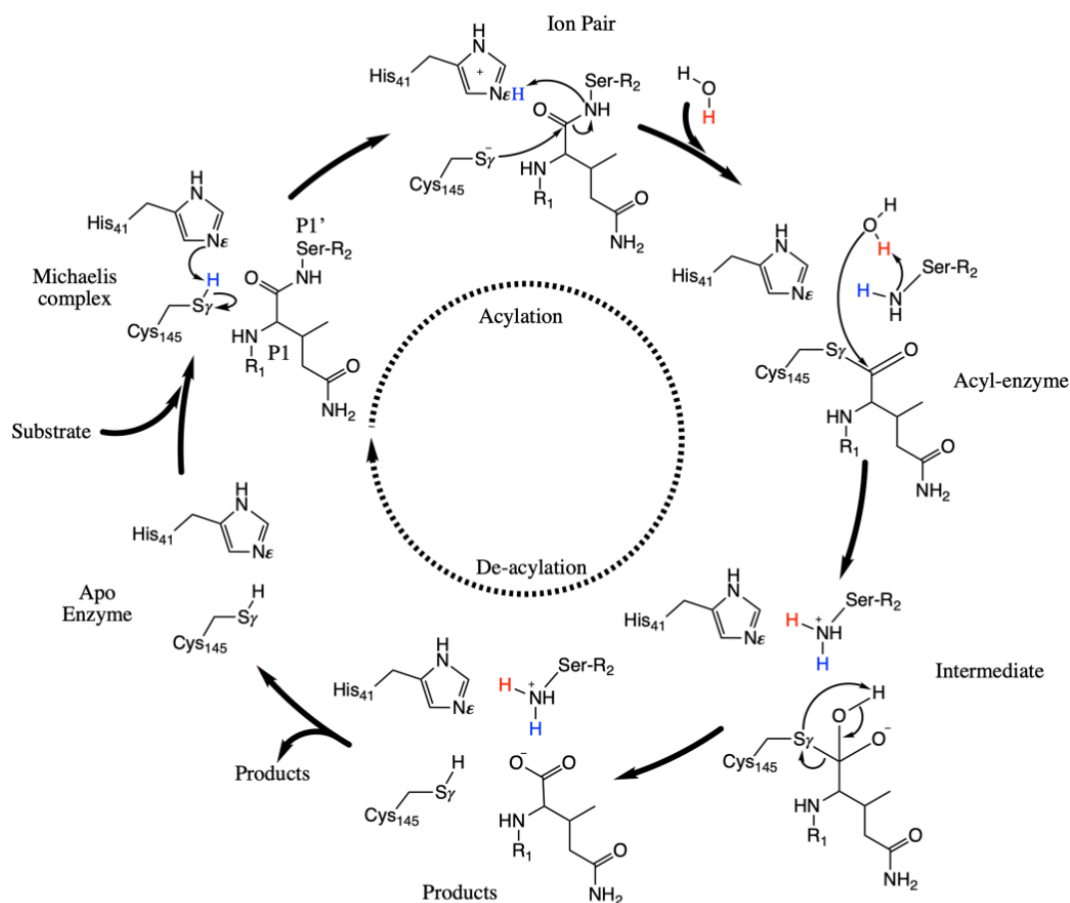
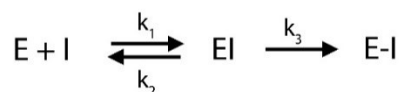


Figure 19. Schematic representation of the proteolysis mechanism.

3.2.2. Inhibitors

Having understood the key interactions in the enzyme-substrate complex as well as their chemical reactivity, it becomes clear that a similar study over the lead compounds that have demonstrated to be effective inhibiting the activity of the 3CL^{Pro} of SARS-CoV 2 will bring valuable information. By this, new strategies could be envisaged to design more potent inhibitors. The inhibitors selected for this study were the N3 Michael acceptor,³² the 11a with an aldehyde warhead,⁷⁰ the nitrile inhibitor PF-07321332 and the hydroxymethyl ketone based inhibitor PF-00835231,¹⁰⁶ see **Figure 20**. These inhibitors follow the next kinetic scheme:



Each inhibitor (I) binds reversibly into the active site of the enzyme (E) forming a non-covalent complex (EI) with a dissociation constant $K_I = k_2 / k_1$. From EI the catalytic cysteine forms a bond between its S_γ atom and the electrophilic carbon in the reaction center of the inhibitor with a k_3 rate constant. The product of this bond formation is the acyl-enzyme complex (E-I). The binding mode and the reactivity of every inhibitor in the EI complexes were studied by classical and hybrid QM/MM molecular dynamic simulations.

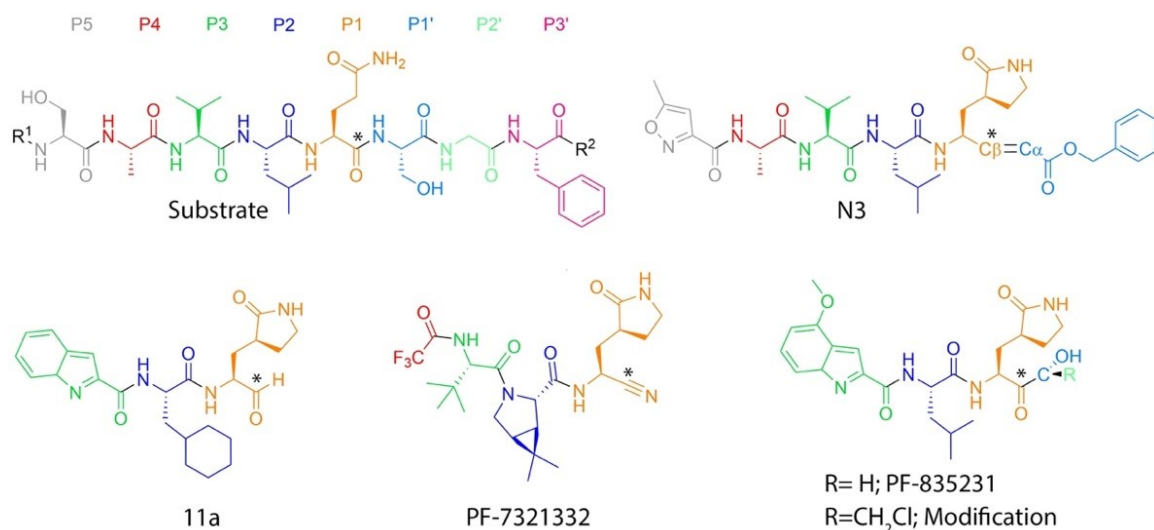


Figure 20. Chemical structures of the 3CL^{pro} of SARS-CoV-2 substrate and inhibitors included in the present work. The inhibitors P fragments are colored using the same color chart for the substrate. The * is pointing the electrophilic carbon in the reaction center.

At first sight, a common aspect in all of them can be identified, the presence of a γ -lactam ring in the P1 side chain position. The role of this P1 group is to keep the interactions observed for the glutamine at P1 position of the substrate in the active site. **Figure 21**, which shows the hydrogen bond contacts between the enzyme and the different substrates, demonstrates that the γ -lactam ring in the P1 position of the inhibitors can establish a very similar interaction pattern than the original substrate.

The hydrogen bond analysis between the 3CL^{pro} of SARS-CoV-2 and its substrates shows that almost all the inhibitors replicate the interactions established by the substrate with the enzyme, especially by the P1, P2 and P3 fragments. The inclusion of P1' and P2' fragments in the development of new inhibitors can be one of the aspects to consider, because most of the inhibitors studied neglect these contributions while the attention stays centered in the regions between P1 to P3. Nevertheless, the inhibitors envisaged to establish better interactions by P1' and P2' must not produce a leaving group after the acyl-enzyme complex is formed.

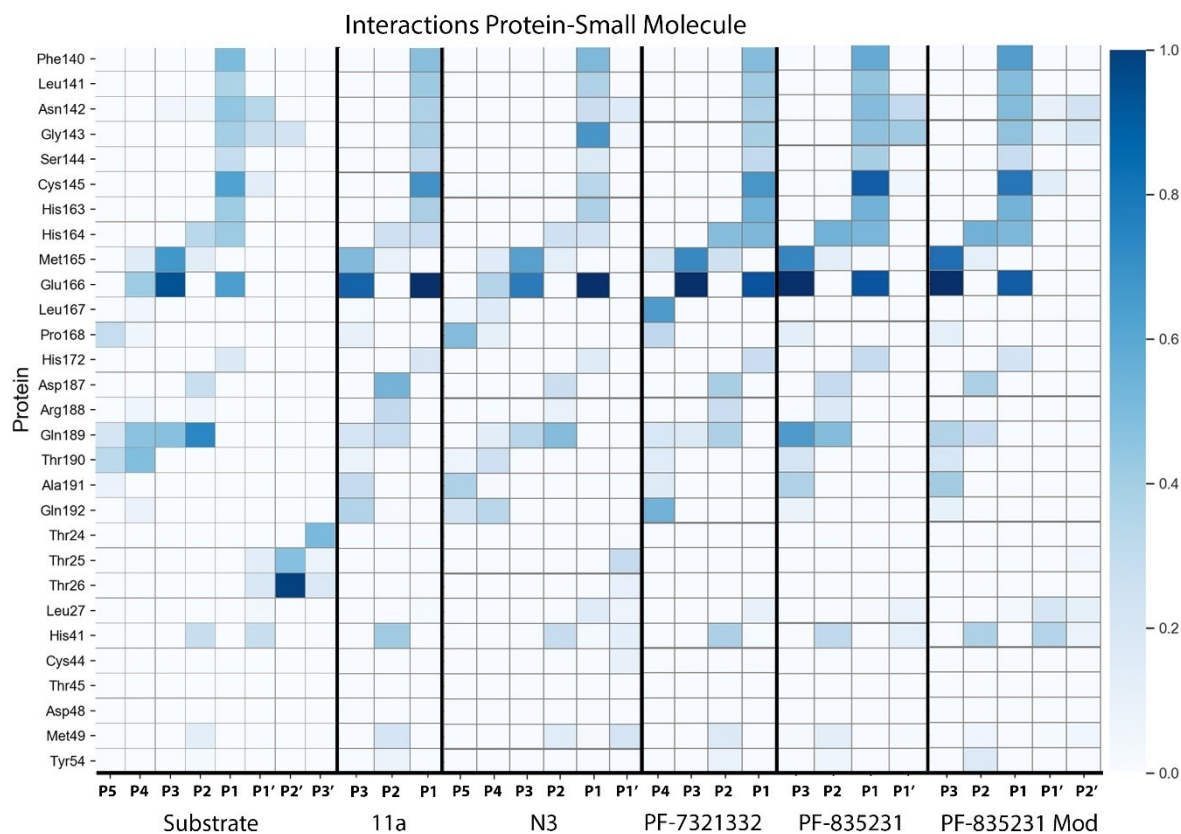


Figure 21. Hydrogen bond analysis between the SARS-CoV-2 of 3CL^{pro} and different small molecules in its active site. The color intensity is proportional to the occupancy of the interactions.

The Michael acceptor inhibitor (N3) and the Hydroxymethyl inhibitor (PF-00835231) take advantage of this feature, as they are the only inhibitors studied with a P1' fragment. However, it must be warranted that a balance between the P1'-enzyme interactions and the degree of solvation the S_{γ} atom of Cys145 can be achieved after the formation of the ion pair. For example, in presence of the aforementioned inhibitors, the energy needed to form the ion pair structure increases compared to this transformation in presence of the peptide substrate or in the apo state (see **Figure 22**). As explained in appendix 3 the ionized catalytic dyad in presence of the N3 inhibitor is stabilized by a water molecule that enters into the active site once the ion pair is formed. In presence of the inhibitor PF-00835231 (see appendix 6) this stabilization is provided by the hydroxyl group in the P1' region. Nevertheless, for the PF-00835231 inhibitor this stabilization is smaller than in the case of the substrate, where it is the serine side chain hydroxyl group of the P1' residue who stabilizes the S_{γ} atom of Cys145 (appendix 2). This is due to the larger conformational freedom of the hydroxymethyl group at the P1' position of PF-00835231 compared to the serine side chain at the equivalent position of the peptide substrate, as the latter is restricted by the substrate backbone. A similar restriction was imposed to the OH group in the PF-00835231 inhibitor when a chloromethyl group was added substituting the pro-R hydrogen of the hydroxymethyl group (see **Figure 20**). Because of this, the energy required to form the ion pair decreased. As seen in **Figure 22** the free energy cost to form the ion pair was even lower than in the case of the 11a inhibitor with a small aldehyde warhead.

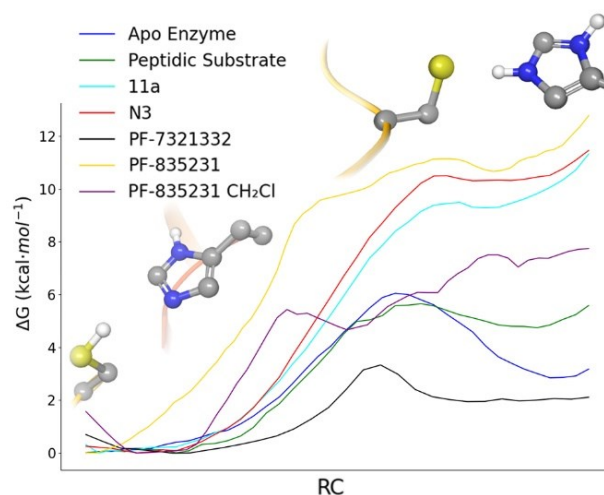


Figure 22. B3LYPD3/6-31+G*/MM free energy profiles for the Ion pair formation. Different RCs were employed in different cases, see appendixes 2-5.

Despite the 11a and the PF-07321332 are the inhibitors with the less bulky warheads of the group of inhibitors studied, aldehyde and nitrile respectively, and both lack the P1' fragment, their ion pair formation energies are very different, $9.3 \text{ kcal}\cdot\text{mol}^{-1}$ in the former and $2.9 \text{ kcal}\cdot\text{mol}^{-1}$ in the later. This difference can be related to the conformational freedom of the warhead, as explained in appendix 4. Rotation of the aldehyde group and the associated terminal hydrogen atom hinders a closer approach of water molecules and their correct positioning to stabilize the ionic form of the catalytic dyad. This aspect was improved in the design of the nitrile inhibitor PF-07321332. The Inclusion of the small and linear nitrile group facilitates the approach of water molecules increasing the stabilization of the negatively charged S_{γ} atom in the ion pair, as explained in the appendix 5.

For the five inhibitors in **Figure 20**, the formation of the covalent enzyme-inhibitor complex (E-I) was studied starting from their non-covalent form (EI). The profiles obtained are shown in **Figure 23**. When available, the predicted activation free energy agree quite well with the values derived from the experimental rate constants.^{107,108} A detailed explanation of the methodology used in these studies can be found in the appendix 3 for N3,¹⁰⁹ appendix 4 for 11a,¹¹⁰ appendix 5 for PF-07321332¹¹¹ and in appendix 6 for the inhibitor PF-00835231 and its chloromethyl modification.¹¹² The free energy profiles shown in **Figure 23** demonstrates how important the ion pair stabilization is in the whole kinetics of the covalent inhibition of SARS-CoV-2 3CL^{pro}. Those compounds presenting the lower free energy cost to form the ion pair in the catalytic dyad, also present the lower activation free energies.

For example, the PF-07321332 inhibitor is the one with the lowest energetic barrier for the proton transfer between Cys145 and His, forming the ion pair (IP1 in **Figure 23c**). Nevertheless, some extra energy has to be invested to separate the ion pair, approaching the nucleophile to the electrophilic carbon atom and to position the water molecule involved in the reaction in order to form the reactive IP2 structure (see **Figures 4c and 4d** in appendix 5). Additionally, when the mobility of the OH group in PF-00835231 is restricted by the inclusion of a chloromethyl group at the P1' position, the ion pair is better stabilized than in the original design (PF-00835231) and the chemical reaction occurs with an activation free energy slightly smaller than that of PF-07321332.

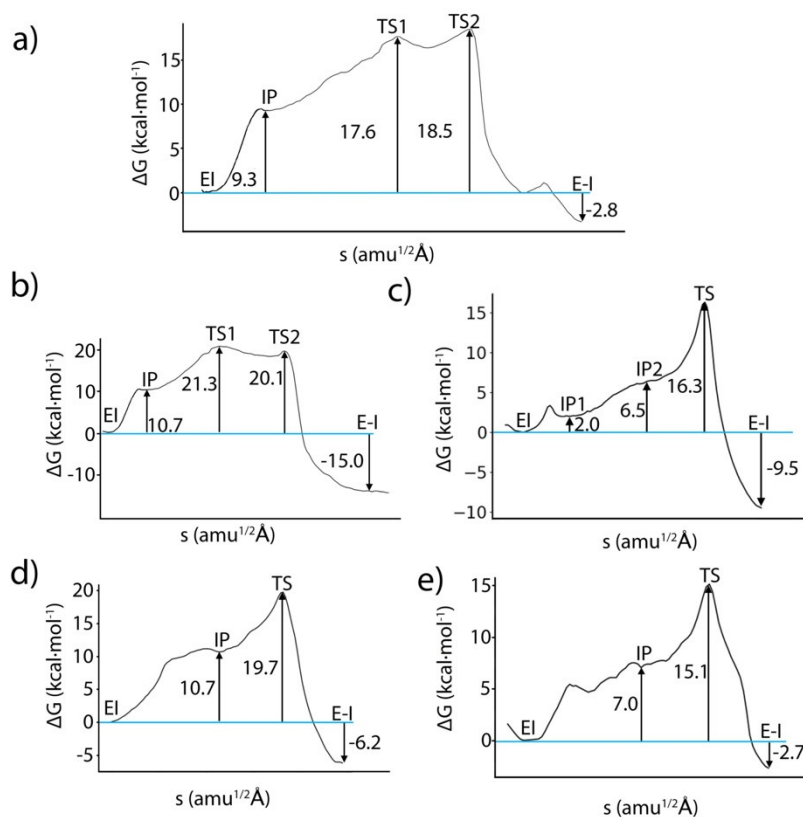


Figure 23. B3LYPD3/6-31+G*/MM free-energy profiles along the path-CV for the whole transformation of the non-covalent complex (EI) to the covalent one (E-I) through formation of the IP for the inhibitors of the enzyme 3CL^{pro} a) 11a, b) N3, c) PF-07321332, d) PF-00835231, e) PF-00835231-CH₂Cl modification.

A comparison among the TS structures of the enzyme-substrate complex and those from the enzyme-inhibitor complexes (see **Figure 24**) shows a key difference between the proteolysis reaction and the covalent inhibition processes. In the former, after formation of the ion pair, the reaction proceeds with the nucleophilic attack of the S_Y atom on the electrophilic carbon atom and the proton transfer from His41 to the leaving amino group and the concomitant breaking of the peptide bond. Instead, in the case of the inhibitors the proton transfer from His41 to the substrate is not direct but mediated by a water molecule or a hydroxyl group (in the case of the PF-00835231 design). Thus, the covalent inhibition process of SARS-CoV-2 3CL^{pro} involves a proton shuttle mechanism. The water molecule involved in the reaction with the 11a, N3 and PF-07321332 inhibitors is taking the role of the N(P1') atom in the reaction enzyme-substrate, i.e., it is abstracting the His41 proton and withdrawing electron density from the C(P1) carbon. During the covalent inhibition process the water molecule occupies the position of the N(P1') atom in the proteolysis reaction (see **Figure 24**). The proton accepted by the water molecule is finally transferred to a carbon atom of the double bond in N3, to the carbonyl oxygen atom in 11a and to the nitrogen atom of the nitrile group in PF-07321332. For all the mechanisms studied here, the charge deficient carbon atom recovers its lost charge by accepting the electrons of the S_Y atom of Cys145. For the N3 inhibitor the S_Y nucleophilic attack is the first step of the chemical reaction followed by the proton transfer. However, the nucleophilic attack and the proton transfer are almost concerted in all the other systems. Maybe this separation between the two steps is the reason why the activation energy for the reaction of the N3 with the enzyme is the highest for the six systems analyzed. In the case of the PF-00835231 inhibitor, the hydroxyl group at P1' is taking the position of water molecule, see **Figure 24**.

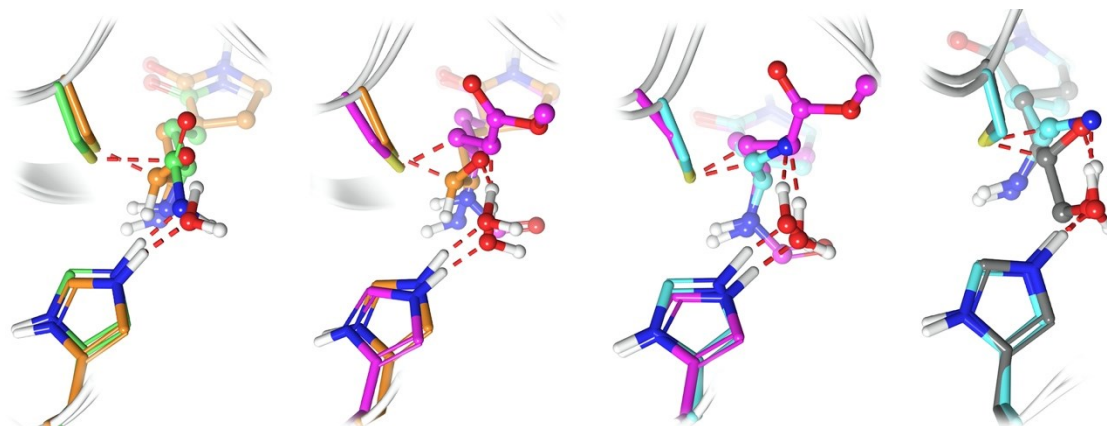


Figure 24. Transition state structures for small molecules studied. Catalytic dyad in licorice. Substrate in green, 11a in orange, N3 in purple, PF-07321332 in cyan and PF-00835231 in gray.

The contribution of the hydroxymethyl group to the binding free energy of the inhibitor PF-00835231 was also evaluated, see **Figure 25**. For that, alchemical transformations were made (the details of this calculation can be found in the table S1 of the appendix 6). A small improvement in the binding free energy was observed by changing the aldehyde hydrogen atom for the hydroxymethyl group. This can be explained by the new protein-inhibitor interactions established between the residues Asn142 and Gly143 in the enzyme and the new P1' fragment included in the inhibitor, see **Figure 21**. And finally, the inclusion of the chloromethyl group, in addition to improving the kinetic profile of the reaction, does not worsen the free energy of binding.

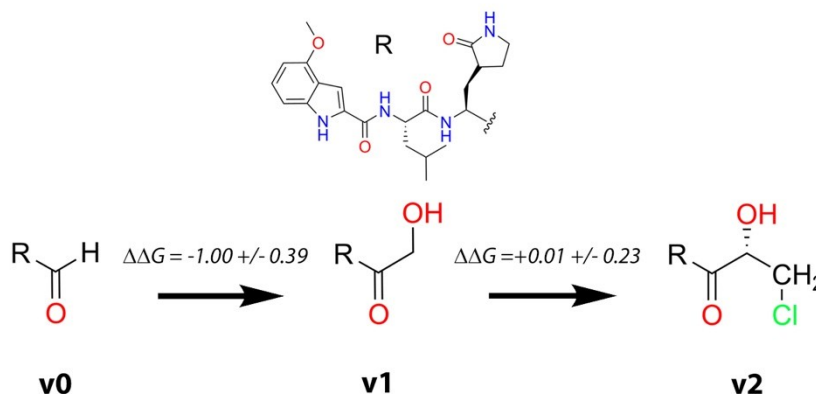


Figure 25. Alchemical transformation made to evaluate the impact of the modification of the hydroxymethyl warhead. The original inhibitor PF-00835231 is denoted as v1, the aldehyde derivative as v0 and the inhibitor with the chloromethyl substituent as v2.

So far, the discussion has been centered on the role of the fragments P1 and P1' in the inhibitors. To get additional information that helps to improve the inhibitors design, an analysis of the contributions to the binding free energy of every side chain in the P_i fragments was made. Alchemical transformations were made over the PF-07321332 inhibitor, as it is the one with the most favorable kinetic profile. This study is described in detail in appendix 5. The study shows that there is still some room for the improvements in the design of this inhibitor. While the fragments with the largest contribution to the binding free energy are the side chains in P1 and P2, the impact of side chain mutations in P3 and P4 by a methyl group have a modest effect in the binding free energy of the inhibitor, see **Figure 26**. That means that the positions P3 and P4 could be changed to improve the binding strength or even removed in order to explore different and stronger interactions towards P1' and P2' without having to stray too far from Lipinski rules.¹¹³

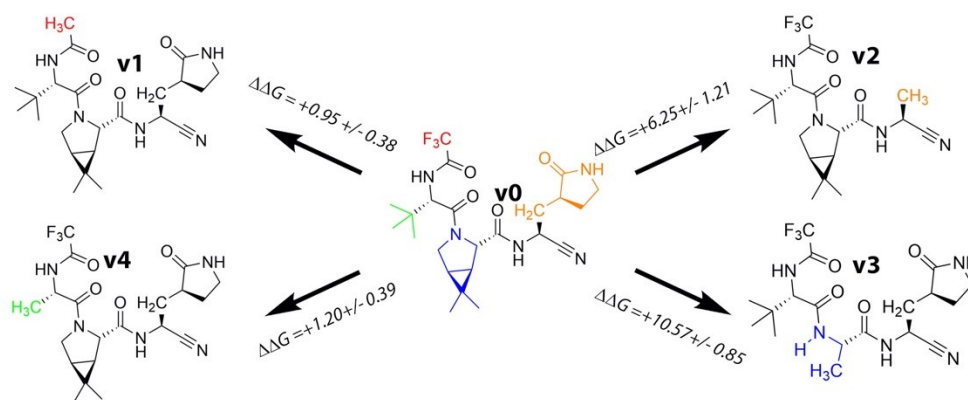


Figure 26. Binding free energy changes associated with the transformation of different P_1 groups of the PF-07321332 inhibitor into methyl groups. P_1 in orange, P_2 in blue, P_3 in green and P_4 in red.

3.3 Final remarks

In this doctoral thesis two enzymes have been studied, the 3CL^{pro} of SARS-CoV-2 and the human caspase-1. Both enzymes catalyze the same proteolytic activity by means of a His-Cys catalytic dyad. The proteolysis involves two steps. In the first one, the acylation, a peptide bond in the substrate is broken. As a consequence, the C-terminal fragment of the substrate is eliminated, while the N-terminal fragment is covalently bonded to the enzyme. In the second step, de-acylation, this N-terminal fragment is released. During the acylation process the S_γ atom of Cys acts as nucleophile, this atom bonds to the electrophilic carbon of the peptide bond that must be broken while the amino group is protonated. Along the de-acylation step a water molecule breaks the sulfur-carbon bond formed between the enzyme and the substrate, releasing the N-terminal fragment of the substrate. Despite these similarities between both enzymes, the results obtained in this doctoral thesis show important mechanistical differences between them.

In the 3CL^{pro} enzyme of SARS-CoV-2 the catalytic cysteine is placed near the catalytic histidine. This positioning facilitates the formation of an ion pair structure after a direct proton transfer from the catalytic cysteine to the histidine. The free energy cost associated to this charge separation process is compensated by the easiness of the subsequent steps: the nucleophilic attack of the cysteine sulfur atom to the electrophilic carbon atom of the substrate and the protonation of the amino group. On the other hand, the architecture of the caspase-1 active site is fundamentally different in terms of the disposition of the two members of the catalytic dyad. Not only the distance between the cysteine and the histidine is larger than the distance found between these two residues in 3CL^{pro} of SARS-CoV-2, but also the substrate is placed in between these two residues impeding the direct proton transfer between them (see **Figure 11**). For that reason, the mechanistical proposal found for the proteolysis in caspase-1 is completely different to that proposed for SARS-CoV-2 3CL^{pro}. The acylation process in caspase-1 involves a direct proton transfer from the catalytic cysteine to the substrate followed by the nucleophilic attack of the sulfur atom on the electrophilic carbon of the substrate. The catalytic histidine in this case acts merely as a stabilizing factor for the protonation of the leaving group. However, this process does not seem to affect the free energy barrier, obtaining similar activation free energy values for the acylation step in the two enzymes.

The stability of the intermediate formed after acylation is another fundamental difference between both enzymes. In the case of the 3CL^{pro} of SARS-CoV-2 the leaving group interacts favorably with the catalytic histidine, assisting the positioning of a water molecule to perform the de-acylation process. As seen in **Figure 27**, in the case of caspase-1, when the peptide bond is being broken the distance between the

amino group of the C-terminal leaving fragment and the catalytic histidine is increased, leading to a less stable intermediate complex. This difference may determine the participation or not of the leaving fragment in the diacylation step.

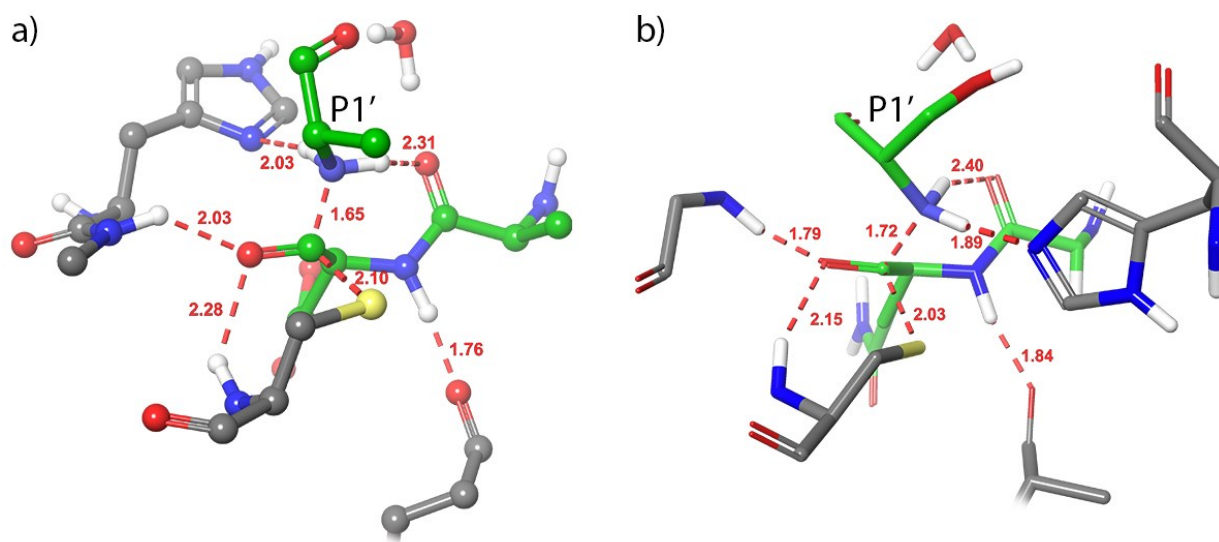


Figure 27. Configurations of the intermediates formed after the acylation reactions. a) caspase-1, b) 3CL^{pro} of SARS-CoV-2.

Another fundamental difference is found in the specificity of both enzymes. The caspase-1 breaks a peptide bond formed between an aspartate and an alanine residues while the 3CL^{pro} of SARS-CoV-2 breaks the peptide bond between a glutamine and a serine. The differences observed in the interactions of both enzymes with their substrates are obviously reflected in the design of their respective inhibitors. Let us recall here the overwhelming presence of γ -lactam ring at the P1 position in 3CL^{pro} inhibitors, which takes advantage of the interactions established between this enzyme and the glutamine P1 group of its substrate. Both, the non-covalent interactions established between the enzyme and its putative inhibitor, and the mechanistic details must be considered in the design of specific inhibitors for each of these cysteine proteases. We hope that the findings of this doctoral thesis will help future advances in this area.

It is finally worthy to add a comment on the reliability of the computational simulation methods employed in this thesis. During the writing of this doctoral thesis Pfizer published new results about the PF-07321332 inhibitor, also known as PAXLOVID™.¹¹⁴ When Pfizer announced that this compound was under clinical trial, in Mars 2021, we decided to include it in our survey on 3CL^{pro} inhibitors. The results of our computational analysis were published in July 2021 (see appendix 5). In our computational study, started from the X-ray structure of a different inhibitor (11a, PDB code 6XHM), we modelled the non-covalent complex and then used QM/MM hybrid methods to study the formation of the covalent complex. The structure found for this complex is presented in **Figure 28**, together with the X-ray structure (with PDB code 7VH8) released several months after our study was published. The agreement between the predicted and the experimental structures is remarkable, showing the maturity reached by computational techniques in the study of biochemical processes and their ability to provide an accurate picture of the processes under analysis.

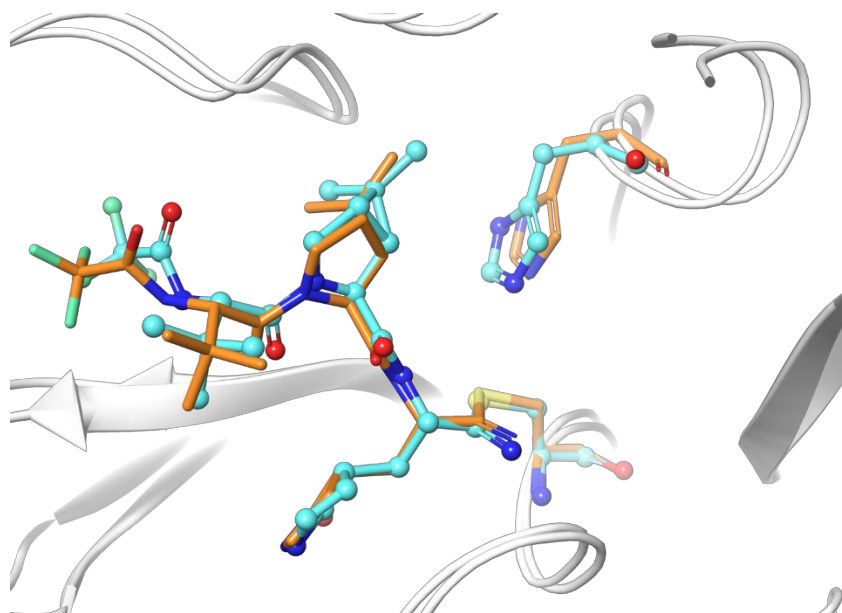


Figure 28. Product structure predicted by the string method for the reaction of 3CLpro with inhibitor PF-07332132, in orange, and the crystallographic structure published with PDB code 7VH8 in cyan.¹¹⁵

Resumen en lengua oficial

Mecanismos patogénicos de la enfermedad de Alzheimer

La Enfermedad de Alzheimer (EA) es una enfermedad neurodegenerativa que provoca problemas de memoria y altera la forma en la que pensamos y nos comportamos. Adicionalmente, en estadios avanzados los pacientes con esta enfermedad pierden el control del sistema nervioso central (SNC), afectando al control de su aparato locomotor y su movilidad. En la actualidad, no existe un consenso sobre la patogenia molecular exacta que conduce a esta muerte neuronal, principalmente porque hay muchos procesos involucrados a lo largo de su desarrollo.¹ Sin embargo, la hipótesis de la cascada del amiloide propuesta en 1992 por Hardy et al.⁷ es la más comúnmente aceptada, ver **Figura R 1**. Según esta hipótesis, la demencia observada es el resultado de la acumulación progresiva del péptido β -amiloide ($A\beta$) en forma de placas en el cerebro. Las pequeñas fibras (fibrillas) de β -amiloide se producen como consecuencia de la eliminación ineficaz de las proteínas en el cerebro.⁸ El proceso de agregación de estas fibrillas en placas desencadena la hiperfosforilación de la proteína tau. Dado que la proteína tau modula la estabilidad de los microtúbulos en los axones, su hiperfosforilación conduce a la pérdida de sus propiedades estructurales, separándola de los microtúbulos.⁹ Como consecuencia de esta separación, los microtúbulos se desmontan provocando la inestabilidad axonal, por lo que los axones no pueden realizar su tarea de transportar información y organelos a lo largo del sistema nervioso central provocando así muerte celular, daño vascular y formación de ovillos neurofibrilares de proteína tau, otra de las marcas característica de la EA.¹⁰ La eliminación de las placas $A\beta$ podría considerarse una forma lógica de tratar la enfermedad ya que, en principio, detendría todos los procesos que son desencadenados por su acumulación. Sin embargo, se ha observado que no existe una correlación directa entre la cantidad de placas $A\beta$ en el cerebro y el grado de demencia experimentado por el paciente en vida.¹¹ Además, algunos individuos con grandes depósitos de $A\beta$ en el cerebro nunca mostraron síntomas de la EA.¹²

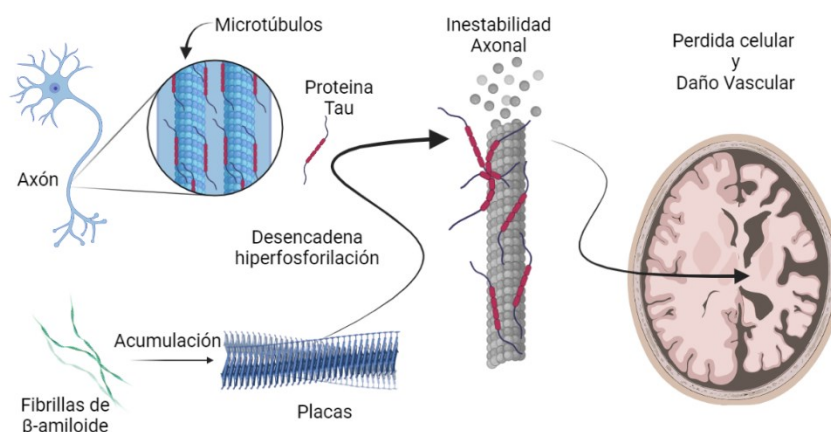


Figura R 1. Hipótesis de la cascada del amiloide. Figura creada con BioRender

Otra manera de abordar la enfermedad proviene de la hipótesis inflamatoria. Esta centra la atención en la activación del sistema inmune durante el desarrollo de la EA.¹³⁻¹⁵ El sistema inmune despliega una respuesta inflamatoria para proteger el tejido afectado de agentes externos como bacterias, infecciones, virus, toxinas o señales internas provenientes de daños celulares, por ejemplo. Esta respuesta se activa mediante la liberación de citocinas proinflamatorias. Mediante este proceso inflamatorio se consigue eliminar el agente externo causante de dicha señal, se depuran las células

necróticas y se repara el tejido.¹⁶ En condiciones normales, este proceso inflamatorio finaliza cuando el ataque ha sido neutralizado. Sin embargo, en la inflamación crónica este proceso regulador se rompe y, aun en ausencia de un agresor externo, el cuerpo humano desencadena un proceso autoinmune durante el cual ataca y destruye sus propias células sanas, las neuronas en el caso de la EA.¹⁷

Como se mencionó anteriormente, la característica principal de la EA es la formación de fibrillas de β -amiloide. En la hipótesis inflamatoria, estas proteínas plegadas irregularmente son fagocitadas por las microglías¹⁸ (ver **Figura R 2**). Este proceso activa el inflammasoma NLRP3 y, a continuación, la cisteína proteasa caspasa-1.¹⁹ La citocina inflamatoria IL-1 β es liberada por la actividad enzimática de la caspasa-1. Esta liberación de IL-1 β desencadena la hiperfosforilación de la proteína tau,²⁰ produciendo finalmente la muerte celular. Este proceso produce desechos neuronales, aumenta el metabolismo de la proteína precursora del amiloide (PPA) y conlleva la formación de más fibrillas de β -amiloide. Todos estos subproductos son reconocidos por las microglías y el proceso continúa una vez más.¹⁶ Se han observado niveles elevados de la cisteína proteasa caspasa-1 activa en tejidos cerebrales de humanos y ratones con esta enfermedad.²¹ Además, esta enzima también es responsable de la inflamación característica observada en enfermedades como la aterosclerosis,²² la artritis reumatoide²³ y la lesión cerebral traumática²⁴ entre otras.²⁵ De esta forma, el bloqueo de la actividad enzimática de la caspasa-1 mejoraría la condición de los pacientes con otras dolencias.

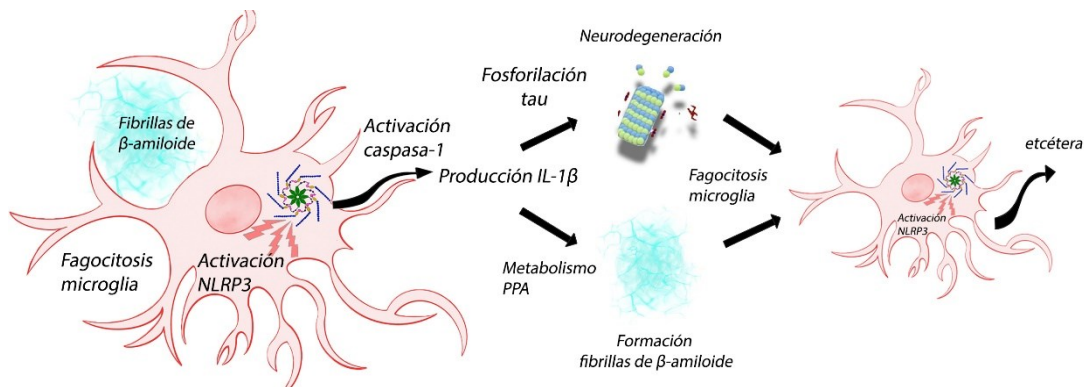


Figura R 2. Papel de la IL-1 β en la respuesta inflamatoria en la enfermedad de Alzheimer

La EA se encuentra entre las principales causas de muerte a nivel mundial¹ y más de 50 millones de personas en todo el mundo la padecen, es decir, el 0,6% de la población.² Si no se encuentra pronto una terapia, la EA tiene el potencial suficiente para afectar significativamente a la economía mundial y la salud,³ ya que el costo total de la atención de los pacientes con Alzheimer se ha estimado en unos 83.900 millones de dólares anuales, una cifra que aumentó en un 35% entre 2010 y 2015.⁴

Mecanismos patogénicos del COVID-19

La enfermedad del Covid-19 es una neumonía viral causada por el coronavirus de tipo 2 responsable del síndrome respiratorio agudo severo, SARS-CoV-2 (por su nombre en inglés *severe acute respiratory syndrome coronavirus 2*). El SARS-CoV-2 es un virus con una cadena sencilla de ARN de sentido positivo de 29.881 pares de bases de longitud que codifica 9860 aminoácidos,²⁶ los cuales forman tanto las proteínas estructurales, como las no estructurales, requeridas para el proceso de replicación del virus y posterior infección del huésped. Para iniciar el ciclo de replicación, la cadena de ARN contenida dentro del virión (partícula del virus que causa la infección) debe ingresar a las células del huésped para ser decodificada por su maquinaria de traducción.

Como se muestra en la **Figura R 3**, este proceso se ve facilitado por la interacción entre el receptor de la enzima convertidora de angiotensina 2, ACE2 (del inglés *angiotensin-converting enzyme 2*) en la membrana de la célula huésped y la proteína de la espícula en la superficie del SARS-CoV-2.²⁷ Después de este proceso de reconocimiento, las membranas del virus y el huésped se fusionan, posteriormente la cadena de ARN se libera al citoplasma para ser decodificada. Este proceso de transcoding es el mismo para todos los virus en el orden Nidovirales.²⁸ Eso significa que un tercio del genoma se transcribe a ARNm subgenómico, mientras que cada uno de los dos tercios restantes de la cadena contienen un marco abierto de lectura, ORF (del inglés *open reading frame*), rep1a y rep1b, que expresan dos grandes poli-proteínas coterminales, pp1a y pp1ab. Estas dos poli-proteínas codifican 16 proteínas no estructurales (nsp1 a nsp16). Dos de estas proteínas no estructurales son la proteasa tipo papaína o PLpro (nsp3) y la cisteína proteasa similar a 3 quimotripsina o 3CL^{pro} (nsp5). La primera se encarga de cortar entre las proteínas nsp1/2, nsp2/3 y nsp3/4, mientras que la segunda se encarga de las restantes 11 rupturas que formarán todo el complejo de transcripción-replicación del virus. A partir de la replicación subgenómica este ARNm se traduce a cuatro proteínas estructurales: la nucleocápside (N), la envoltura (E), la membrana (M) y la espícula (S). Estas proteínas estructurales se agrupan para formar el virión en el compartimento intermedio entre el retículo endoplasmático y el aparato de Golgi para formar los viriones maduros.^{29,30}

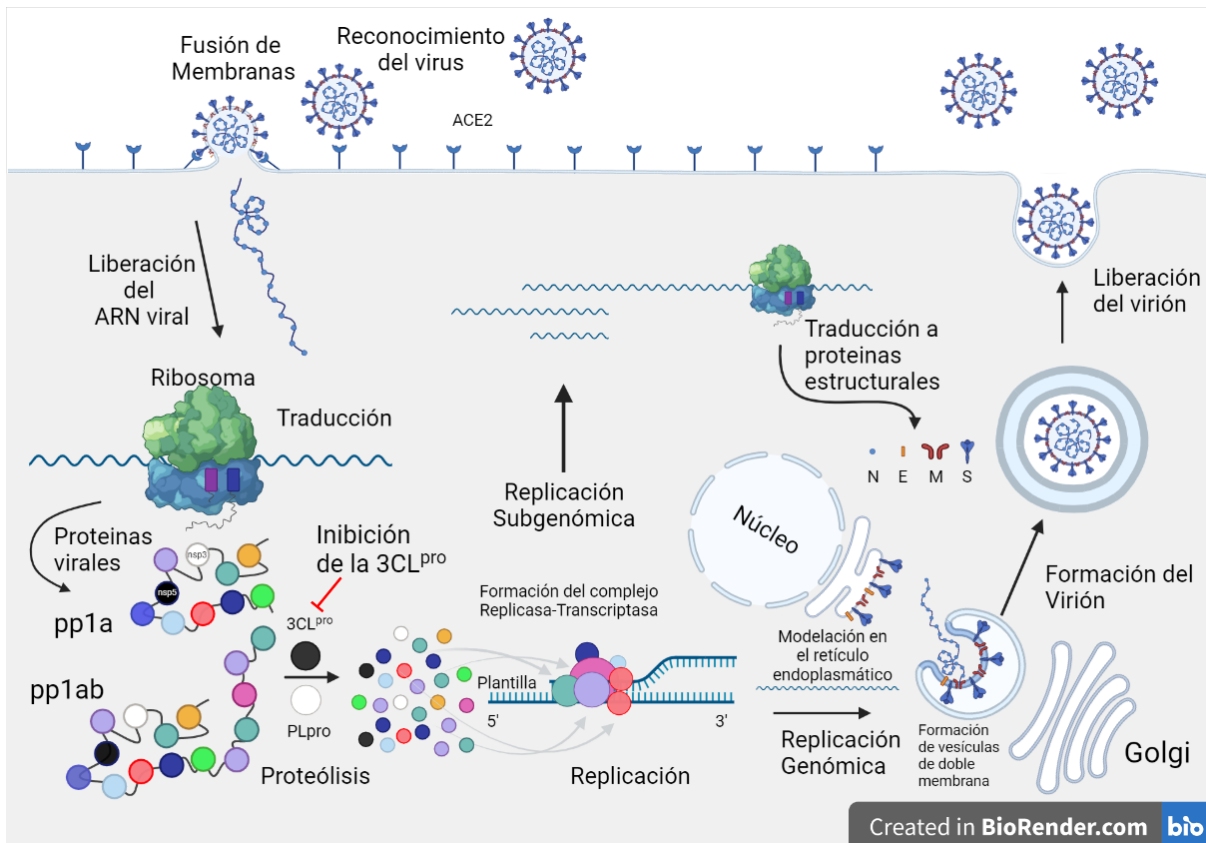


Figura R 3. Ciclo de replicación del SARS-CoV-2, en rojo se muestra donde el ciclo terminaría si se inhibiera la enzima 3CL^{pro}. Nucleocápside (N), la envoltura (E), la membrana (M) y la espícula (S).

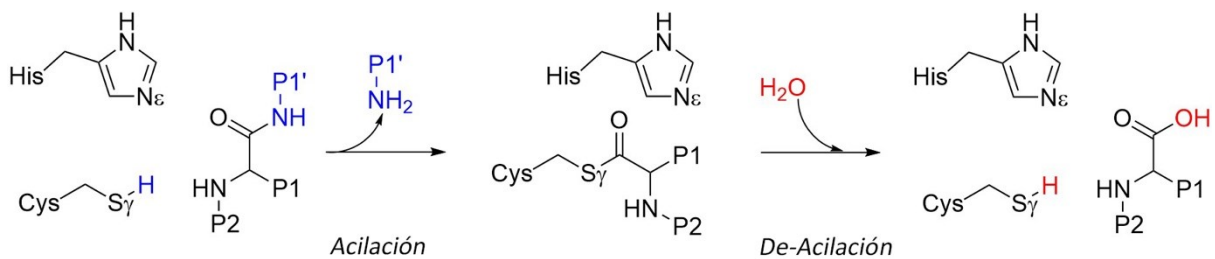
Este proceso se repite hasta que las células infectadas no pueden realizar sus funciones naturales y finaliza con la muerte del huésped. Por esa razón, el bloqueo de la cisteína proteasa 3CL^{pro} es un objetivo terapéutico importante contra la enfermedad Covid-19.^{31,32} Dado que el sitio activo de la 3CL^{pro} está altamente conservado entre los coronavirus humanos, el estudio de su actividad catalítica podría extrapolarse a otros coronavirus humanos.³³

La enfermedad de Alzheimer y el COVID-19 son muy diferentes entre sí, sin embargo, desde el punto de vista molecular tienen algo en común: la presencia de una cisteína proteasa en sus mecanismos patogénicos.

Cisteína proteasas

Las cisteína proteasas son enzimas con una cisteína cuyo grupo tiol se encarga de la actividad proteolítica de la enzima. Las cisteína proteasas contienen una díada catalítica conformada por el grupo imidazolio de una histidina y por el grupo tiol mencionado anteriormente. El mecanismo de reacción putativo para las cisteína proteasas en general se describe en el esquema R 1. Este mecanismo se divide en dos pasos: acilación y de-acilación.⁴⁵

Las cisteína proteasas se asocian en clanes, familias y subfamilias de acuerdo a las similitudes en el arreglo y secuencia de los amino ácidos alrededor de los residuos catalíticos. Para los lectores curiosos, una clara explicación y clasificación de cada una de las más de 100 cisteína proteasas encontradas hasta 2013 se resumen en el volumen 2 del Handbook of Proteolytic Enzymes.³⁴



Esquema R 1. Mecanismo de reacción propuesto para las cisteínas proteasas.⁴⁵ Siendo P1 la cadena lateral del Asp en la caspasas y Gln en la 3CLpro del SARS-CoV-2.

Caspasa-1

El nombre caspasa proviene de cisteinil-aspartato proteasa. Estas enzimas son muy específicas hidrolizando el enlace peptídico del sustrato formado entre un aspartato en la posición P1 (primera posición del sustrato anterior al enlace peptídico que se rompe) y el aminoácido adyacente en P1' (primera posición posterior). Los miembros de la familia de las caspasas se diferencian de otras peptidasas por esta estricta necesidad de un residuo de aspartato en la posición P1. Para el caso específico de la caspasa-1 y su sustrato la pro-IL-1 β , esta hidrólisis se realiza entre las líneas verticales en las secuencias Asp27-|-Gly28-Pro29 y Asp116-|-Ala117-Pro118.⁴⁶ Este último corte da como resultado la generación de la forma activa de 17,5 kDa de la IL-1 β .

El esquema R 1 muestra un mecanismo de reacción proveniente de otras cisteína proteasas que ha sido también adaptado a las caspasas. En esta propuesta mecanística, la cisteína se activa en el complejo de Michaelis tras la desprotonación de su grupo tiol por medio de la histidina presente en la díada catalítica, formando de esta manera el par iónico tiolato-imidazolio.⁴⁵ Esto puede ser probable para algunas cisteína proteasas como la catepsina B, la actinidina y la papaína, por ejemplo, en donde la distancia entre los átomos donador-aceptor de la díada están lo suficientemente cerca como para reaccionar (3,7 Å en la catepsina B, 3,2 Å en la actinidina y 3,9 Å en la papaína), ver **Figura R 4a**. Sin embargo, la situación en las caspasas es bastante diferente, ver **Figura R 4b**. Como se ha señalado previamente,^{52,53} los átomos donador-aceptor de la díada catalítica están más lejos que en las enzimas mencionadas anteriormente y adicionalmente el sustrato se sitúa entre la díada catalítica, lo que hace que la transferencia directa del hidrógeno sea poco probable (ver **Figura R 4b** y **Figura R 4c**). En múltiples estructuras cristalográficas de

la caspasa-1, estos átomos se encuentran a una distancia promedio de 5,2 Å, ver **Figura R 4c**, lo que hace poco probable que la activación ocurra por transferencia protónica directa entre la díada catalítica.

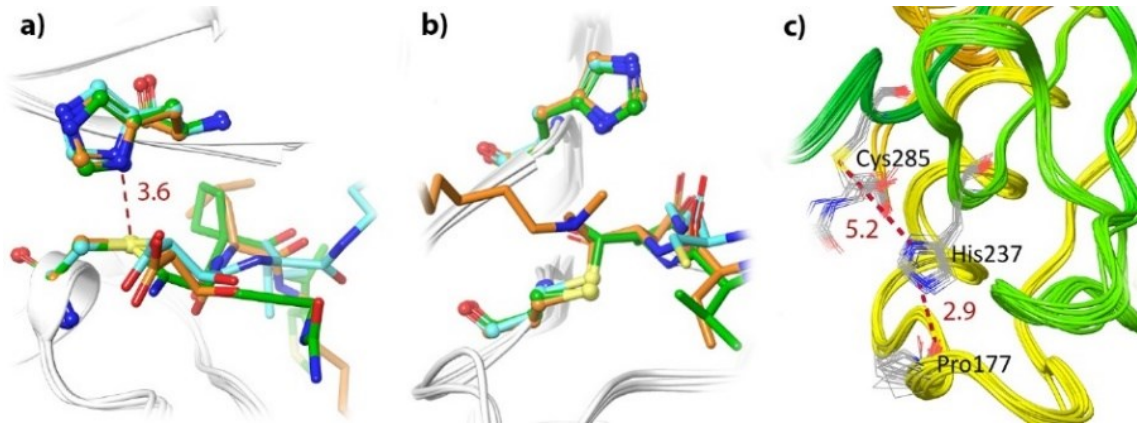


Figura R 4. Diada catalítica en diferentes cisteína proteasas a) papaína en cian, actinidina en naranja y catepsina B en verde (códigos PDB: 1PPP, 1AEC y 6AY2 respectivamente) b) caspasa -1 en naranja, caspasa-3 en cian (1RHJ) y caspasa-7 en verde (1F1J) c) Centro catalítico de estructuras cristalinas de caspasa-1 humana 'wild type' cocristalizadas con inhibidores que contienen un Asp en el fragmento central (códigos PDB: 6BZ9, 6F6R, 5MTK, 5MMV, 3NS7, 2HBQ, 1RWK, 1RWM, 1RWN, 1RWO, 1RWP, 1RWV, 1RWW, 1RWX, 1IBC y 1ICE) En rojo, las distancias medias de los átomos de nitrógeno del anillo imidazol de la His237 al átomo de azufre de la Cys285 y al átomo de oxígeno de Pro177 (distancias en Å).

Cabe mencionar que la His237 tiene, en principio, tres posibles estados de protonación: dos neutros, con los nitrógenos en épsilon o delta protonado y uno doblemente protonado, con carga neta positiva. En el mecanismo propuesto por Wilson et al.,⁴⁷ el estado protomérico asumido para el paso inicial involucra a la histidina con un protón en posición delta. Sin embargo, como se muestra en la **Figura R 4c**, el nitrógeno épsilon está a una distancia promedio de 2,9 Å del oxígeno del grupo carbonilo del aminoácido en el *loop* cercano a este (Pro177 en caspasa-1), lo que sugiere que el nitrógeno épsilon podría estar protonado. Este aspecto se abordó en detalle en el presente trabajo⁵³ antes de iniciar los estudios cinéticos.

3CL^{pro} del SARS-CoV-2

Las enzimas 3CL^{pro} están muy relacionadas estructuralmente con las cisteína proteasas de quimotripsina y picornavirus 3C, de donde tomaron su nombre. La 3CL^{pro} codificada en el ARN del SARS-CoV-2 comparte una estructura tridimensional bastante conservada comparada con su ortóloga en el SARS-CoV,⁵⁸ siendo esto una característica esperada ya que comparten el 96% de su secuencia de aminoácidos. Además, los residuos implicados en la catálisis, la unión y la dimerización están completamente conservados.⁵⁹ La 3CL^{pro} es una proteína no estructural, nsp (del inglés *non-structural protein*), de las poliproteínas replicasa 1a y 1ab (pp1a y pp1ab, respectivamente). Esta enzima también se conoce como la proteasa principal (M^{pro}) ya que participa en 11 de los 14 eventos de hidrólisis realizados sobre las poliproteínas traducidas del ARN viral.³⁰ La secuencia sobre la que realiza su actividad enzimática se ha conservado entre los coronavirus y es altamente selectiva para la secuencia Leu-Gln en las posiciones P2-P1 y Ser, Gly o Ala en la posición P1'. Desde P2 hacia la posición N terminal en el sustrato, la especificidad comienza a perderse. Lo mismo ocurre con las posiciones P2', P3' y así sucesivamente hacia la región del C terminal. Los estudios realizados sobre la 3CL^{pro} del virus de la bronquitis infecciosa aviar muestran que esta enzima es tan específica para la Gln en P1 que no se observa actividad in vitro cuando se cambia este aminoácido.⁶⁰

El mecanismo de reacción propuesto para esta enzima se divide en dos pasos, acilación y de-acilación, como se muestra en el esquema R 1, en el primer paso, el fragmento C terminal del sustrato se libera

después del rompimiento del enlace peptídico entre los fragmentos P1 y P1'. El fragmento N terminal del sustrato forma un complejo covalente con la enzima después del ataque nucleofílico del átomo S_{γ} de la Cys145 sobre el carbono del carbonilo de la Gln en P1. Para ello, este átomo de azufre debe ser desprotonado previamente y esto lo lleva a cabo la His41 en la díada catalítica.⁶¹ En este caso, a diferencia de las caspasas, los átomos donador y aceptor de la díada catalítica están cercanos y el sustrato no se posiciona entre los residuos de la díada catalítica, de tal modo que una transferencia directa del protón de la Cys145 a la His41 parece factible. Además, es bien sabido que este es el único papel químico que juega esta His en el paso de acilación, ya que los estudios experimentales muestran una falta de efecto cinético isotópico cuando la acilación se lleva a cabo en presencia de óxido de deuterio.⁶²

Inhibición enzimática

Como se explicó anteriormente, tanto en la patogénesis molecular de la EA como en la del COVID-19 interviene una enzima proteolítica, la caspasa-1 en la EA²¹ y la 3CL^{pro} en el COVID-19.⁶⁴ En principio, una posible estrategia para detener el avance de alguna de estas enfermedades consistiría en bloquear la actividad de la respectiva enzima. Dicho bloqueo se puede llevar a cabo usando una pequeña molécula llamada inhibidor. Este inhibidor se diseña de tal modo que se una a la enzima en el sitio activo en el que se lleva a cabo la acción enzimática sobre el sustrato natural. Es decir, esta molécula competirá con el sustrato por ocupar el sitio activo de la enzima. Por esta razón, este mecanismo de inhibición se conoce como inhibición competitiva (ver **Figura R 5**). La inhibición no competitiva es también otra estrategia utilizada, la cual está fuera del alcance del presente trabajo.

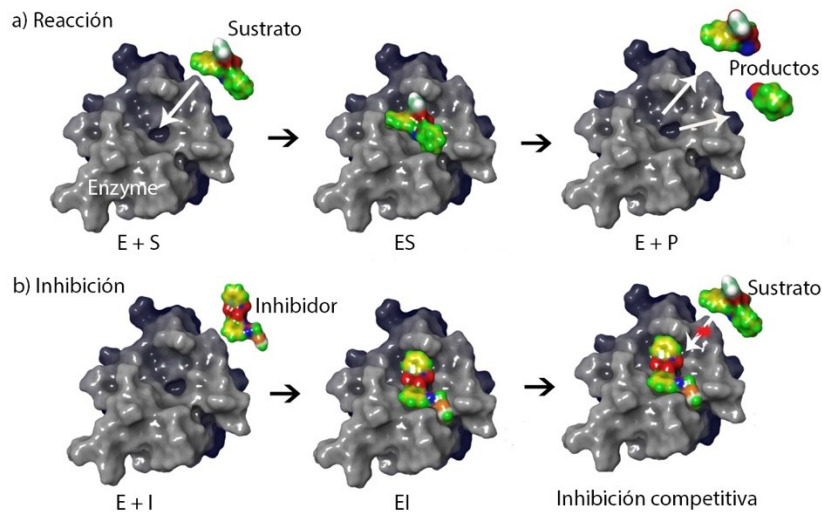


Figura R 5. Mecanismo de inhibición competitiva a) Reacción química catalizada natural entre la enzima y su sustrato en el sitio activo b) Mecanismo de inhibición por un inhibidor diseñado que previene la unión del sustrato en el sitio activo.

La inhibición enzimática puede ser reversible o irreversible. La mayoría de los inhibidores enzimáticos que se utilizan hoy en día como fármacos pertenecen al primer grupo.⁶⁵ Los inhibidores reversibles, en general, no están sujetos a una transformación química al interactuar con la enzima. Para esta clase de compuestos el complejo enzima-inhibidor está formado principalmente por interacciones no covalentes, como enlaces de hidrógeno, puentes salinos, interacciones de van der Waals y/o interacciones hidrofóbicas. Por el contrario, en la inhibición irreversible suele formarse un enlace covalente entre la enzima y el inhibidor. Estos reciben también el nombre de inhibidores covalentes. Dentro de lo posible, se ha evitado el uso de este tipo de inhibidores como fármacos porque suelen ser inespecíficos y pueden reaccionar con otro tipo de enzimas presentes en el organismo. Por esta razón,

su administración como terapia a largo plazo puede dar lugar a efectos secundarios graves.^{66,67} Sin embargo, existen nuevas evidencias que apuntan en la dirección opuesta.^{68,69} Para las dos enzimas contempladas en este trabajo el diseño de los inhibidores se ha basado en la estructura del sustrato^{67,70} y pueden ser reversibles o irreversibles dependiendo de la naturaleza del enlace covalente formado entre la enzima y el inhibidor. La fuerza de las interacciones covalentes y no covalentes puede evaluarse mediante métodos computacionales evitando esfuerzos en síntesis y ensayos enzimáticos.

Por lo expuesto hasta ahora sabemos que detrás de cada enfermedad hay una desregulación molecular y que al corregir estas desregulaciones, la probabilidad de curar la enfermedad es alta. ¿Sería posible desarrollar una cura para estas enfermedades utilizando sólo métodos computacionales? La respuesta corta es no. Porque hasta ahora no contamos con modelos lo suficientemente robustos que nos permitan calcular todos los parámetros farmacocinéticos y farmacodinámicos de los medicamentos en humanos. Sin embargo, las herramientas computacionales se incluyen hoy en día de forma rutinaria en todas las etapas del proceso de diseño de fármacos,⁷¹ aumentando así la velocidad y disminuyendo los costos asociados al desarrollo de moléculas candidatas a convertirse en fármacos. La inclusión de los métodos computacionales en el flujo de trabajo del diseño de fármacos aumenta las probabilidades de éxito y optimiza el proceso. Ambas enzimas, la caspasa-1 y la 3CL^{pro}, carecen de fármacos en el mercado que sean capaces de tratar las enfermedades en las que estas están involucradas.^{32,56} Por esto, es importante comprender la forma en que estas enzimas están interactuando con sus sustratos naturales y potenciales inhibidores. Identificando las interacciones clave enzima-sustrato y elucidando el mecanismo de reacción de dichas enzimas tanto con sus sustratos naturales como con sus correspondientes inhibidores covalentes. Sólo con esta descripción atomística del proceso de unión y una comprensión clara de la estructura del estado de transición (TS) involucrada en la reacción química, será posible diseñar moléculas estables capaces de unirse más estrechamente a la enzima que su sustrato natural.⁷²

Métodos computacionales en el diseño de medicamentos

Dinámicas Moleculares Clásicas

Las simulaciones de dinámica molecular clásica son el resultado de aplicar las leyes de la mecánica clásica para predecir las estructuras y propiedades de las moléculas.⁸⁰ En estas, la energía potencial del sistema está relacionada con variables geométricas tales como longitudes de enlace, ángulos de enlace, ángulos de torsión, etc. Para realizar estas simulaciones se omite la estructura electrónica. De esta manera los átomos se representan como esferas de van der Waals cargadas conectadas por osciladores o muelles. Así, la mecánica newtoniana clásica se puede utilizar para obtener las fuerzas que actúan sobre los átomos del sistema y predecir su desplazamiento en el tiempo. Las coordenadas iniciales de los átomos se obtienen mediante técnicas experimentales como resonancia magnética nuclear (RMN), cristalografía por difracción de rayos X o mediante la construcción de estructuras desconocidas haciendo uso del modelamiento por homología a partir de estructuras conocidas.⁸¹ La expresión matemática adecuada que describe la función de potencial recibe el nombre de campo de fuerza y tiene la siguiente forma:⁸⁴

$$V(r) = \sum_{\text{enlaces}} k_b(r - r_0)^2 + \sum_{\text{ángulos}} k_\theta(\theta - \theta_0)^2 + \sum_{\text{diedros}} k_\phi[1 + \cos(n\phi - \phi_0)] + \sum_{\text{no-enlace}} \left\{ 4\epsilon_{ij} \left[\left(\frac{\sigma_{0ij}}{R_{ij}} \right)^{12} - \left(\frac{\sigma_{0ij}}{R_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r R_{ij}} \right\}$$

Donde los términos denominados enlaces, ángulos y diedros describen las interacciones de enlace, mientras que los términos restantes describen las interacciones intermoleculares entre pares atómicos. Los dos primeros términos describen la contribución de energía potencial, como potencial armónico, debido al estiramiento y flexión de los enlaces cuando se desplazan de sus valores de referencia r_0 , y θ_0 ; siendo k_b y k_θ sus respectivas constantes de fuerza. El tercer término describe la interacción entre dos átomos separados por tres enlaces consecutivos que forman un ángulo diedro de rotación ϕ alrededor de un enlace central. La rotación a lo largo de este enlace se describe como una serie de cosenos de M términos para cada ángulo diedro, con una constante de fuerza k_ϕ para esta rotación, siendo ϕ_0 el ángulo de fase. El siguiente término es el potencial de Lennard-Jones 6-12 que describe las interacciones de van der Waals intermoleculares entre parejas de átomos i y j , siendo R_{ij} la distancia entre los dos átomos, σ_0 es la distancia a la que la interacción desaparece y $\varepsilon_{i,j}$ la energía de interacción en el fondo del pozo potencial. $\varepsilon_{i,j}$ y $\sigma_{0\ ij}$ son específicos para cada par de átomos i y j . Finalmente, el último término es un potencial de Coulomb que describe la interacción electrostática entre los átomos i y j con cargas parciales q_i y q_j a una distancia R_{ij} , siendo ε_0 y ε_r la permitividad del vacío y la constante dieléctrica del medio, respectivamente. Las interacciones electrostáticas y de van der Waals no se evalúan entre todos los posibles pares de átomos del sistema, sino solo para aquellos que se encuentran dentro de un radio de corte establecido durante el cálculo. Los términos k_b , r_0 , k_θ , θ_0 , k_ϕ , ϕ_0 , $\varepsilon_{i,j}$, σ_0 , q_i y q_j son parámetros que se especifican en el campo de fuerza.

El campo de fuerza utilizado en este trabajo fue el ff14SB.⁸⁵ Este contiene parámetros para los átomos que conforman una proteína hecha de aminoácidos estándar. Las moléculas de agua se describieron utilizando el modelo flexible del campo de fuerza TIP3P.⁸⁶

Dinámicas moleculares QM-MM

En las metodologías clásicas la distribución electrónica de las moléculas no se describe explícitamente. Por esta razón, los cambios en la estructura electrónica presentes en las reacciones químicas no serán posibles de describir mediante estos métodos. Para este tipo de estudios la mecánica cuántica es la herramienta indicada para abordar el problema, ya que la mecánica cuántica permite describir con precisión la estructura electrónica molecular. Sin embargo, en el caso de los procesos enzimáticos, realizar un cálculo de mecánica cuántica involucrando todos los átomos del sistema es una tarea no solamente imposible de realizar con los recursos computacionales disponibles, sino también innecesaria, ya que la mayoría de los cambios en la estructura electrónica durante una reacción química se dan sólo en un número reducido de átomos. De este modo, una descripción híbrida de Mecánica Cuántica/Mecánica Molecular (QM/MM, Quantum Mechanics/Molecular Mechanics) del sistema puede ser una estrategia adecuada. Con la metodología híbrida QM/MM, la mayor parte del sistema puede describirse de manera clásica mediante un campo de fuerzas, mientras que los átomos involucrados en la reacción química recibirán un tratamiento cuántico y la interacción entre ambas regiones viene descrita por la siguiente expresión:⁸⁶

$$E_{QM/MM} = \sum_{A \in MM} \int \rho_e \frac{Q_A}{|r - R_A|} dr + \sum_{A \in QM, B \in MM} \frac{Z_B Q_A}{R_{AB}} + \sum_{A \in QM, B \in MM} 4\varepsilon_{AB} \left[\left(\frac{\sigma_{AB}}{R_{AB}} \right)^{12} - \left(\frac{\sigma_{AB}}{R_{AB}} \right)^6 \right]$$

siendo el primer término las interacciones electrostáticas entre la densidad electrónica QM $\rho_e(r)$ y las cargas puntuales MM Q_A , el siguiente término describe la interacción entre la carga de los núcleos en la región QM Z_B , y las cargas puntuales MM Q_A y finalmente el último término corresponde a la interacción de van der Waals entre los átomos QM y MM.

Diferencias de energía libre

La probabilidad de encontrar un sistema molecular en un estado (A) u otro (B) está determinada por la diferencia de energía libre entre esos dos estados. O lo que es lo mismo, conociendo la probabilidad relativa de dos estados, se pueden deducir sus diferencias de energía libre:

$$\Delta G_{BA} = G_B - G_A$$

La energía libre de cualquier estado se puede obtener por medio de la relación proveniente de la mecánica estadística:

$$G_A = -k_B T \ln \left(Z^{-1} \int_A e^{\frac{-V(\mathbf{r})}{k_B T}} d\mathbf{r} \right)$$

Como no es posible evaluar esta integral en todo el espacio de fases, una alternativa práctica es calcular la diferencia de energía libre entre estados del sistema. Para ello se realizan simulaciones moleculares de estados cercanos en el espacio de fases.⁹¹ Dentro de las técnicas utilizadas en esta tesis para realizar un muestreo de dichos estados y así obtener las diferencias de energía libre se encuentran: la integración termodinámica,⁹² las simulaciones de MM/GBSA,⁹³ y “umbrella sampling”.⁹⁸ Una explicación detallada de la metodología utilizada en cada uno de los estudios llevados a cabo se puede encontrar en los apéndices que acompañan a este documento.

Resultados y conclusiones

Caspasa-1

Se evaluó la energía libre de unión entre una serie de ligandos congenéricos y la caspasa-1 usando MM/GBSA. Por medio de esta técnica fue posible establecer el estado de protonación más probable para la His237. Se observó que los valores calculados para la energía libre de unión se correlacionan mejor con los resultados experimentales cuando se supone que los residuos que conforman la díada catalítica, cisteína e histidina, se encuentran en estado neutro, con el protón situado en el N ϵ de la histidina y no en el N δ , véase la **Figura R 6**. En el presente estudio también fue posible establecer los residuos de la enzima que forman las interacciones clave en los complejos enzima-inhibidor, véase el apéndice 1. Se observó que la principal diferencia entre los inhibidores potentes y los débiles se relaciona con las interacciones que estos establecen con los residuos His342, Pro343 y Arg383 en las regiones S2 a S4 del sitio activo de la caspasa-1.⁵³ Por tanto, una posible estrategia para el diseño de nuevos y más potentes inhibidores para esta enzima debería centrarse en fortalecer las interacciones con esta región.

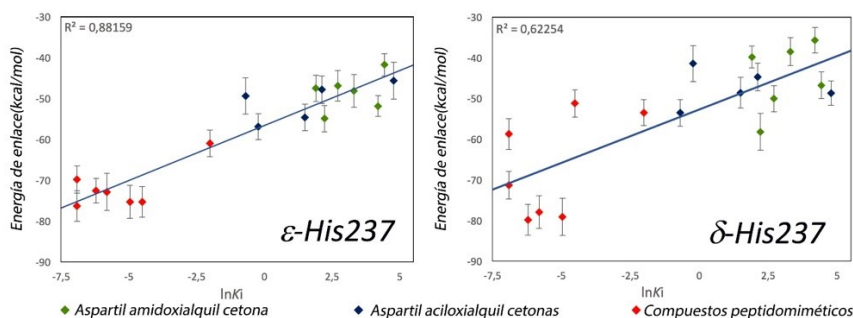


Figura R 6. Correlación entre la K_i experimental y la energía de enlace MM/GBSA calculada para los dos sistemas en dos diferentes estados tautoméricos para la His237. Panel izquierdo: ϵ -His237; panel derecho: δ -His237.

Utilizando el método de la cuerda adaptativa⁹⁷ exploramos la superficie de energía libre multidimensional para la primera etapa de la reacción catalizada por la caspasa-1 determinada por las cinco variables colectivas mostradas en el panel derecho de la **Figura R 7**. El panel izquierdo de la misma figura muestra una estructura intermedia asociada al mecanismo que se ha determinado en este trabajo. El mecanismo consta de dos pasos: *i*) La primera barrera corresponde a la transferencia protónica desde la cisteína catalítica al átomo de nitrógeno del enlace peptídico del sustrato (ver CV1 y CV2 en la **Figura R 7**); posteriormente se da el ataque nucleofílico del átomo de azufre sobre el carbono electrófilo del grupo carbonilo (como se muestra en la evolución de CV4). El intermedio se estabiliza mediante un fuerte puente de hidrógeno entre el nitrógeno del péptido protonado y la histidina catalítica (ver la evolución de CV5). *ii*) En el segundo paso del mecanismo de acilación se rompe el enlace peptídico (CV3 aumenta) formando el complejo E-S y liberando el primer fragmento del péptido original. La barrera de energía libre obtenida para el paso de acilación es de 14,6 kcal·mol⁻¹, que corresponde al primer paso de la reacción global. La k_{cat} reportada para la reacción global de esta enzima con el sustrato Ac-Tyr-Val-Ala-Asp-pNA es 0,78 s⁻¹.¹⁰² De acuerdo con la teoría del estado de transición, este valor es equivalente a una barrera de energía libre de activación de 17,6 kcal·mol⁻¹. De esta forma, el resultado obtenido por el método de la cuerda adaptativa está en buena concordancia con el resultado experimental. A partir de esta primera etapa se pueden contemplar al menos dos posibles escenarios. En el primero, el grupo saliente amino está involucrado en el proceso de de-acilación como se observó en el mecanismo de reacción de la 3CL^{pro} del SARS-CoV-2,¹⁰³ (ver apéndice 2) activando una molécula de agua que actuaría de nucleófilo para romper el enlace E-S. En el segundo, el grupo saliente se difunde al seno del solvente disminuyendo la energía del sistema. A partir de aquí, la reacción de de-acilación podría ocurrir mediante el mecanismo clásico establecido para el paso de de-acilación en otras cisteínas proteasas, donde la histidina catalítica activa la molécula de agua nucleofílica. Este segundo escenario sería posible ya que las interacciones del grupo saliente amino con la histidina no son tan favorables como las observadas en el caso del grupo saliente de la enzima 3CL^{pro} del SARS-CoV-2 (apéndice 2).¹⁰³

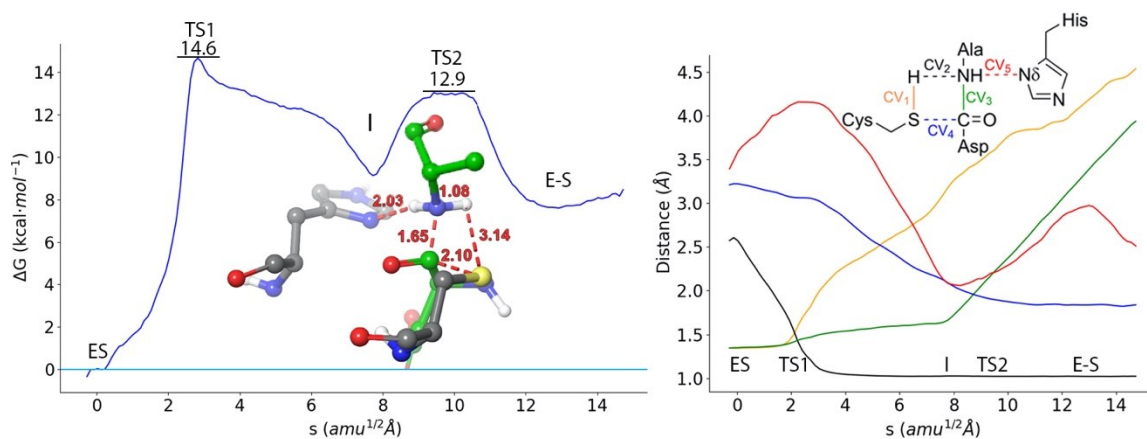


Figura R 7. Proceso de acilación para la caspasa-1. Izquierda) perfil de energía libre a nivel B3LYPD3/6-31G*/MM a lo largo de la coordenada de camino para la formación del complejo covalente E-S a partir del no covalente (ES), la estructura mostrada corresponde a la del intermedio I. Derecha) Evolución de las variables colectivas seleccionadas a lo largo del camino de mínima energía libre para la reacción

3CL^{pro} del SARS-CoV-2

En el caso de la proteasa principal del SARS-CoV-2, la díada catalítica está formada por los residuos His41 y Cys145. En primer lugar, se estudió el papel de la His41 durante la catálisis y la posibilidad de encontrar la díada catalítica en la configuración de par iónico (Cys/HisH⁺) tanto en los estados *apo* como *holo*. Para esto se utilizó un esquema de cálculo B3LYPD3/MM con el conjunto de base 6-31+G*. Para determinar

la diferencia de energía libre entre ambas formas. En todos los casos encontramos que la configuración neutra de la díada catalítica es más estable que el par iónico tanto en la forma *apo* como en la forma *holo* aunque la diferencia de energía libre depende del sustrato que se encuentre en el centro activo.

Partiendo de la estructura *holo* con la díada catalítica en la configuración de par iónico se estudió el mecanismo de reacción de la enzima en presencia de un péptido sustrato. Se encontró que el primer evento implica la abstracción del protón en el Nε de la His41 protonada por parte del átomo N(P1') del sustrato. Mientras este protón es transferido, se produce el ataque nucleofílico y se rompe el enlace peptídico. Estos eventos ocurren de manera concertada pero asincrónica. La barrera de energía libre obtenida para la etapa de acilación fue de 14,6 kcal·mol⁻¹.¹⁰³ Este valor incluye la energía necesaria para formar el par iónico (curvas roja y negra en la **Figura R 8**). Los resultados de este primer paso de la reacción ayudan a explicar la observación experimental que establece que hay dos protones intercambiándose durante el paso de acilación.⁶¹ Uno de ellos correspondería a la etapa inicial de la reacción (formación del par iónico) y el otro estaría asociado a la estructura del estado de transición, es decir, cuando la His41 es desprotonada por el átomo N(P1') del sustrato.

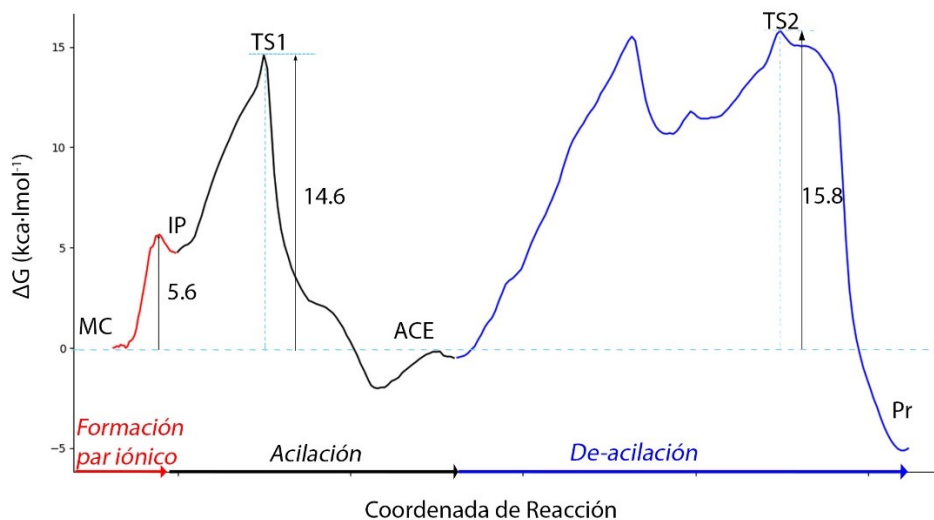


Figura R 8. Perfil de energía libre para el mecanismo de proteólisis de la 3CLpro del SARS-CoV-2 nivel B3LYPD3/6-31G*/MM

A partir del comportamiento de las simulaciones de dinámica molecular clásica en la estructura del complejo acil-enzima se propuso un mecanismo de reacción alternativo al estándar para el proceso de de-acilación en cisteína proteasas.¹⁰³ En estas simulaciones se observó que una molécula de agua se posiciona entre el grupo saliente P'-NH₂ y el complejo acil-enzima, lo que sugiere que la molécula de agua puede ser desprotonada por este grupo terminal NH₂ presente en el grupo saliente y no necesariamente por la His41. En la literatura se ha determinado que el grupo NH₂ terminal es una mejor base que la cadena lateral de histidina.¹⁰⁵ La energía de activación de 17,5 kcal·mol⁻¹ obtenida para este paso concuerda con los valores experimentales medidos en su enzima ortóloga del SARS-CoV (de 16,2 a 17,2 kcal·mol⁻¹).⁶¹ El perfil de energía completo para este sistema se puede ver en la **Figura R 8** y la representación esquemática del mecanismo de reacción se presenta en la **Figura R 9**. Una descripción detallada de este estudio se puede encontrar en el apéndice 2.

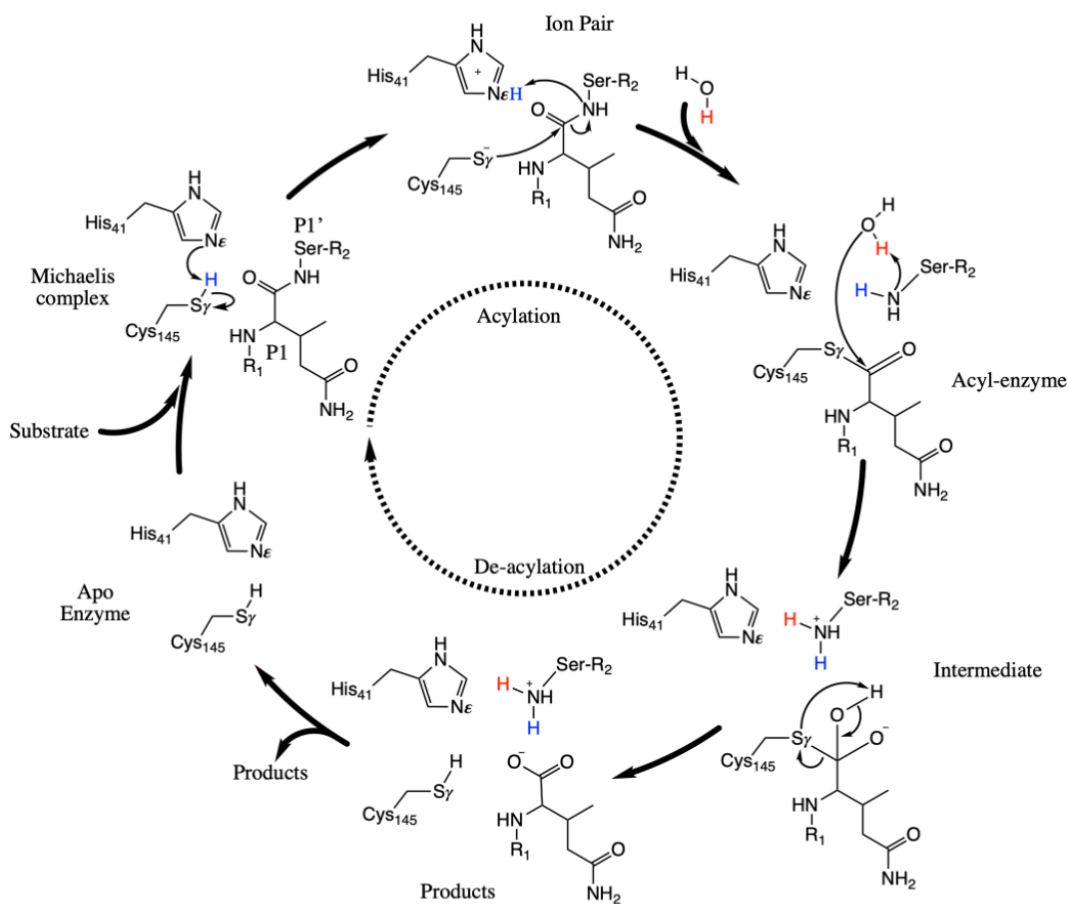


Figura R 9. Representación esquemática del mecanismo de proteólisis

Inhibidores de la 3CL^{pro}

Los inhibidores seleccionados para este estudio fueron el aceptor de Michael N3,³² el 11a con un grupo aldehído como cabeza de serie,⁷⁰ el inhibidor nitrilo PF-07321332 y el inhibidor basado en una hidroximetilcetona PF-00835231,¹⁰⁶ ver Figura R 10.

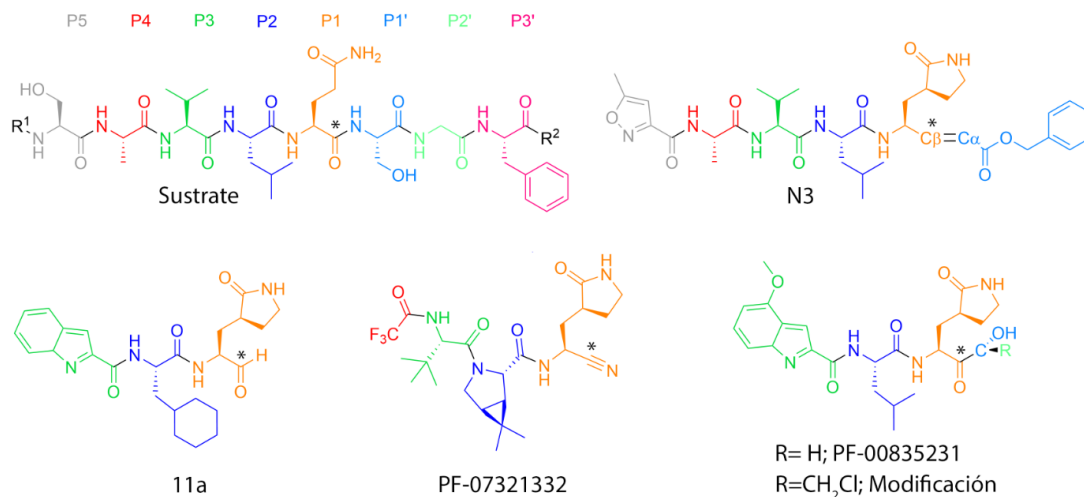


Figura R 10. Estructuras químicas del sustrato e inhibidores de la 3CL^{pro} del SARS-CoV-2 incluidos en el presente trabajo. Los fragmentos de inhibidores P se colorean utilizando la misma tabla de colores que para el sustrato. El * apunta al carbón electrófilo en el centro de reacción.

El análisis de puentes de hidrógeno entre la 3CL^{pro} del SARS-CoV-2 y sus sustratos, **Figura R 11**, muestra que casi todos los inhibidores replican las interacciones establecidas por el sustrato con la enzima, especialmente mediante los fragmentos P1, P2 y P3.

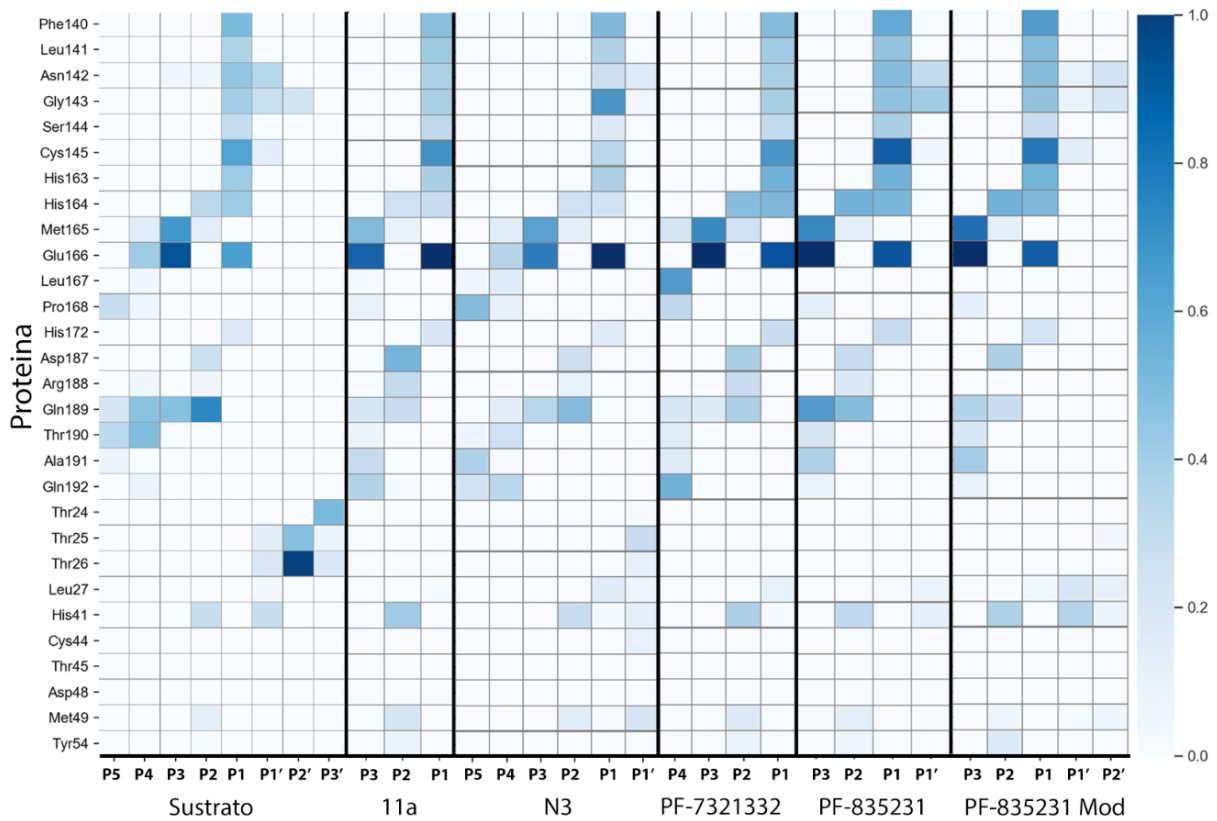


Figura R 11. Análisis de puentes de hidrógeno entre la 3CL^{pro} del SARS-CoV-2 y diferentes inhibidores o sustrato en su sitio activo. La intensidad de coloración correlaciona la fracción de tiempo que el puente de hidrógeno estuvo presente durante la simulación.

En presencia de los inhibidores mencionados anteriormente, la energía necesaria para formar la estructura del par iónico aumenta en comparación con dicha transformación en presencia del sustrato peptídico o en el estado apo (ver **Figura R 12**). Como se explica en el apéndice 3, la díada catalítica ionizada en presencia del inhibidor de N3 se estabiliza mediante una molécula de agua que entra en el sitio activo una vez que se forma el par iónico.¹⁰⁹ En presencia del inhibidor PF-00835231 (véase el apéndice 6), esta estabilización la proporciona el grupo hidroxilo en la región P1'.¹¹² Sin embargo, en el caso del inhibidor PF-00835231 esta estabilización es menor que en el sustrato, donde es el grupo hidroxilo de la cadena lateral de la serina en la posición P1' quien estabiliza el átomo de S_γ de la Cys145 (apéndice 2). Esto se debe a la mayor libertad conformacional del grupo hidroximetilo en la posición P1' del inhibidor PF-00835231 en comparación con la libertad conformacional de la cadena lateral de la serina, posición equivalente del sustrato peptídico, ya que este último está restringido por el esqueleto del sustrato. Con el objetivo de mejorar las características del inhibidor, se impuso una restricción similar al grupo OH en el inhibidor PF-00835231 añadiendo un grupo clorometilo sustituyendo el hidrógeno pro-R del grupo hidroximetilo (ver **Figura R 10**). Mediante este cambio, la energía libre requerida para formar el par iónico disminuyó. Como se ve en la **Figura R 12**, el costo de energía para formar el par iónico fue incluso menor que en el caso del inhibidor 11a que contenía una cabeza de serie pequeña como lo es el grupo aldehído.

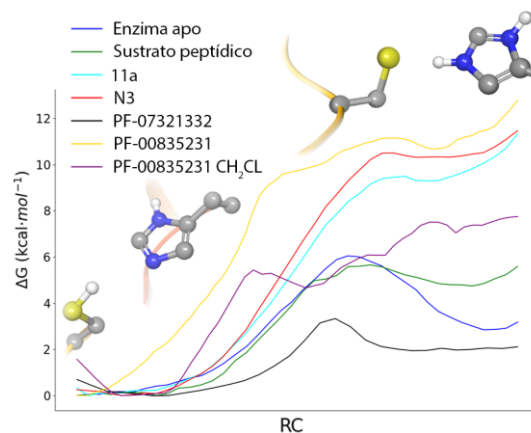


Figura R 12. Perfiles de energía libre B3LYPD3/6-31+G*/MM para la formación del par iónico en la díada catalítica de la enzima 3CLpro en la forma apo y en presencia de diferentes sustratos. Se emplearon diferentes coordenadas de reacción (RC) en cada caso, véanse los apéndices 2-5.

Para los cinco inhibidores de la **Figura R 10**, se estudió la formación del complejo covalente enzima-inhibidor (E-I) partiendo de su forma no covalente (EI). Los perfiles obtenidos se muestran en la **Figura R 13**. Las energías libres de activación predichas concuerdan con los valores derivados de las constantes de velocidad experimentales.^{107,108} Una explicación detallada de la metodología utilizada en estos estudios se puede encontrar en el apéndice 3 para N3,¹⁰⁹ apéndice 4 para 11a,¹¹⁰ apéndice 5 para PF-07321332¹¹¹ y en el apéndice 6 para el inhibidor PF-00835231 y su modificación con la inclusión del grupo clorometilo.¹¹² Los perfiles de energía libre que se muestran en la **Figura R 13** demuestran la importancia de la estabilización del par iónico durante todo el proceso cinético de inhibición covalente de la 3CL^{pro} del SARS-CoV-2. Así, aquellos compuestos que presentan el menor costo de energía libre para formar el par iónico en la díada catalítica, también presentan las energías libres de activación más bajas.

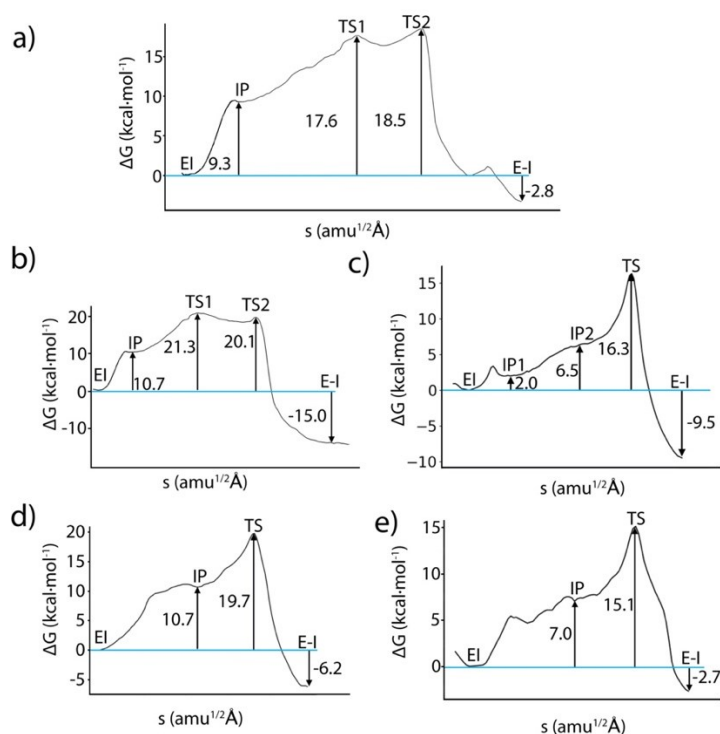


Figura R 13. Perfiles de energía libre B3LYPD3/6-31+G*/MM a lo largo del de la coordenada de camino para la transformación del complejo no covalente (EI) en covalente (E-I) a través de la formación del par iónico IP para los inhibidores de la enzima 3CL^{pro} a) 11a, b) N3, c) PF-07321332, d) PF-00835231, e) PF-00835231 Modificación -CH₂Cl.

Una comparación entre las estructuras de los estados de transición del complejo enzima-sustrato y las de los complejos enzima-inhibidor (ver **Figura R 14**) muestra una diferencia clave entre la reacción de proteólisis (con un péptido sustrato) y los mecanismos de inhibición covalente. En el primer caso, después de la formación del par iónico la reacción prosigue con el ataque nucleofílico del átomo $S\gamma$ de la Cys145 sobre el átomo de carbono electrófilo y la transferencia protónica de la His41 al grupo amino saliente y la correspondiente ruptura del enlace peptídico. En cambio, en el caso de los inhibidores, la transferencia protónica desde la His41 al sustrato no es directa, sino que está mediada por una molécula de agua o un grupo hidroxilo (en el caso del diseño PF-00835231). De este modo, el proceso de inhibición covalente de la 3CL^{pro} del SARS-CoV-2 implica un mecanismo de intercambio de protones. La molécula de agua involucrada en la reacción con los inhibidores 11a, N3 y PF-07321332 está tomando el papel del átomo de N(P1') en la reacción enzima-sustrato, es decir, está abstrayendo el protón de la His41 y disminuyendo la densidad electrónica del carbono C(P1), aumentando su carácter electrófilo. Durante el proceso de inhibición covalente, la molécula de agua ocupa la posición del átomo de N(P1') en la reacción de proteólisis (ver **Figura R 14**). El protón aceptado por la molécula de agua (o el grupo hidroxilo) se transfiere finalmente a un átomo de carbono del doble enlace en el caso del N3, al átomo de oxígeno del carbonilo en el del 11a y del PF-00835231 y al átomo de nitrógeno del grupo nitrilo en el del PF-07321332. Para todos los mecanismos estudiados aquí, el átomo de carbono deficiente en carga recupera su carga perdida aceptando los electrones del átomo $S\gamma$ de la Cys145. Para el inhibidor N3, el ataque nucleofílico del $S\gamma$ es el primer paso de la reacción química seguido de la transferencia de protones.

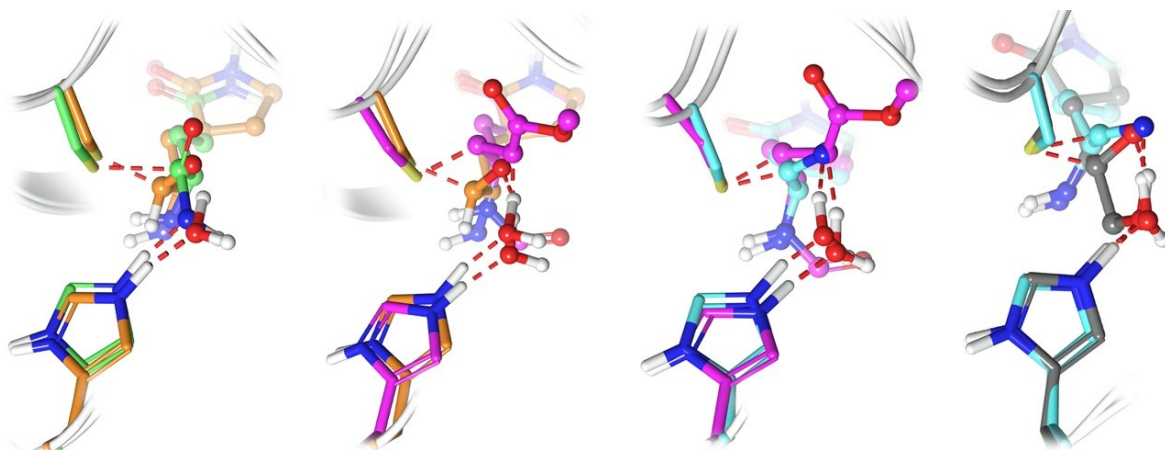


Figura R 14. Estructuras de los estados de transición para las moléculas estudiadas para la reacción de la enzima 3CL^{pro} del SARS-CoV-2 con diferentes compuestos. Diada catalítica en bastones. Sustrato peptídico en verde, inhibidor 11a en naranja, N3 en violeta, PF-07321332 en cian y PF-00835231 en gris.

Para obtener información adicional que ayude a mejorar el diseño de los inhibidores, se evaluó la contribución del grupo hidroximetilo a la energía libre de unión del inhibidor PF-00835231, ver **Figura R 15**. Para esto, se realizaron transformaciones alquímicas (los detalles de este cálculo se pueden encontrar en el apéndice 6). Se observó una mejora en la energía libre de unión al cambiar el átomo de hidrógeno del aldehído por el grupo hidroximetilo. Adicionalmente, la inclusión del grupo clorometilo sobre la posición pro-*R*, además de mejorar el perfil cinético de la reacción, no empeora la energía libre de unión.

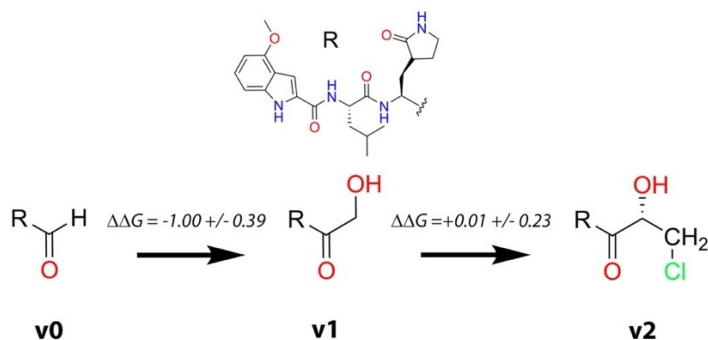


Figura R 15. Transformación alquímica realizada para evaluar el impacto de la modificación del hidroximetilo como cabeza de serie. El inhibidor original PF-00835231 se indica como v1, el derivado de aldehído como v0 y el inhibidor con el sustituyente clorometilo como v2.

Finalmente, se realizó un análisis de las contribuciones a la energía libre de unión de cada cadena lateral en los fragmentos de Pi. Se realizaron transformaciones alquímicas sobre el inhibidor PF-07321332, ya que es el que posee el perfil cinético más favorable. Este estudio se describe en detalle en el apéndice 5 y muestra que todavía hay espacio para mejorar su diseño modificando algunos de los fragmentos. Mientras que los fragmentos con la mayor contribución a la energía libre de unión son las cadenas laterales P1 y P2, el impacto de las mutaciones de la cadena lateral P3 y P4 por un grupo metilo tiene un efecto modesto en la energía libre de unión del inhibidor, ver **Figura R 16**. Esto significa que las posiciones P3 y P4 podrían cambiarse para mejorar la fuerza de unión, o inclusive eliminarse en el caso de los inhibidores N3, aceptor de Michael y el PF-00835231, hidroximetilcetona para explorar interacciones diferentes y más fuertes hacia P1' y P2' sin tener que alejarse demasiado de las reglas de Lipinski.¹¹³

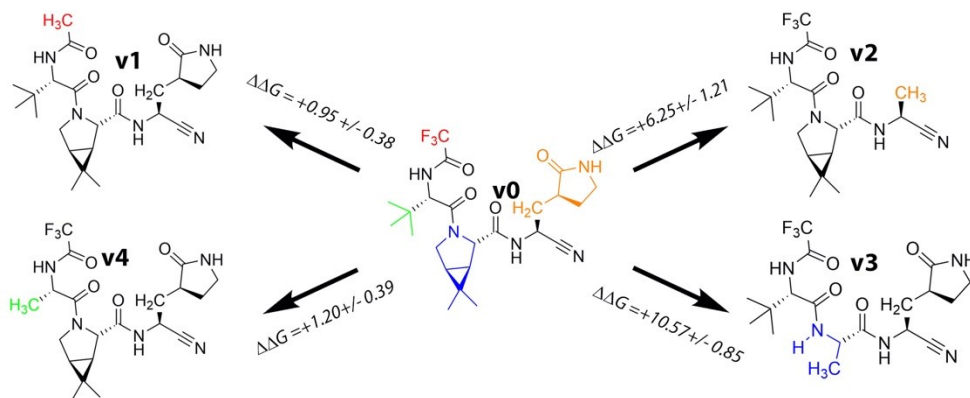


Figura R 16. Cambios de energías libres de unión asociados con la transformación de diferentes grupos Pi del inhibidor PF-07321332 en grupos metilo. P1 en naranja, P2 en azul, P3 en verde y P4 en rojo

Observaciones finales

En esta tesis doctoral se estudiaron dos enzimas cisteína proteasas: la 3CL^{pro} del SARS-CoV-2 y la caspasa-1 humana. Ambas enzimas catalizan la misma actividad proteolítica por medio de una díada catalítica His-Cys. La proteólisis implica dos pasos: En el primero, la acilación, se rompe un enlace peptídico en el sustrato. Como consecuencia, se elimina el fragmento C-terminal del sustrato, mientras que el fragmento N-terminal se une covalentemente a la enzima. En el segundo paso, la de-acilación, se libera este fragmento N-terminal. Durante el proceso de acilación, el átomo S_γ de la Cys actúa como nucleófilo, este átomo se une al carbono electrófilo del enlace peptídico que debe romperse mientras el grupo amino es protonado. Durante la de-acilación, una molécula de agua rompe el enlace azufre-carbono

formado entre la enzima y el sustrato, liberando el fragmento N-terminal del sustrato. A pesar de estas similitudes entre ambas enzimas, los resultados obtenidos en esta tesis doctoral muestran importantes diferencias mecanísticas entre ellas.

En la enzima 3CL^{pro} del SARS-CoV-2, la cisteína catalítica se coloca cerca de la histidina catalítica. Este posicionamiento facilita la formación de una estructura de par iónico como resultado de una transferencia protónica directa desde la cisteína catalítica a la histidina. El coste energético asociado a este proceso de separación de cargas se compensa con la mayor reactividad de la díada para los pasos siguientes: el ataque nucleofílico del átomo de azufre de la cisteína al átomo de carbono electrófilo del sustrato y la protonación del grupo amino. Por otro lado, la organización del sitio activo en la caspasa-1 es fundamentalmente diferente en términos de la disposición de los dos miembros de la díada catalítica. No sólo la distancia entre la cisteína y la histidina es mayor que la distancia encontrada entre estos dos residuos en la 3CL^{pro} del SARS-CoV-2, sino que adicionalmente el sustrato se coloca entre estos dos residuos, impidiendo la transferencia directa de protón entre ellos (ver **Figura R 17**). Por esa razón, la propuesta mecanística encontrada para la proteólisis en la caspasa-1 es completamente diferente a la propuesta para la 3CL^{pro} del SARS-CoV-2. El proceso de acilación en la caspasa-1 implica una transferencia directa de protón desde la cisteína catalítica al sustrato seguida por el ataque nucleofílico del átomo de azufre sobre el carbono electrófilo del sustrato. La histidina catalítica en este caso actúa simplemente como un factor estabilizador para la protonación del grupo saliente. Sin embargo, estas diferencias no parecen afectar la barrera de energía libre, ya que se obtienen valores de energía libre de activación similares para el paso de acilación en ambas enzimas.

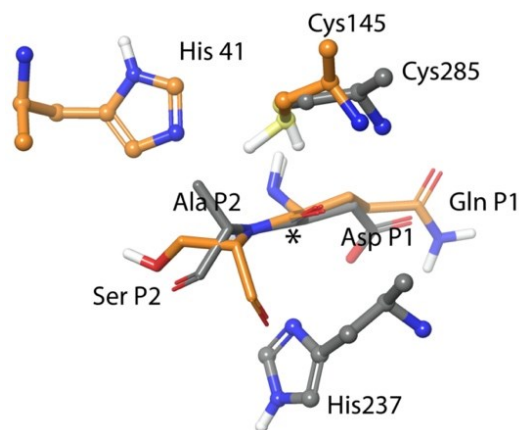


Figura R 17. Comparación de la díada catalítica en la caspasa-1 en gris y en la 3CL^{pro} del SARS-CoV-2 en naranja. El sustrato está representado como bastones y los residuos de la enzima como barras y esferas, el * está en el carbono electrófilo en ambos sustratos. Ambas estructuras se extrajeron de simulaciones de dinámica molecular. Los códigos PDB utilizados como puntos de partida fueron el 6F6R para la caspasa-1 y el 6Y2F para la 3CL^{pro} del SARS-CoV-2.

La estabilidad del intermedio formado después de la acilación es otra diferencia fundamental entre ambas enzimas. En el caso de la 3CL^{pro} del SARS-CoV-2, el grupo saliente interactúa favorablemente con la histidina catalítica, ayudando al posicionamiento de una molécula de agua para realizar el proceso de de-acilación. Como se ve en la **Figura R 18**, en el caso de la caspasa-1, cuando se rompe el enlace peptídico, la distancia entre el grupo amino del grupo saliente C-terminal y la histidina catalítica aumenta, lo que conduce a un complejo intermedio menos estable. Esta diferencia puede determinar la participación o no del grupo saliente en el paso de de-acilación posterior.

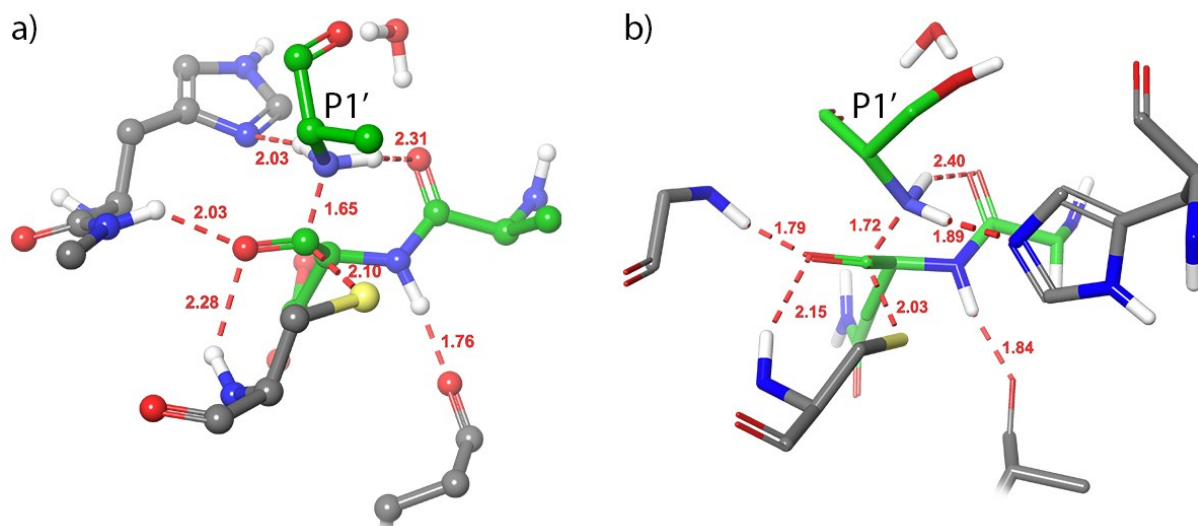


Figura R 18. Configuraciones de los intermedios formados después de las reacciones de acilación. a) caspasa-1, b) 3CL^{pro} de SARS-CoV-2

Otra diferencia fundamental se encuentra en la especificidad de ambas enzimas. La caspasa-1 rompe un enlace peptídico formado entre un aspartato y una alanina, mientras que la 3CL^{pro} del SARS-CoV-2 rompe el enlace peptídico entre una glutamina y una serina. Las diferencias observadas en las interacciones de ambas enzimas con sus sustratos se reflejan obviamente en el diseño de sus respectivos inhibidores. Recordemos aquí la persistente presencia de un anillo γ -lactámico en la posición P1 en los inhibidores de la 3CL^{pro}, aprovechando las interacciones que se establecen entre esta enzima y la glutamina en la posición P1 de su sustrato. Tanto las interacciones no covalentes establecidas entre la enzima y su inhibidor putativo como los detalles mecanísticos deben considerarse en el diseño de inhibidores específicos para cada una de estas cisteínas proteasas. Esperamos que los hallazgos de esta tesis doctoral ayuden a futuros avances en esta área.

Finalmente, vale la pena agregar un comentario sobre la confiabilidad de los métodos de simulación computacional empleados en estos estudios. Durante la redacción de esta tesis doctoral, Pfizer publicó nuevos resultados sobre el inhibidor PF-07321332, también conocido como PAXLOVIDTM.¹¹⁴ Cuando Pfizer anunció que este compuesto estaba en ensayo clínico, en marzo de 2021, decidimos incluirlo en nuestro estudio sobre inhibidores de la enzima 3CL^{pro}. Los resultados de nuestro análisis computacional se publicaron en julio de 2021 (ver apéndice 5). En nuestro estudio computacional, a partir de la estructura de rayos X de un inhibidor diferente (11a, código PDB 6XHM), modelamos el complejo no covalente y luego usamos metodologías QM/MM para estudiar la formación del complejo covalente. La estructura encontrada para este complejo se presenta en la **Figura R 19**, junto con la estructura de rayos X (con código PDB 7VH8) publicada varios meses después de nuestro estudio. La concordancia entre las estructuras predichas y las experimentales es notable, mostrando la madurez alcanzada por las técnicas computacionales en el estudio de los procesos bioquímicos y su capacidad para proporcionar una imagen precisa de los procesos bajo análisis.

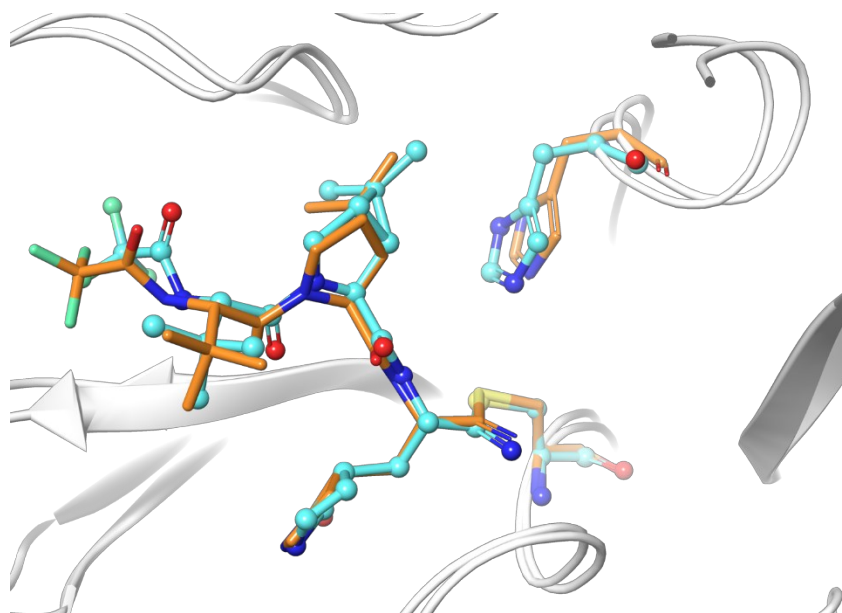


Figura R 19. Estructura del producto determinada mediante el método de la cuerda para la reacción del inhibidor PF-07332132 con la 3CL^{pro} en naranja, y la estructura cristalográfica publicada con código PDB 7VH8 en cyan.¹¹⁵

Bibliography

- (1) The top 10 causes of death <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death> (accessed Jun 24, 2021).
- (2) Prince, M.; Wimo, A.; Guerchet, M.; Gemma-Claire, A.; Wu, Y.-T.; Prina, M. *World Alzheimer Report 2015: The Global Impact of Dementia - An Analysis of Prevalence, Incidence, Cost And Trends*; 2015. <https://doi.org/10.1111/j.0963-7214.2004.00293.x>.
- (3) Wimo, A.; Winblad, B. Health Economical Aspects of Alzheimer Disease and Its Treatment. *Psychogeriatrics* **2001**, *1* (3), 189–193. <https://doi.org/10.1111/j.1479-8301.2001.tb00047.x>.
- (4) Cummings, J. L. Alzheimer's Disease. *Engl. J. Med.* **2004**, No. 351, 56–67.
- (5) Scully, J. L. What Is a Disease? *EMBO Rep.* **2004**, *5* (7), 650–653. <https://doi.org/10.1038/sj.embor.7400195>.
- (6) Storandt M. Towards a Multifactorial Model of Alzheimer Disease. *Neurobiol. Aging* **2012**, *33* (10), 2262–2271. <https://doi.org/10.1016/j.neurobiolaging.2011.11.029>.
- (7) Hardy, J. . A.; Higgins, G. . A. Alzheimer's Disease: The Amyloid Cascade Hypothesis. *Science (80-.)*. **1992**, *256* (5054), 184–185. <https://doi.org/10.1126/science.1566067>.
- (8) Krstic, D.; Knuesel, I. Deciphering the Mechanism Underlying Late-Onset Alzheimer Disease. *Nat. Rev. Neurol.* **2012**, *9* (1), 25–34. <https://doi.org/10.1038/nrneuro.2012.236>.
- (9) Bramblett, G. T.; Goedert, M.; Jakes, R.; Merrick, S. E.; Trojanowski, J. Q.; Lee, V. M. Y. Abnormal Tau Phosphorylation at Ser396 in Alzheimer's Disease Recapitulates Development and Contributes to Reduced Microtubule Binding. *Neuron* **1993**, *10* (6), 1089–1099. [https://doi.org/10.1016/0896-6273\(93\)90057-X](https://doi.org/10.1016/0896-6273(93)90057-X).
- (10) Goedert, M.; Jakes, R.; Spillantini, M. G.; Hasegawa, M.; Smith, M. J.; Crowther, R. A. Assembly of Microtubule-Associated Protein Tau into Alzheimer-like Filaments Induced by Sulphated Glycosaminoglycans. *Nature* **1996**, *383* (6600), 550–553. <https://doi.org/10.1038/383550a0>.
- (11) Kawas, C. H. Early Alzheimer's Disease. *N. Engl. J. Med.* **2003**, *349* (11), 1056–1063. <https://doi.org/10.1056/NEJMcp022295>.
- (12) Hardy, J. The Amyloid Hypothesis of Alzheimer's Disease: Progress and Problems on the Road to Therapeutics. *Science (80-.)*. **2002**, *297* (5580), 353–356. <https://doi.org/10.1126/science.1072994>.
- (13) Stuchbury, G.; Münch, G. Alzheimer's Associated Inflammation, Potential Drug Targets and Future Therapies. *J. Neural Transm.* **2005**, *112* (3), 429–453. <https://doi.org/10.1007/s00702-004-0188-x>.
- (14) Zhang, B.; Gaiteri, C.; Bodea, L. G.; Wang, Z.; McElwee, J.; Podtelezhnikov, A. A.; Zhang, C.; Xie, T.; Tran, L.; Dobrin, R.; Fluder, E.; Clurman, B.; Melquist, S.; Narayanan, M.; Suver, C.; Shah, H.; Mahajan, M.; Gillis, T.; Mysore, J.; MacDonald, M. E.; Lamb, J. R.; Bennett, D. A.; Molony, C.; Stone, D. J.; Gudnason, V.; Myers, A. J.; Schadt, E. E.; Neumann, H.; Zhu, J.; Emilsson, V. Integrated Systems Approach Identifies Genetic Nodes and Networks in Late-Onset Alzheimer's Disease. *Cell* **2013**, *153* (3), 707–720. <https://doi.org/10.1016/j.cell.2013.03.030>.
- (15) Currais, A.; Quehenberger, O.; M Armando, A.; Daugherty, D.; Maher, P.; Schubert, D. Amyloid Proteotoxicity Initiates an Inflammatory Response Blocked by Cannabinoids. *npj Aging Mech. Dis.* **2016**, *2* (February), 16012. <https://doi.org/10.1038/npjamd.2016.12>.

- (16) Netea, M. G.; Balkwill, F.; Chonchol, M.; Cominelli, F.; Donath, M. Y.; Giamarellos-Bourboulis, E. J.; Golenbock, D.; Gresnigt, M. S.; Heneka, M. T.; Hoffman, H. M.; Hotchkiss, R.; Joosten, L. A. B. B.; Kastner, D. L.; Korte, M.; Latz, E.; Libby, P.; Mandrup-Poulsen, T.; Mantovani, A.; Mills, K. H. G. G.; Nowak, K. L.; O'Neill, L. A.; Pickkers, P.; van der Poll, T.; Ridker, P. M.; Schalkwijk, J.; Schwartz, D. A.; Siegmund, B.; Steer, C. J.; Tilg, H.; van der Meer, J. W. M. M.; van de Veerdonk, F. L.; Dinarello, C. A. A Guiding Map for Inflammation. *Nat. Immunol.* **2017**, *18* (8), 826–831. <https://doi.org/10.1038/ni.3790>.
- (17) Rose, N. R. Autoimmune Diseases. In *International Encyclopedia of Public Health*; Elsevier, 2017; pp 192–195. <https://doi.org/10.1016/B978-0-12-803678-5.00029-1>.
- (18) Bamberger, M. E.; Harris, M. E.; McDonald, D. R.; Husemann, J.; Landreth, G. E. A Cell Surface Receptor Complex for Fibrillar β -Amyloid Mediates Microglial Activation. *J. Neurosci.* **2003**, *23* (7), 2665–2674. <https://doi.org/10.1523/JNEUROSCI.23-07-02665.2003>.
- (19) Halle, A.; Hornung, V.; Petzold, G. C.; Stewart, C. R.; Monks, B. G.; Reinheckel, T.; Fitzgerald, K. A.; Latz, E.; Moore, K. J.; Golenbock, D. T. The NALP3 Inflammasome Is Involved in the Innate Immune Response to Amyloid- β . *Nat. Immunol.* **2008**, *9* (8), 857–865. <https://doi.org/10.1038/ni.1636>.
- (20) de Calignon, A.; Fox, L. M.; Pitstick, R.; Carlson, G. A.; Bacskai, B. J.; Spires-Jones, T. L.; Hyman, B. T. Caspase Activation Precedes and Leads to Tangles. *Nature* **2010**, *464* (7292), 1201–1204. <https://doi.org/10.1038/nature08890>.
- (21) Heneka, M. T.; Kummer, M. P.; Stutz, A.; Delekate, A.; Schwartz, S.; Vieira-Saecker, A.; Griep, A.; Axt, D.; Remus, A.; Tzeng, T.-C.; Gelpi, E.; Halle, A.; Korte, M.; Latz, E.; Golenbock, D. T. NLRP3 Is Activated in Alzheimer's Disease and Contributes to Pathology in APP/PS1 Mice. *Nature* **2012**, *493* (7434), 674–678. <https://doi.org/10.1038/nature11729>.
- (22) Duewell, P.; Kono, H.; Rayner, K. J.; Sirois, C. M.; Vladimer, G.; Bauernfeind, F. G.; Abela, G. S.; Franchi, L.; Nuñez, G.; Schnurr, M.; Espevik, T.; Lien, E.; Fitzgerald, K. A.; Rock, K. L.; Moore, K. J.; Wright, S. D.; Hornung, V.; Latz, E. NLRP3 Inflammasomes Are Required for Atherogenesis and Activated by Cholesterol Crystals. *Nature* **2010**, *464* (7293), 1357–1361. <https://doi.org/10.1038/nature08938>.
- (23) Mathews, R. J.; Robinson, J. I.; Battellino, M.; Wong, C.; Taylor, J. C.; Eyre, S.; Churchman, S. M.; Wilson, A. G.; Isaacs, J. D.; Hyrich, K.; Barton, A.; Plant, D.; Savic, S.; Cook, G. P.; Sarzi-Puttini, P.; Emery, P.; Barrett, J. H.; Morgan, A. W.; McDermott, M. F. Evidence of NLRP3-Inflammasome Activation in Rheumatoid Arthritis (RA); Genetic Variants within the NLRP3-Inflammasome Complex in Relation to Susceptibility to RA and Response to Anti-TNF Treatment. *Ann. Rheum. Dis.* **2014**, *73* (6), 1202 LP – 1210. <https://doi.org/10.1136/annrheumdis-2013-203276>.
- (24) Brough, D.; Tyrrell, P. J.; Allan, S. M. Regulation of Interleukin-1 in Acute Brain Injury. *Trends Pharmacol. Sci.* **2011**, *32* (10), 617–622. <https://doi.org/10.1016/j.tips.2011.06.002>.
- (25) Ozaki, E.; Campbell, M.; Doyle, S. L. Targeting the NLRP3 Inflammasome in Chronic Inflammatory Diseases: Current Perspectives. *J. Inflamm. Res.* **2015**, *8*, 15–27. <https://doi.org/10.2147/JIR.S51250>.
- (26) Chen, L.; Liu, W.; Zhang, Q.; Xu, K.; Ye, G.; Wu, W.; Sun, Z.; Liu, F.; Wu, K.; Zhong, B.; Mei, Y.; Zhang, W.; Chen, Y.; Li, Y.; Shi, M.; Lan, K.; Liu, Y. RNA Based MNGS Approach Identifies a Novel Human Coronavirus from Two Individual Pneumonia Cases in 2019 Wuhan Outbreak. *Emerg. Microbes Infect.* **2020**, *9* (1), 313–319. <https://doi.org/10.1080/22221751.2020.1725399>.
- (27) Huang, Y.; Yang, C.; Xu, X.; Xu, W.; Liu, S. Structural and Functional Properties of SARS-CoV-2 Spike Protein: Potential Antivirus Drug Development for COVID-19. *Acta Pharmacol. Sin.* **2020**,

- 41 (9), 1141–1149. <https://doi.org/10.1038/s41401-020-0485-4>.
- (28) Gorbalenya, A. E.; Snijder, E. J.; Ziebuhr, J. Virus-Encoded Proteinases and Proteolytic Processing in the Nidovirales. *J. Gen. Virol.* **2000**, *81* (4), 853–879. <https://doi.org/10.1099/0022-1317-81-4-853>.
- (29) de Wit, E.; van Doremalen, N.; Falzarano, D.; Munster, V. J. SARS and MERS: Recent Insights into Emerging Coronaviruses. *Nat. Rev. Microbiol.* **2016**, *14* (8), 523–534. <https://doi.org/10.1038/nrmicro.2016.81>.
- (30) Fehr, A. R.; Perlman, S. *Coronaviruses: An Overview of Their Replication and Pathogenesis*; Maier, H. J., Bickerton, E., Britton, P., Eds.; Methods in Molecular Biology; Springer New York: New York, NY, 2015; Vol. 1282. <https://doi.org/10.1007/978-1-4939-2438-7>.
- (31) Anand, K. Coronavirus Main Proteinase (3CLpro) Structure: Basis for Design of Anti-SARS Drugs. *Science* (80-.). **2003**, *300* (5626), 1763–1767. <https://doi.org/10.1126/science.1085658>.
- (32) 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.; You, T.; Liu, X.; Yang, X.; Bai, F.; Liu, H.; Liu, 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>.
- (33) Shitrit, A.; Zaidman, D.; Kalid, O.; Bloch, I.; Doron, D.; Yarnizky, T.; Buch, I.; Segev, I.; Ben-Zeev, E.; Segev, E.; Kobilier, O. Conserved Interactions Required for Inhibition of the Main Protease of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). *Sci. Rep.* **2020**, *10* (1), 20808. <https://doi.org/10.1038/s41598-020-77794-5>.
- (34) *Handbook of Proteolytic Enzymes*; Rawlings, N. D., Salvesen, G. S., Eds.; Elsevier, 2013.
- (35) Grant, K.; Kassai, M. Major Advances in the Hydrolysis of Peptides and Proteins by Metal Ions and Complexes. *Curr. Org. Chem.* **2006**, *10* (9), 1035–1049. <https://doi.org/10.2174/138527206777435535>.
- (36) Seelmeier, S.; Schmidt, H.; Turk, V.; von der Helm, K. Human Immunodeficiency Virus Has an Aspartic-Type Protease That Can Be Inhibited by Pepstatin A. *Proc. Natl. Acad. Sci.* **1988**, *85* (18), 6612–6616. <https://doi.org/10.1073/pnas.85.18.6612>.
- (37) Cushman, D. W.; Cheung, H. S.; Sabo, E. F.; Ondetti, M. A. Design of Potent Competitive Inhibitors of Angiotensin-Converting Enzyme. Carboxyalkanoyl and Mercaptoalkanoyl Amino Acids. *Biochemistry* **1977**, *16* (25), 5484–5491. <https://doi.org/10.1021/bi00644a014>.
- (38) Vassar, R.; Kovacs, D. M.; Yan, R.; Wong, P. C. The -Secretase Enzyme BACE in Health and Alzheimer’s Disease: Regulation, Cell Biology, Function, and Therapeutic Potential. *J. Neurosci.* **2009**, *29* (41), 12787–12794. <https://doi.org/10.1523/JNEUROSCI.3657-09.2009>.
- (39) Riemann, D.; Kehlen, A.; Langner, J. CD13—Not Just a Marker in Leukemia Typing. *Immunol. Today* **1999**, *20* (2), 83–88. [https://doi.org/10.1016/S0167-5699\(98\)01398-X](https://doi.org/10.1016/S0167-5699(98)01398-X).
- (40) Lew, R. A.; Mustafa, T.; Ye, S.; McDowall, S. G.; Chai, S. Y.; Albiston, A. L. Angiotensin AT4 Ligands Are Potent, Competitive Inhibitors of Insulin Regulated Aminopeptidase (IRAP). *J. Neurochem.* **2004**, *86* (2), 344–350. <https://doi.org/10.1046/j.1471-4159.2003.01852.x>.
- (41) Yamamoto, N.; Nakayama, J.; Yamakawa-Kobayashi, K.; Hamaguchi, H.; Miyazaki, R.; Arinami, T. Identification of 33 Polymorphisms in the Adipocyte-Derived Leucine Aminopeptidase (ALAP) Gene and Possible Association with Hypertension. *Hum. Mutat.* **2002**, *19* (3), 251–257. <https://doi.org/10.1002/humu.10047>.

- (42) Bhutani, N.; Venkatraman, P.; Goldberg, A. L. Puromycin-Sensitive Aminopeptidase Is the Major Peptidase Responsible for Digesting Polyglutamine Sequences Released by Proteasomes during Protein Degradation. *EMBO J.* **2007**, *26* (5), 1385–1396. <https://doi.org/10.1038/sj.emboj.7601592>.
- (43) Hamai, K.; Ikeda, R.; Sumi, H.; Mihara, H. Carboxypeptidase Activity in Human Urine from Healthy Subjects and Renal Disease Patients. *Clin. Chim. Acta* **1990**, *188* (3), 233–241. [https://doi.org/10.1016/0009-8981\(90\)90205-7](https://doi.org/10.1016/0009-8981(90)90205-7).
- (44) Schechter, I.; Berger, A. On the Size of the Active Site in Proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **1967**, *27* (2), 157–162. [https://doi.org/10.1016/S0006-291X\(67\)80055-X](https://doi.org/10.1016/S0006-291X(67)80055-X).
- (45) 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](https://doi.org/10.1016/S0167-7306(09)60016-8).
- (46) Howard, A. D.; Kostura, M. J.; Thornberry, N.; Ding, G. J.; Limjuco, G.; Weidner, J.; Salley, J. P.; Hogquist, K. A.; Chaplin, D. D.; Mumford, R. a. IL-1-Converting Enzyme Requires Aspartic Acid Residues for Processing of the IL-1 Beta Precursor at Two Distinct Sites and Does Not Cleave 31-KDa IL-1 Alpha. *J. Immunol.* **1991**, *147* (9), 2964–2969.
- (47) Wilson, K. P.; Black, J.-A. F.; Thomson, J. A.; Kim, E. E.; Griffith, J. P.; Navia, M. A.; Murcko, M. A.; Chambers, S. P.; Aldape, R. A.; Raybuck, S. A.; Livingston, D. J. Structure and Mechanism of Interleukin-L β Converting Enzyme. *Nature* **1994**, *370* (6487), 270–275. <https://doi.org/10.1038/370270a0>.
- (48) Pop, C.; Salvesen, G. S. Human Caspases: Activation, Specificity, and Regulation. *J. Biol. Chem.* **2009**, *284* (33), 21777–21781. <https://doi.org/10.1074/jbc.R800084200>.
- (49) Denault, J.-B.; Békés, M.; Scott, F. L.; Sexton, K. M. B.; Bogyo, M.; Salvesen, G. S. Engineered Hybrid Dimers: Tracking the Activation Pathway of Caspase-7. *Mol. Cell* **2006**, *23* (4), 523–533. <https://doi.org/10.1016/j.molcel.2006.06.020>.
- (50) Datta, D.; Scheer, J. M.; Romanowski, M. J.; Wells, J. A. An Allosteric Circuit in Caspase-1. *J. Mol. Biol.* **2008**, *381* (5), 1157–1167. <https://doi.org/10.1016/j.jmb.2008.06.040>.
- (51) Roskoski, R. Michaelis-Menten Kinetics. In *Reference Module in Biomedical Sciences*; Elsevier, 2015; pp 1–9. <https://doi.org/10.1016/B978-0-12-801238-3.05143-6>.
- (52) Miscione, G. Pietro; Calvaresi, M.; Bottoni, A. Computational Evidence for the Catalytic Mechanism of Caspase-7. A DFT Investigation. *J. Phys. Chem. B* **2010**, *114* (13), 4637–4645. <https://doi.org/10.1021/jp908991z>.
- (53) Ramos-Guzmán, C. A.; Zinovjev, K.; Tuñón, I. Modeling Caspase-1 Inhibition: Implications for Catalytic Mechanism and Drug Design. *Eur. J. Med. Chem.* **2019**, *169*, 159–167. <https://doi.org/10.1016/j.ejmech.2019.02.064>.
- (54) Galatsis, P.; Caprathe, B.; Downing, D.; Gilmore, J.; Harter, W.; Hays, S.; Kostlan, C.; Linn, K.; Lunney, E.; Para, K.; Thomas, A.; Warmus, J.; Allen, H.; Brady, K.; Talanian, R.; Walker, N. Inhibition of Interleukin-1 β Converting Enzyme (ICE or Caspase 1) by Aspartyl Acyloxyalkyl Ketones and Aspartyl Amidooxyalkyl Ketones. *Bioorg. Med. Chem. Lett.* **2010**, *20* (17), 5089–5094. <https://doi.org/10.1016/j.bmcl.2010.07.031>.
- (55) Parui, A. L.; Bose, K. Caspases: Regulatory Mechanisms and Their Implications in Pathogenesis and Therapeutics. In *Pathophysiological Aspects of Proteases*; Springer Singapore: Singapore, 2017; pp 423–488. https://doi.org/10.1007/978-981-10-6141-7_18.
- (56) Cade, C. E.; Clark, A. C. Caspases – Key Players in Apoptosis. In *Proteases in Apoptosis*:

- Pathways, Protocols and Translational Advances*; Bose, K., Ed.; Springer International Publishing: Cham, 2015; pp 31–51. https://doi.org/10.1007/978-3-319-19497-4_2.
- (57) Sulpizi, M.; Laio, A.; VandeVondele, J.; Cattaneo, A.; Rothlisberger, U.; Carloni, P. Reaction Mechanism of Caspases: Insights from QM/MM Car-Parrinello Simulations. *Proteins Struct. Funct. Genet.* **2003**, *52* (2), 212–224. <https://doi.org/10.1002/prot.10275>.
- (58) 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>.
- (59) Chen, Y. W.; Yiu, C.-P. B.; Wong, K.-Y. Prediction of the SARS-CoV-2 (2019-NCoV) 3C-like Protease (3CLpro) Structure: Virtual Screening Reveals Velpatasvir, Ledipasvir, and Other Drug Repurposing Candidates. *F1000Research* **2020**, *9*, 129. <https://doi.org/10.12688/f1000research.22457.1>.
- (60) Tibbles, K. W.; Brierley, I.; Cavanagh, D.; Brown, T. D. Characterization in Vitro of an Autocatalytic Processing Activity Associated with the Predicted 3C-like Proteinase Domain of the Coronavirus Avian Infectious Bronchitis Virus. *J. Virol.* **1996**, *70* (3), 1923–1930. <https://doi.org/10.1128/jvi.70.3.1923-1930.1996>.
- (61) 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>.
- (62) POLGAR, L. Deuterium Isotope Effects on Papain Acylation. Evidence for Lack of General Base Catalysis and for Enzyme-Leaving-Group Interaction. *Eur. J. Biochem.* **1979**, *98* (2), 369–374. <https://doi.org/10.1111/j.1432-1033.1979.tb13196.x>.
- (63) Taranto, A. G.; Carvalho, P.; Avery, M. A. QM/QM Studies for Michael Reaction in Coronavirus Main Protease (3CLPro). *J. Mol. Graph. Model.* **2008**, *27* (3), 275–285. <https://doi.org/10.1016/j.jmgm.2008.05.002>.
- (64) Wu, F.; Zhao, S.; Yu, B.; Chen, Y.-M.; Wang, W.; Song, Z.-G.; Hu, Y.; Tao, Z.-W.; Tian, J.-H.; Pei, Y.-Y.; Yuan, M.-L.; Zhang, Y.-L.; Dai, F.-H.; Liu, Y.; Wang, Q.-M.; Zheng, J.-J.; Xu, L.; Holmes, E. C.; Zhang, Y.-Z. A New Coronavirus Associated with Human Respiratory Disease in China. *Nature* **2020**, *579* (7798), 265–269. <https://doi.org/10.1038/s41586-020-2008-3>.
- (65) Bradley, P. B. Characteristics of Drug Action. In *Introduction to Neuropharmacology*; Elsevier, 1989; pp 3–16. <https://doi.org/10.1016/B978-0-7236-1271-1.50006-7>.
- (66) Dougall, I. G.; Unitt, J. Evaluation of the Biological Activity of Compounds. In *The Practice of Medicinal Chemistry*; Elsevier, 2015; pp 15–43. <https://doi.org/10.1016/B978-0-12-417205-0.00002-X>.
- (67) Le, G.; Abbenante, G. Inhibitors of TACE and Caspase-1 As Anti-Inflammatory Drugs. *Curr. Med. Chem.* **2005**, *12* (25), 2963–2977. <https://doi.org/10.2174/092986705774462851>.
- (68) Baillie, T. A. Targeted Covalent Inhibitors for Drug Design. *Angew. Chemie Int. Ed.* **2016**, *55* (43), 13408–13421. <https://doi.org/10.1002/anie.201601091>.
- (69) Singh, J.; Petter, R. C.; Baillie, T. A.; Whitty, A. The Resurgence of Covalent Drugs. *Nat. Rev. Drug Discov.* **2011**, *10* (4), 307–317. <https://doi.org/10.1038/nrd3410>.
- (70) 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), 1331–1335. <https://doi.org/10.1126/science.abb4489>.
- (71) Jorgensen, W. L. The Many Roles of Computation in Drug Discovery. *Science* (80-.). **2004**, *303* (5665), 1813–1818. <https://doi.org/10.1126/science.1096361>.
- (72) Schramm, V. L. Enzymatic Transition States and Transition State Analog Design. *Annu. Rev. Biochem.* **1998**, *67* (1), 693–720. <https://doi.org/10.1146/annurev.biochem.67.1.693>.
- (73) Marlier, J. F. Multiple Isotope Effects on the Acyl Group Transfer Reactions of Amides and Esters. *Acc. Chem. Res.* **2001**, *34* (4), 283–290. <https://doi.org/10.1021/ar000054d>.
- (74) Jorgensen, W. L. Efficient Drug Lead Discovery and Optimization. *Acc. Chem. Res.* **2009**, *42* (6), 724–733. <https://doi.org/10.1021/ar800236t>.
- (75) Shoichet, B. K.; McGovern, S. L.; Wei, B.; Irwin, J. J. Lead Discovery Using Molecular Docking. *Curr. Opin. Chem. Biol.* **2002**, *6* (4), 439–446. [https://doi.org/10.1016/S1367-5931\(02\)00339-3](https://doi.org/10.1016/S1367-5931(02)00339-3).
- (76) DeWitte, R. S.; Shakhnovich, E. I. SMOG: De Novo Design Method Based on Simple, Fast, and Accurate Free Energy Estimates. 1. Methodology and Supporting Evidence. *J. Am. Chem. Soc.* **1996**, *118* (47), 11733–11744. <https://doi.org/10.1021/ja960751u>.
- (77) PAULING, L. Nature of Forces between Large Molecules of Biological Interest*. *Nature* **1948**, *161* (4097), 707–709. <https://doi.org/10.1038/161707a0>.
- (78) Warshel, A. Energetics of Enzyme Catalysis. *Proc. Natl. Acad. Sci.* **1978**, *75* (11), 5250–5254. <https://doi.org/10.1073/pnas.75.11.5250>.
- (79) Herschlag, D. The Role of Induced Fit and Conformational Changes of Enzymes in Specificity and Catalysis. *Bioorg. Chem.* **1988**, *16* (1), 62–96. [https://doi.org/10.1016/0045-2068\(88\)90038-7](https://doi.org/10.1016/0045-2068(88)90038-7).
- (80) McCammon, J. A.; Gelin, B. R.; Karplus, M. Dynamics of Folded Proteins. *Nature* **1977**, *267* (5612), 585–590. <https://doi.org/10.1038/267585a0>.
- (81) Durrant, J. D.; McCammon, J. A. Molecular Dynamics Simulations and Drug Discovery. *BMC Biol.* **2011**, *9*, 71. <https://doi.org/10.1186/1741-7007-9-71>.
- (82) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. . C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* **1977**, *23* (3), 327–341. [https://doi.org/10.1016/0021-9991\(77\)90098-5](https://doi.org/10.1016/0021-9991(77)90098-5).
- (83) Verlet, L. Computer “Experiments” on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.* **1967**, *159* (1), 98–103. <https://doi.org/10.1103/PhysRev.159.98>.
- (84) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102* (18), 3586–3616. <https://doi.org/10.1021/jp973084f>.
- (85) 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>.

- (86) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79* (2), 926. <https://doi.org/10.1063/1.445869>.
- (87) Mulholland, A. J. The QM/MM Approach to Enzymatic Reactions. In *Theoretical and Computational Chemistry*; Elsevier Masson SAS, 2001; Vol. 9, pp 597–653. [https://doi.org/10.1016/S1380-7323\(01\)80015-7](https://doi.org/10.1016/S1380-7323(01)80015-7).
- (88) Warshel, A.; Levitt, M. Theoretical Studies of Enzymic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme. *J. Mol. Biol.* **1976**, *103* (2), 227–249. [https://doi.org/10.1016/0022-2836\(76\)90311-9](https://doi.org/10.1016/0022-2836(76)90311-9).
- (89) Field, M. J.; Bash, P. A.; Karplus, M. A Combined Quantum Mechanical and Molecular Mechanical Potential for Molecular Dynamics Simulations. *J. Comput. Chem.* **1990**, *11* (6), 700–733. <https://doi.org/10.1002/jcc.540110605>.
- (90) Zuckerman, D. M. *Statistical Physics of Biomolecules: An Introduction.*; CRC Press, 2010.
- (91) van Gunsteren, W. F.; Daura, X.; Mark, A. E. Computation of Free Energy. *Helv. Chim. Acta* **2002**, *85* (10), 3113–3129. [https://doi.org/10.1002/1522-2675\(200210\)85:10<3113::AID-HLCA3113>3.0.CO;2-0](https://doi.org/10.1002/1522-2675(200210)85:10<3113::AID-HLCA3113>3.0.CO;2-0).
- (92) Straatsma, T. P.; McCammon, J. A. Computational Alchemy. *Annu. Rev. Phys. Chem.* **1992**, *43* (1), 407–435. <https://doi.org/10.1146/annurev.pc.43.100192.002203>.
- (93) Massova, I.; Kollman, P. a. Combined Molecularmechanical and Continuum Solvent Approach (MM-PBSA/GBSA) to Predict Ligand Binding. *Perspect. Drug Discov. Des.* **2000**, *18* (1), 113–135. <https://doi.org/10.1023/A:1008763014207>.
- (94) Srinivasan, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A. Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate - DNA Helices. *J. Am. Chem. Soc.* **1998**, *120* (37), 9401–9409. <https://doi.org/10.1021/ja981844+>.
- (95) Pearlman, D. A. Evaluating the Molecular Mechanics Poisson–Boltzmann Surface Area Free Energy Method Using a Congeneric Series of Ligands to P38 MAP Kinase. *J. Med. Chem.* **2005**, *48* (24), 7796–7807. <https://doi.org/10.1021/jm050306m>.
- (96) Schenter, G. K.; Garrett, B. C.; Truhlar, D. G. Generalized Transition State Theory in Terms of the Potential of Mean Force. *J. Chem. Phys.* **2003**, *119* (12), 5828–5833. <https://doi.org/10.1063/1.1597477>.
- (97) 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>.
- (98) 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](https://doi.org/https://doi.org/10.1016/0021-9991(77)90121-8).
- (99) Zinovjev, K.; Tuñón, I. Exploring Chemical Reactivity of Complex Systems with Path-Based Coordinates: Role of the Distance Metric. *J. Comput. Chem.* **2014**, *35* (23), 1672–1681. <https://doi.org/10.1002/jcc.23673>.
- (100) Sugita, Y.; Kitao, A.; Okamoto, Y. Multidimensional Replica-Exchange Method for Free-Energy Calculations. *J. Chem. Phys.* **2000**, *113* (15), 6042–6051. <https://doi.org/10.1063/1.1308516>.
- (101) 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>.

- (102) Giegel, D. A. ICE Processing and Kinetic Mechanism. *J. Cell. Biochem.* **1997**, *18*, 11–18.
- (103) Ramos-Guzmán, C. A.; Ruiz-Pernía, J. J.; Tuñón, I. Unraveling the SARS-CoV-2 Main Protease Mechanism Using Multiscale Methods. *ACS Catal.* **2020**, *10* (21), 12544–12554. <https://doi.org/10.1021/acscatal.0c03420>.
- (104) Akaji, K.; Konno, H.; Mitsui, H.; Teruya, K.; Shimamoto, Y.; Hattori, Y.; Ozaki, T.; Kusunoki, M.; Sanjoh, A. Structure-Based Design, Synthesis, and Evaluation of Peptide-Mimetic SARS 3CL Protease Inhibitors. *J. Med. Chem.* **2011**, *54* (23), 7962–7973. <https://doi.org/10.1021/jm200870n>.
- (105) Grimsley, G. R.; Scholtz, J. M.; Pace, C. N. A Summary of the Measured PK Values of the Ionizable Groups in Folded Proteins. *Protein Sci* **2009**, *18*, 247–251. <https://doi.org/10.1002/pro.19>.
- (106) Hoffman, R. L.; Kania, R. S.; Brothers, M. A.; Davies, J. F.; Ferre, R. A.; Gajiwala, K. S.; He, M.; Hogan, R. J.; Kozminski, K.; Li, L. Y.; Lockner, J. W.; Lou, J.; Marra, M. T.; Mitchell, L. J.; Murray, B. W.; Nieman, J. A.; Noell, S.; Planken, S. P.; Rowe, T.; Ryan, K.; Smith, G. J.; Solowiej, J. E.; Stepan, C. M.; Taggart, B. Discovery of Ketone-Based Covalent Inhibitors of Coronavirus 3CL Proteases for the Potential Therapeutic Treatment of COVID-19. *J. Med. Chem.* **2020**, *63* (21), 12725–12747. <https://doi.org/10.1021/acs.jmedchem.0c01063>.
- (107) Yang, H.; Xie, W.; Xue, X.; Yang, K.; Ma, J.; Liang, W.; Zhao, Q.; Zhou, Z.; Pei, D.; Ziebuhr, J.; Hilgenfeld, R.; Kwok, Y. Y.; Wong, L.; Gao, G.; Chen, S.; Chen, Z.; Ma, D.; Bartlam, M.; Rao, Z. Design of Wide-Spectrum Inhibitors Targeting Coronavirus Main Proteases. *PLoS Biol.* **2005**, *3* (10). <https://doi.org/10.1371/journal.pbio.0030324>.
- (108) 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>.
- (109) Ramos-Guzmán, C. A.; Ruiz-Pernía, J. J.; Tuñón, I. A Microscopic Description of SARS-CoV-2 Main Protease Inhibition with Michael Acceptors. Strategies for Improving Inhibitor Design. *Chem. Sci.* **2021**, *12* (10), 3489–3496. <https://doi.org/10.1039/D0SC04978F>.
- (110) Ramos-Guzmán, C. A.; Ruiz-Pernía, J. J.; Tuñón, I. Multiscale Simulations of SARS-CoV-2 3CL Protease Inhibition with Aldehyde Derivatives. Role of Protein and Inhibitor Conformational Changes in the Reaction Mechanism. *ACS Catal.* **2021**, *11* (7), 4157–4168. <https://doi.org/10.1021/acscatal.0c05522>.
- (111) Ramos-Guzmán, C. A.; Ruiz-Pernía, J. J.; Tuñón, I. Computational Simulations on the Binding and Reactivity of a Nitrile Inhibitor of the SARS-CoV-2 Main Protease. *Chem. Commun.* **2021**, 2–5. <https://doi.org/10.1039/D1CC03953A>.
- (112) Ramos-Guzmán, C. A.; Ruiz-Pernía, J. J.; Tuñón, I. Inhibition Mechanism of SARS-CoV-2 Main Protease with Ketone-Based Inhibitors Unveiled by Multiscale Simulations: Insights for Improved Designs**. *Angew. Chemie Int. Ed.* **2021**, 2–11. <https://doi.org/10.1002/anie.202110027>.
- (113) Lipinski, C. A. Lead- and Drug-like Compounds: The Rule-of-Five Revolution. *Drug Discov. Today Technol.* **2004**, *1* (4), 337–341. <https://doi.org/10.1016/j.ddtec.2004.11.007>.
- (114) Owen, D. R.; Allerton, C. M. N.; Anderson, A. S.; Aschenbrenner, L.; Avery, M.; Berritt, S.; Boras, B.; Cardin, R. D.; Carlo, A.; Coffman, K. J.; Dantonio, A.; Di, L.; Eng, H.; Ferre, R.; Gajiwala, K. S.; Gibson, S. A.; Greasley, S. E.; Hurst, B. L.; Kadar, E. P.; Kalgutkar, A. S.; Lee, J. C.; Lee, J.; Liu, W.;

Mason, S. W.; Noell, S.; Novak, J. J.; Obach, R. S.; Ogilvie, K.; Patel, N. C.; Pettersson, M.; Rai, D. K.; Reese, M. R.; Sammons, M. F.; Sathish, J. G.; Singh, R. S. P.; Stepan, C. M.; Stewart, A. E.; Tuttle, J. B.; Updyke, L.; Verhoest, P. R.; Wei, L.; Yang, Q.; Zhu, Y. An Oral SARS-CoV-2 M pro Inhibitor Clinical Candidate for the Treatment of COVID-19. *Science* (80-.). **2021**.
<https://doi.org/10.1126/science.abl4784>.

- (115) Zhao, Y.; Fang, C.; Zhang, Q.; Zhang, R.; Zhao, X.; Duan, Y.; Wang, H.; Zhu, Y.; Feng, L.; Zhao, J.; Shao, M.; Yang, X.; Zhang, L.; Peng, C.; Yang, K.; Ma, D.; Rao, Z.; Yang, H. Crystal Structure of SARS-CoV-2 Main Protease in Complex with Protease Inhibitor PF-07321332. *Protein Cell* **2021**.
<https://doi.org/10.1007/s13238-021-00883-2>.

ANNEXES

Appendix 1

Modeling Caspase-1 Inhibition: Implications for Catalytic Mechanism and Drug Design



Research paper

Modeling caspase-1 inhibition: Implications for catalytic mechanism and drug design

Carlos A. Ramos-Guzmán, Kirill Zinovjev, Iñaki Tuñón*

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

ARTICLE INFO

Article history:

Received 9 November 2018

Received in revised form

21 February 2019

Accepted 23 February 2019

Available online 27 February 2019

Keywords:

Caspase-1

MM/GBSA

ICE inhibition

Binding free energy

ABSTRACT

The metabolic product of caspase-1, IL-1 β , is an important mediator in inflammation and pyroptosis cell death process. Alzheimer's disease, septic shock and rheumatoid arthritis are IL-1 β mediated diseases, making the caspase-1 an interesting target of pharmacological value. Many inhibitors have been developed until now, most of them are peptidomimetic with improved potency. In the present study, all-atom molecular dynamics simulations and the MM/GBSA method were employed to reproduce and interpret the results obtained by in vitro experiments for a series of inhibitors. The analysis shows that the tautomeric state of the catalytic His237 impact significantly the performance of the prediction protocol, providing evidence for a His237 tautomeric state different to the proposed in the putative mechanism. Additionally, analysis of inhibitor-enzyme interactions indicates that the differences in the inhibitory potency of the tested ligands can be explained mainly by the interaction of the inhibitors with the S2–S4 protein region. These results provide guidelines for subsequent studies of caspase-1 catalytic reaction mechanism and for the design of novel inhibitors.

© 2019 Elsevier Masson SAS. All rights reserved.

1. Introduction

The enzyme caspase-1, also known as interleukin-1 β converting enzyme (ICE) hydrolyses the biologically inactive form of the proinflammatory cytokine pro-interleukin-1 β (Pro-IL-1 β) into its active form, the IL-1 β . Evidence suggests that inflammatory processes related with IL-1 β play an important role in the early stages of the development of Alzheimer's disease (AD) [1]. Additionally, a relationship between the presence of the metabolic product of caspase-1 (the IL-1 β) and the amyloid- β deposition has been observed, the latter being a characteristic hallmark of the AD. In this way the inhibition of caspase-1 is a promising therapeutic strategy for AD, for the brain injury/diseases in general and for other IL-1 β mediated diseases as septic shock and rheumatoid arthritis [2].

Pro-IL-1 β proteolysis is divided in two steps: the substrate acylation by the enzyme and the hydrolysis of the acyl-enzyme intermediate (see Fig. 1). The formation of the acyl-enzyme intermediate is supported by crystallographic evidence and involves the participation of two residues of the enzyme: Cys285 and His237 [3]. In the putative mechanism [3] the first step consists in the

deprotonation of Cys285 by His237. Once deprotonated, the sulfur atom of Cys285 would perform a nucleophilic attack on the carbonylic carbon of Asp116 of the substrate. Then His237 would protonate the amide nitrogen atom of Ala117 breaking the peptide bond between these two consecutive amino acids of the substrate and leading to the formation of an acyl-enzyme complex. In the second step, a water molecule would perform a concerted nucleophilic attack on the carbon of the C-S bond. The leaving Cys285 would then be protonated by His237 [3–5] generating the 17.5 kDa mature IL-1 β , as depicted in Fig. 1. In this mechanism proposed by Wilson et al. [3] His237 is drawn in a rotameric state different to that found in the X-ray structure and the unprotonated ϵ -nitrogen of the catalytic histidine is in charge of Cys285 ionization, similarly to what has been proposed for other cysteine proteases. However, crystallographic structures show that the distance between the catalytic dyad is too long in the case of caspases, between 6 and 7 Å, and that inhibitors that mimic the substrate are placed in between them [3,6–12]. While a δ -protonated His237 is still reproduced in many publications as the basic mechanism for the caspase family [4,13–15], some computational mechanistic analysis [13,14,16] suggest that the most probable tautomeric state in caspases could be protonated at the ϵ -position. In addition, Wilson et al. [3] found that when the His237 is mutated to glutamine, alanine or lysine the reaction rate goes down, but the activity does not disappear

* Corresponding author.

E-mail address: Ignacio.tunon@uv.es (I. Tuñón).

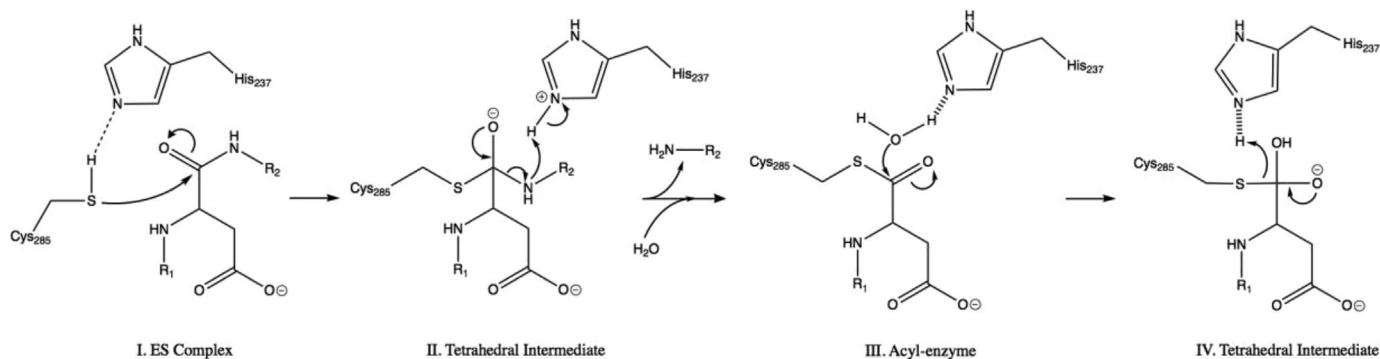


Fig. 1. Putative reaction mechanism for the hydrolysis of the Pro-IL-1 β by caspase-1 [3].

completely. Thus, the role of the catalytic histidine in caspases (and then its protonation and tautomeric states) could be different from other cysteine proteases.

Today there are many inhibitors patented for the enzyme caspase-1. However, until now none of them has been commercialized due to their poor pharmacokinetic profile and their high toxicity [17]. Many of them are substrate peptidomimetics, like the first potent inhibitor reported [18] for caspase-1, AcYVAD(CH₂)₅Ph. Fig. 2 shows the general scaffold of known caspase-1 inhibitors. The core fragment is an aspartate or aspartate like moiety, the P4-P2 fragments are usually peptidomimetic scaffolds of the substrate, while the prime side fragment is typically replaced by leaving groups that facilitate the nucleophilic attack of Cys285, resulting in irreversible inhibition. Despite the bioactivity advantages of the irreversible inhibitors they may exhibit larger toxicity because of their high reactivity: irreversible inhibitors can react easily with another nucleophile present in the body. As a result, their administration as a long term therapy may lead to severe side-effects [19].

Fig. 3 shows AcYVAD(CH₂)₅Ph inside caspase-1 active site, highlighting the three main components of caspase-1 inhibitors: a core fragment (green), a P4-P2 fragment (brown) and a prime side fragment (yellow). The prime side fragment interacts with the hydrophobic cavity formed by Ile176, Pro177, Ile239 and His248; the core fragment is inside the hole formed by the catalytic residues Cys285 and His237 (S1 region) and the carboxylate pocket Arg179, Gly238, Gln283, Ser339, Arg341, and finally the P4-P2 fragment interacts with Trp340, Arg341, His342, Pro343, Val348, Arg383 (S2–S4 region).

In order to evaluate the possibilities of the prime side modifications to develop novel caspase-1 inhibitors, Galatsis et al. [20] developed a SAR campaign using a carbobenzyloxy aspartic acid (Cbz) as core fragment (Fig. 4). As prime side fragments these authors used a series of acyloxyalkyl and amidooxyalkyl ketones, exploring in this way protein interactions available for the prime side fragment and obtaining molecules with improved inhibitory potency against caspase-1. These authors reported the inhibition constant for the full set of inhibitors.

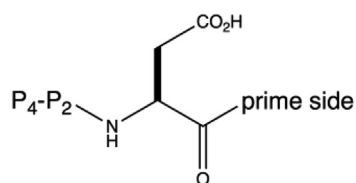


Fig. 2. Scaffold of caspase-1 inhibitors: P4-P2 fragment, core fragment and a prime side fragment.

To reduce the computational cost, a subset of the inhibitors synthesized by Galatsis et al. [20] was selected for the present study (see Table 1). The values of their inhibition constants (K_i values) were related to the binding free energy obtained from molecular simulations. To name the molecules, the same notation as in Ref. [20] was used. To establish a systematic comparison among a congeneric series of compounds the inhibitors presented in that work were divided into three families: aspartyl acyloxyalkyl ketones (4a, 4b, 4i, 4p and 5), aspartyl amidooxyalkyl ketones (6, 7b, 7d, 7g, 7h and 7i) and derivatized peptidomimetics inhibitors (1, 11a, 11b, 12a, 12b, 12c and 13). Compounds with prime sides analogous in at least two of the three families were selected. On the other hand, compounds with prime side present only in one family were rejected for our study. It should be pointed out that molecule 1 corresponds to the potent caspase-1 inhibitor AcYVAD(CH₂)₅Ph. All the selected compounds are reversible inhibitors, with the exception of compound 11b that shows bimodal kinetics corresponding to fast inhibition followed by a slower irreversible inactivation. However, the faster reversible inhibition can be separated and the inhibition constant determined [20]. In this work we have correlated the inhibition constants with the binding free energy of the inhibitor, corresponding to the formation of the non-covalent complex.

In computer-aided drug design studies, the accurate computation of binding free energy for protein-ligand complexes is the holy grail, since this quantity determines the inhibition ranking of a series of candidate inhibitors. However, a rigorous evaluation of the binding free energy is usually computationally expensive because of the sampling required to properly evaluate some of the contributions to the binding process, such as the active site desolvation and conformational changes taking place along the binding process [21–23]. Thus, approximate methods are often used. A first approach to the study of inhibition can be obtained by means of docking studies that sort compounds from less to more active using a scoring function [24,25]. A more successful strategy is the combination of molecular dynamics simulations of the protein-ligand complex in discrete solvents with continuum solvent models to evaluate the solvation free energy of the protein and the ligand. Thus, molecular mechanics simulations can be combined with the generalized-Born surface area method (MM/GBSA) or the Poisson-Boltzmann surface area (MM/PBSA) models [26]. MM/GBSA and MM/PBSA have shown that both provide binding free energies that have good correlation with experimental data [27–30]. In this work, the MM/GBSA method was used to obtain the binding free energies of those molecules presented in Table 1 in the active site of caspase-1. Our simulations were validated by comparison with the experimental inhibition constants and those residues contributing more to differentiate the inhibitors according to the affinity

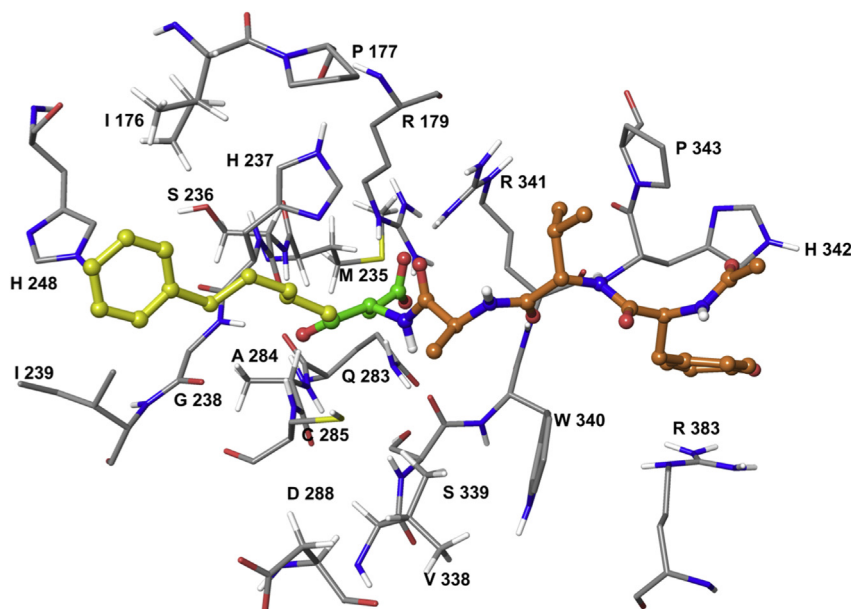


Fig. 3. Representation of AcYVAD(CH₂)₅Ph in the active site of caspase-1. The carbon atoms of the prime side, the core and the P4-P2 fragments are represented in yellow, green and brown, respectively [20]. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

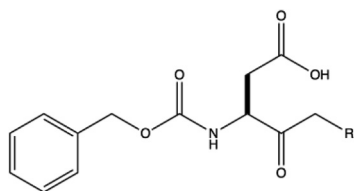


Fig. 4. Cbz-Asp-R (Carbobenzyloxy-aspartic acid-R. R is an acyloxyalkyl or an amidoxyalkyl ketone). The prime side (R), core fragment (Asp) and P4-P2 (Cbz).

between the inhibitor and the protein were identified. This study can be then useful for future design of new compounds with inhibitory capacity for caspase-1.

2. Methods

2.1. Classical molecular dynamic simulations

The inhibitors listed in Table 1 were parametrized using the antechamber program [31] from the AmberTools18 [32] package, based on the general amber force field [33]. The ligand-enzyme complexes were built with Maestro (V. 10.5.014) [34] using the structure with PDB code 1ICE [3] as template. This structure corresponds to the human caspase-1 cocrystallized with the inhibitor Ac-Tyr-Val-Ala-Asp-Aldehyde. The inhibitors selected for this study were aligned to the structure of AcYVAD(CH₂)₅Ph. The crystallographic data for this inhibitor was provided in the supporting information of ref. [20]. To align the ligands the flexible shape-based alignment option available in the ligand alignment tool in Maestro was used [35]. In this, for a given pair of molecules, the best alignment is selected after considering hundreds of possible alignments based on the distribution of similarity triplets of atoms.

Determination of the appropriate protonation states for titratable residues of proteins can be made using different strategies [36–39]. The family of caspase enzymes is a clear example of the importance of the assumed protonation state in the determination

of the reaction mechanism. In this work hydrogen atoms were added using the Protein Preparation Wizard tool from Maestro and the protonation states were calculated with PROPKA3.0 [38]. It is worth to mention that according to PROPKA predictions the catalytic His237 is neutral at pH = 7, with the proton found in the ϵ -position (ϵ -His237) and not in the δ -position (δ -His237). This last is the tautomeric state assume for His237 in the putative mechanism proposed by Wilson et al. [3]. In the present work we used the MM/GBSA methodology as an additional tool to determine the tautomeric state of neutral His237, performing the analysis of the binding free energies for the two possibilities. Additionally, and in spite of the results obtained with PROPKA, it could be possible to consider that in the caspase-inhibitor complex Cys285 and His237 are already ionized. Previous Molecular Dynamics simulations of the protein with this protonation states resulted in an unstable system [16]. In this work we ran a 50 ns simulation of the reference ligand **1** (AcYVAD(CH₂)₅Ph) in the active site with an unprotonated Cys285 and a protonated His237. This simulation showed that the Michaelis complex was not stable. The inhibitor prime and core sides move away from the active site, being in contact with the enzyme only by the P4-P2 fragment (see the corresponding RMSD plot in Fig. S1). We thus discarded this possibility for the rest of the study. Regarding the rotameric state of the catalytic histidine we always used as starting point the X-ray structure, regardless of the assumed tautomeric state. This rotameric state is kept during the simulations (see Fig. S2).

Once the protein-ligands complexes were constructed, the simulation systems were built with the tleap tool from AmberTools18. Each protein-ligand system, described with the ff14SB force field [40], was solvated by a box of TIP3P water molecules [41] with a buffer region of at least 12 Å from any protein or ligand atom to the limits of the simulation box. As all the simulated systems were neutral, no counter ions were needed. Initial geometry relaxations of the systems were carried out using 500 steps of steepest descent method followed by 500 steps of conjugate gradient minimization for a maximum of 1000 minimization steps with a termination threshold of 10^{-4} kcal mol⁻¹ Å⁻¹ for the root mean square of the gradient. After heating the system at 300 K

Table 1
Caspase-1 inhibitors selected for analysis in this work.

Entry	Structure	K _i (μM)	Entry	Structure	K _i (μM)
1		0.011	5		119
4a		8.3	7g		66.5
4b		4.5	7h		27.5
4i		0.5	7i		84.0
4p		0.8	7b		9.20
6		15.0	7d		6.77
11a		0.003	12a		0.007
11b		0.001	12b		0.002
13		0.134	12c		0.001

during 1 ns of NVT simulation, 50 ns of NPT simulation at 300 K and 1 bar were performed. A time step of 2 fs was employed and the bonds involving hydrogens were constrained using SHAKE [42]. Long range electrostatic interactions were treated by the Particle Mesh Ewald method [43,44] with a cutoff for the short range interactions of 10 Å. Temperature and pressure were controlled by Langevin thermostat and Berendsen barostat, respectively. Classical molecular dynamics simulations were performed using the GPU version of PMEMD from Amber17 [45,46]. The RMSD values obtained for the simulations corresponding to ε-His237 and δ-His237 are gathered in Fig. S3 and Fig. S4 for the inhibitors, and in Fig. S5 and Fig. S6 for the protein backbone. These plots show that the simulations are stable for both states.

2.2. MM/GBSA free energy calculations

MM/GBSA [26] is a popular methodology to compute binding free energies in protein-ligand complexes [47–51]. This is an end point sampling method, meaning that only the free energies of bound and unbound states are involved in the calculations. The binding free energy, ΔG_{bind} , is calculated as:

$$\Delta G_{bind} = G_{complex} - G_{protein} - G_{ligand} \quad (1)$$

where every term corresponds to the value for the ensemble average for every point. The free energy is approximated as [26]:

$$G = E_{MM} + G_{solv} - T \cdot S \quad (2)$$

where E_{MM} is the gas phase molecular mechanics energy, G_{solv} describes the polar and non-polar contributions to the solvation free energy, computed via the Generalized Born approximation and using a linear function of the Solvent-Accessible Surface Area. The last term in eq. (2) is the entropic contribution that is often neglected in studies of congeneric series or if only a relative ranking of ligand binding affinities is required. This typically leads to too favorable binding free energies although it is a reasonable approximation when comparing ligands of similar size [52,53], as in the present case.

The MMPBSA.py tool [54] from AmberTools18 was used to perform the MM/GBSA calculations. Once the systems were relaxed after the 50 ns of NPT simulation at 300 K and 1 bar, 50 short trajectories of 500 ps were generated for each one using the same conditions stated above and different starting velocities for each replica. These 50 trajectories were employed to obtain the average values of the quantities appearing in eq. (2), using a total of 2500 frames per ligand.

3. Results

3.1. Binding free energies

In order to validate our simulations, the binding energy of each ligand present in Table 1 and the enzyme was calculated using the MM/GBSA method. These calculations were performed using the structures of the enzyme with the His237 protonated in the δ -position, according to the accepted mechanism, and with the His237 protonated in the ϵ -position. The correlations between the logarithm of the experimental K_i and the calculated binding energies of the 18 selected ligands and the protein in the two tautomeric states are shown in Fig. 5 (the numerical data is provided as Supplementary Material, Table S1).

Fig. 5 shows that the correlation obtained between the experimental results and the calculated binding free energies are considerably better when His237 is assumed to be protonated at the ϵ -position. Until now, computational studies for the development of new inhibitors have not discussed explicitly the problem of the tautomeric state of His237 [55–57]. The results of the present

study confirm that the tautomeric state of His237 play an important role in substrate-enzyme interactions. Knowing the correct tautomeric state of the His237 is of paramount importance for virtual screening studies, where, usually, all the plausible protonation states of the ligands are considered, but, in general, only one protonation or tautomeric state is assigned for every residue in the receptor. The difficulty in determining the most likely tautomeric state for histidine lies in the similarity of their pKas, which also are close to the physiological pH.

While in the putative mechanism, Fig. 1, the δ -nitrogen of His237 deprotonates Cys285, analysis of currently available X-ray structures suggests a different scenario for the reaction mechanism, as mentioned before. Fig. 6 shows an overlap of crystallized structures of caspase-1 with inhibitors containing an aspartate residue in the core fragment, as in the natural substrates. According to these structures, the closest nitrogen atom of the imidazole ring of His237 (δ -nitrogen) and the sulfur atom of Cys285 are separated for more than 5 Å. This means that it is unlikely that the Cys285 could be deprotonated by His237. Also, the average distance

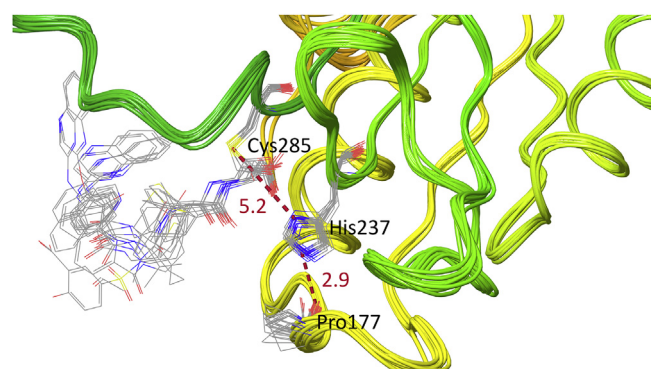


Fig. 6. Catalytic center of wild type crystal structures of human caspase-1 co-crystallized with inhibitors containing an aspartate in the core fragment (PDB codes: 6BZ9, 6F6R, 5MTK, 5MMV, 3NS7, 2HBQ, 1RWK, 1RWM, 1RWN, 1RWO, 1RWP, 1RWV, 1RWW, 1RWX, 1IBC and 1ICE) In red averages distances from nitrogen atoms of the imidazole ring of His237 to the sulfur atom of Cys285 and to the oxygen atom of Pro177 in crystals (Å). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

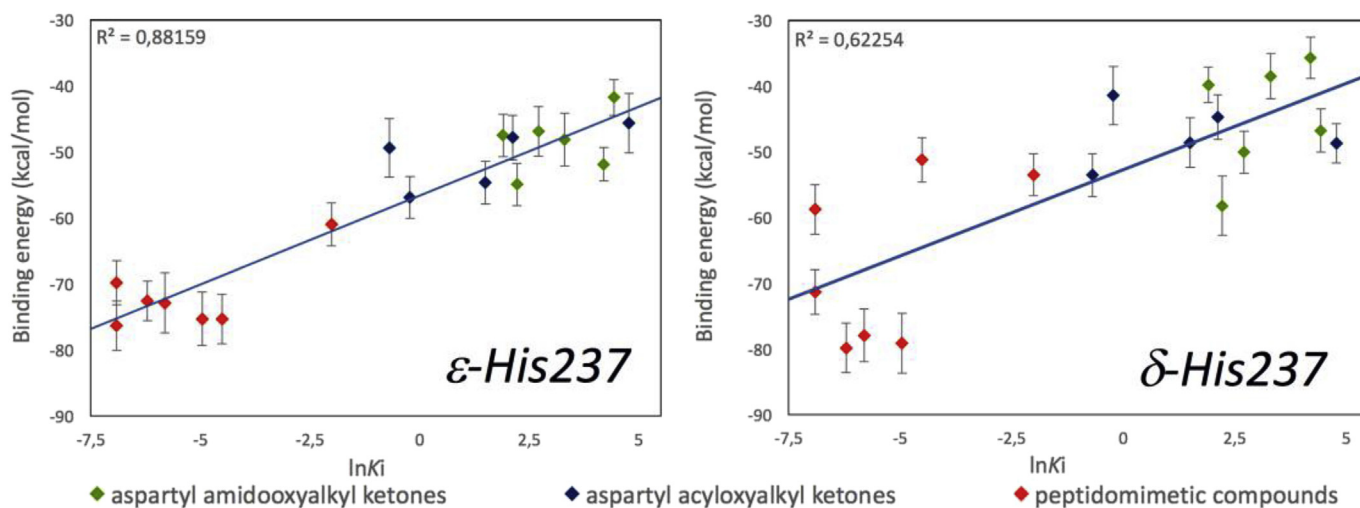


Fig. 5. Correlation between the experimental K_i and the MM/GBSA binding energy calculated for the two systems in different tautomeric states for the His237. Left panel: ϵ -His237; right panel: δ -His237. Simulations of the system with His237 protonated at the ϵ -position reproduce better the experimental inhibition constants.

between the ϵ -nitrogen atom of His237 and the oxygen atom of Pro177 in the nearest loop, is 2.9 Å. This suggests that the His237 orientation could be stabilized by a hydrogen bond between these two atoms and then His237 would be preferentially protonated at the ϵ -position.

Fig. 7 compares the simulations of the protein-ligand complexes in the two tautomeric states with available X-ray structures. This figure shows the averaged distances between the ϵ -nitrogen atom of His237 and the oxygen atom of Pro177 along 50 ns of simulations for every ligand and with the protein in the two tautomeric states (ϵ -His237 and δ -His237). While the RMSD plots (Figs. S3–S6) shows that simulations using the two tautomeric states are stable, only those simulations performed with His237 protonated in the ϵ -position provide structures that are compatible with the X-ray observations. The formation of this hydrogen bond between His237 and Pro177 assist the positioning of the loop containing Pro177 in such a way that the residues of this loop can establish favorable interactions with the ligands (see below).

3.2. Protein-ligand interaction analysis

The interactions of the 18 selected inhibitors with the residues in the active site were analyzed only for the ϵ -His237 case. The per-residue contribution to the binding free energy was obtained for every complex. From our simulations, the following residues were found to contribute more than 1% to the total averaged binding free energy: Ile176, Pro177, Arg179, Met235, Ser236, His237, Gly238, Ile239, His248, Gln283, Ala284, Cys285, Asp 288, Val338, Ser339, Trp340, Arg341, His342, Pro343, and Arg383. The per-residue contribution to the binding free energy is represented in Fig. 8 and their numerical values are given in Table S2. Fig. S7 represents the histograms for every residue in every complex present in Table S2.

According to Fig. 8 and Table S2 those residues that are closer to

the core fragment of the inhibitors, Arg179 and Arg341, are also the residues with the largest contribution to the total binding energy, both for weak and potent inhibitors. Residues Gln283, His237, Trp340, Ser339 and Cys285 also contribute similarly to the binding of all the inhibitors presented here. Most of these residues are found in the S1 region or in the carboxylate pocket. Additionally, it is possible to observe that Arg383, His342 and Pro343 residues (found in the S2-S4 protein region) play an important role differentiating between the subsets of potent (11b, 12c, 12b, 11a, 12a and 1) and weak inhibitors. As seen in Fig. 8, these residues found in the S2-S4 protein region contribute more to the binding energy for the subset of most potent inhibitors. On the other hand, the remaining and less active inhibitors barely interact with this region. This finding suggests the relevance of the interactions established with the S2-S4 region for the inhibitory capacity of the ligands analyzed in this work. Fig. 9 compares the binding modes of a potent (1) and a weak (7i) caspase-1 inhibitors. The P4-P2 fragment in the potent inhibitor interacts with Arg383, His342 and Pro343, while this interaction is not observed in the case of the weak inhibitor. This figure shows two different binding modes in which the inhibitor interacts or not with the S2-S4 region. Potent inhibitors (as compound 1) interact with this region while weaker inhibitors (such as compound 7i) bind in a different pose interacting with the S1 region and the carboxylate pocket.

4. Conclusions

We have used MD simulations and the MM/GBSA method to study the binding mode of a series of inhibitors of the human enzyme caspase-1. We have considered different protonation states for His237, a key residue for the reaction mechanism of this enzyme: His237 positively charged and neutral but with the proton at the ϵ - or δ -nitrogen atoms (ϵ -His237 and δ -His237). We found that the computed binding free energies correlate much better with

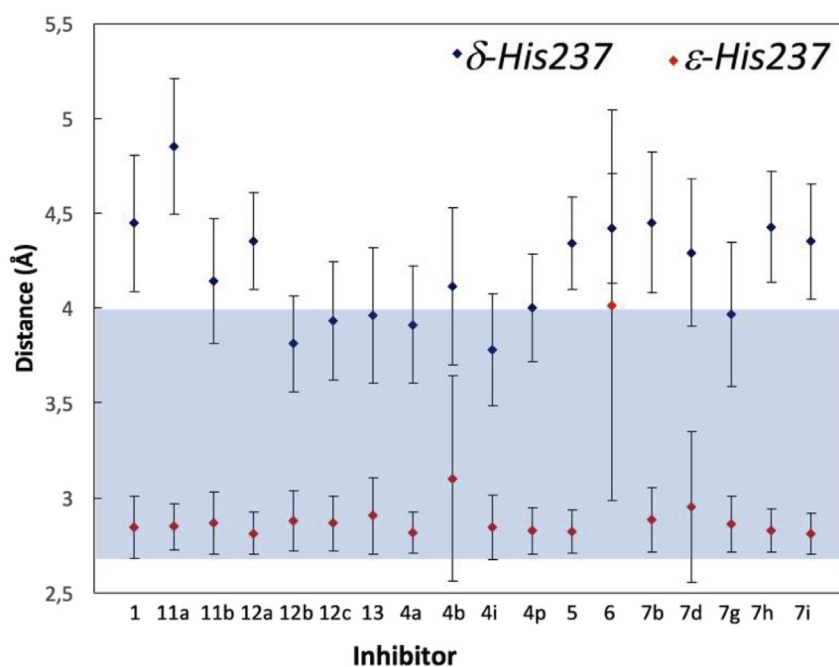


Fig. 7. Average distance between the ϵ -nitrogen of His237 and the oxygen of Pro177 from simulations of the protein-ligand complexes. Red and blue dots correspond to the values obtained for the simulations of ϵ -His237 and δ -His237, respectively. Vertical lines correspond to standard deviations. The blue band corresponds to the maximum and minimum values observed for this distance in the X-ray structures shown in Fig. 6. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

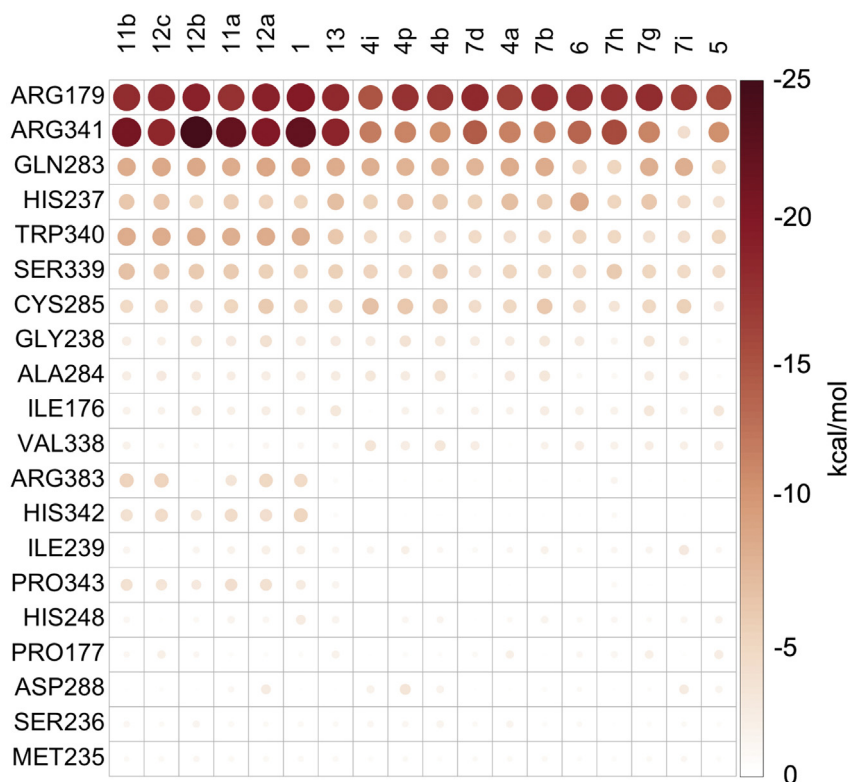


Fig. 8. Per-residue contributions to the binding free energy. The size and color of the circles are determined by the value of the energy contribution. The inhibitors are arranged from more active (left) to less active (right), while residues are ordered from top to bottom according to their average contribution to the binding energy. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

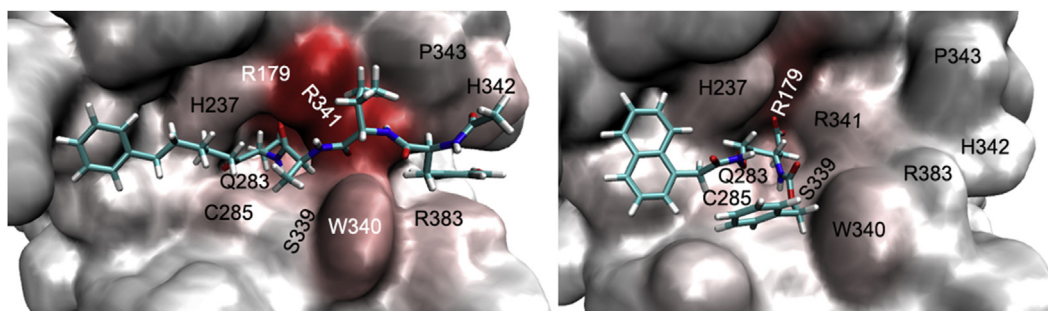


Fig. 9. Left panel: binding mode of a potent inhibitor (**1**) in the active site of caspase-1. Right panel: binding mode of a weak inhibitor (**7i**). The residues labeled are Arg179, His237, Cys285 (catalytic site), Gln283, Ser339, Trp340 and Arg341 (residues with the largest contribution to the binding energy), Pro343, His342 and Arg383 (residues that contribute more to the binding energy for the subset of the most potent inhibitors). Residues are coloured according to their contribution to the free energy.

the experimental inhibition constants for the simulations of the ϵ -His237 for of the enzyme. This tautomeric state is supported by the analysis of X-ray structures of caspase-1 with different inhibitors and would suggest that, contrarily to what is proposed for other cysteine proteases, His237 cannot be the base in charge of Cys285 deprotonation. Instead, His237 could play an important role positioning the loop containing Pro177 and during other steps of the reaction mechanism.

The protein-ligand interaction analysis shows that there are three residues essential to understand the differences between the inhibitory potency of the 18 ligands analyzed here: His342, Pro343 and Arg383, all of them placed in the protein S2-S4 region. A possible strategy for the design of new and more potent inhibitors of caspase-1 should focus in strengthening the interactions established with this region.

Acknowledgements

The authors gratefully acknowledge financial support from Ministerio de Economía y Competitividad of Spain (project CTQ2015-66223-C2-2-P) and Generalitat Valenciana (project AICO/2018/238). The authors acknowledge computational facilities of the Servei d'Informàtica de la Universitat de València in the 'Tirant' supercomputer.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ejmech.2019.02.064>.

References

- [1] M.T. Heneka, M.P. Kummer, A. Stutz, A. Delekate, S. Schwartz, A. Vieira-Saecker, A. Griep, D. Axt, A. Remus, T.-C. Tzeng, E. Gelpi, A. Halle, M. Korte, E. Latz, D.T. Golenbock, NLRP3 is activated in Alzheimer's disease and contributes to pathology in APP/PS1 mice, *Nature* 493 (2012) 674–678, <https://doi.org/10.1038/nature11729>.
- [2] A. Denes, G. Lopez-Castejon, D. Brough, Caspase-1: is IL-1 just the tip of the ICEberg? *Cell Death Dis.* 3 (2012) <https://doi.org/10.1038/cddis.2012.86> e338–e338.
- [3] K.P. Wilson, J.-A.F. Black, J.A. Thomson, E.E. Kim, J.P. Griffith, M.A. Navia, M.A. Murcko, S.P. Chambers, R.A. Aldape, S.A. Raybuck, D.J. Livingston, Structure and mechanism of interleukin-1 β converting enzyme, *Nature* 370 (1994) 270–275, <https://doi.org/10.1038/370270a0>.
- [4] C.E. Cade, A.C. Clark, Caspases – key players in apoptosis, in: K. Bose (Ed.), *Proteases Apoptosis Pathways*, Protoc. Transl. Adv. Springer International Publishing, Cham, 2015, pp. 31–51, https://doi.org/10.1007/978-3-319-19497-4_2.
- [5] A.L. Parui, K. Bose, Caspases: regulatory mechanisms and their implications in pathogenesis and therapeutics, in: *Pathophysiol. Asp. Proteases*, Springer Singapore, Singapore, 2017, pp. 423–488, https://doi.org/10.1007/978-981-10-6141-7_18.
- [6] J. Yang, Z. Liu, C. Wang, R. Yang, J.K. Rathkey, O.W. Pinkard, W. Shi, Y. Chen, G.R. Dubyak, D.W. Abbott, T.S. Xiao, Mechanism of gasdermin D recognition by inflammatory caspases and their inhibition by a gasdermin D-derived peptide inhibitor, *Proc. Natl. Acad. Sci. Unit. States Am.* 115 (2018) 6792–6797, <https://doi.org/10.1073/pnas.1800562115>.
- [7] J.-F. Fournier, L. Clary, S. Chambon, L. Dumais, C.S. Harris, C. Millois, R. Pierre, S. Talano, É. Thoreau, J. Aubert, M. Aurelly, C. Bouix-Peter, A. Brethon, L. Chantalat, O. Christin, C. Comino, G. El-Bazbouz, A.-L. Ghilini, T. Isabet, C. Lardy, A.-P. Luzy, C. Mathieu, K. Mebrouk, D. Orfila, J. Pascau, K. Reverse, D. Roche, V. Rodeschini, L.F. Hennequin, Rational drug design of topically administered caspase 1 inhibitors for the treatment of inflammatory acne, *J. Med. Chem.* 61 (2018) 4030–4051, <https://doi.org/10.1021/acs.jmedchem.8b00067>.
- [8] P. Galatsis, B. Caprathe, J. Gilmore, A. Thomas, K. Linn, S. Sheehan, W. Harter, C. Kostlan, E. Lunnay, C. Stankovic, J. Rubin, K. Brady, H. Allen, R. Talanian, Succinic acid amides as P2–P3 replacements for inhibitors of interleukin-1 β converting enzyme (ICE or caspase 1), *Bioorg. Med. Chem. Lett* 20 (2010) 5184–5190, <https://doi.org/10.1016/j.bmcl.2010.07.004>.
- [9] J.M. Scheer, M.J. Romanowski, J.A. Wells, A common allosteric site and mechanism in caspases, *Proc. Natl. Acad. Sci. U.S.A.* 103 (2006) 7595–7600, <https://doi.org/10.1073/pnas.0602571103>.
- [10] T. O'Brien, B.T. Fahr, M.M. Sopko, J.W. Lam, N.D. Waal, B.C. Raimundo, H.E. Purkey, P. Pham, M.J. Romanowski, Structural analysis of caspase-1 inhibitors derived from tethering, *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* 61 (2005) 451–458, <https://doi.org/10.1107/S1744309105010109>.
- [11] B.T. Fahr, T. O'Brien, P. Pham, N.D. Waal, S. Baskaran, B.C. Raimundo, J.W. Lam, M.M. Sopko, H.E. Purkey, M.J. Romanowski, Tethering identifies fragment that yields potent inhibitors of human caspase-1, *Bioorg. Med. Chem. Lett* 16 (2006) 559–562, <https://doi.org/10.1016/j.bmcl.2005.10.048>.
- [12] T.A. Rano, T. Timkey, E.P. Peterson, J. Rotonda, D.W. Nicholson, J.W. Becker, K.T. Chapman, N.A. Thornberry, A combinatorial approach for determining protease specificities: application to interleukin-1 β converting enzyme (ICE), *Chem. Biol.* 4 (1997) 149–155, [https://doi.org/10.1016/S1074-5521\(97\)90258-1](https://doi.org/10.1016/S1074-5521(97)90258-1).
- [13] G. Pietro Miscione, M. Calvaresi, A. Bottoni, Computational evidence for the catalytic mechanism of caspase-7. A DFT investigation, *J. Phys. Chem. B* 114 (2010) 4637–4645, <https://doi.org/10.1021/jp908991z>.
- [14] M. Sulpizi, A. Laio, J. VandeVondele, A. Cattaneo, U. Rothlisberger, P. Carloni, Reaction mechanism of caspases: insights from QM/MM Car-Parrinello simulations, *Proteins Struct. Funct. Genet.* 52 (2003) 212–224, <https://doi.org/10.1002/prot.10275>.
- [15] B. Elsässer, F.B. Zauner, J. Messner, W.T. Soh, E. Dall, H. Brandstetter, Distinct roles of catalytic cysteine and histidine in the protease and ligase mechanisms of human legumain as revealed by DFT-based QM/MM simulations, *ACS Catal.* 7 (2017) 5585–5593, <https://doi.org/10.1021/acscatal.7b01505>.
- [16] M. Sulpizi, U. Rothlisberger, P. Carloni, Molecular dynamics studies of caspase-3, *Biophys. J.* 84 (2003) 2207–2215, [https://doi.org/10.1016/S0006-3495\(03\)75026-7](https://doi.org/10.1016/S0006-3495(03)75026-7).
- [17] H. Lee, E.A. Shin, J.H. Lee, D. Ahn, C.G. Kim, J.H. Kim, S.H. Kim, Caspase inhibitors: a review of recently patented compounds (2013–2015), *Expert Opin. Ther. Pat.* 28 (2018) 47–59, <https://doi.org/10.1080/13543776.2017.1378426>.
- [18] A.M.M. Mjalli, K.T. Chapman, M. MacCoss, N.A. Thornberry, Phenylalkyl ketones as potent reversible inhibitors of interleukin-1 β converting enzyme, *Bioorg. Med. Chem. Lett* 3 (1993) 2689–2692, [https://doi.org/10.1016/S0960-894X\(01\)80743-5](https://doi.org/10.1016/S0960-894X(01)80743-5).
- [19] G. Le, G. Abbenante, Inhibitors of TACE and caspase-1 as anti-inflammatory drugs, *Curr. Med. Chem.* 12 (2005) 2963–2977, <https://doi.org/10.2174/092986705774462851>.
- [20] P. Galatsis, B. Caprathe, D. Downing, J. Gilmore, W. Harter, S. Hays, C. Kostlan, K. Linn, E. Lunnay, K. Para, A. Thomas, J. Warmus, H. Allen, K. Brady, R. Talanian, N. Walker, Inhibition of interleukin-1 β converting enzyme (ICE or caspase 1) by aspartyl acyloxyalkyl ketones and aspartyl amidoxyalkyl ketones, *Bioorg. Med. Chem. Lett* 20 (2010) 5089–5094, <https://doi.org/10.1016/j.bmcl.2010.07.031>.
- [21] S. Genheden, U. Ryde, The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities, *Expert Opin. Drug Discov.* 10 (2015) 449–461, <https://doi.org/10.1517/17460441.2015.1032936>.
- [22] Z. Courmia, B. Allen, W. Sherman, Relative Binding Free Energy Calculations in Drug Discovery: Recent Advances and Practical Considerations, 2017, <https://doi.org/10.1021/acs.jcim.7b00564>.
- [23] M. De Vivo, M. Masetti, G. Bottegoni, A. Cavalli, The role of molecular dynamics and related methods in drug discovery, *J. Med. Chem.* (2016), <https://doi.org/10.1021/acs.jmedchem.5b01684>.
- [24] M.D. Eldridge, C.W. Murray, T.R. Auton, G.V. Paolini, R.P. Mee, Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes, *J. Comput. Aided Mol. Des.* 11 (1997) 425–445, <https://doi.org/10.1023/A:1007996124545>.
- [25] S.-Y. Huang, S.Z. Grinter, X. Zou, Scoring functions and their evaluation for protein-ligand docking: recent advances and future directions, *Phys. Chem. Chem. Phys.* 12 (2010) 12899–12908, <https://doi.org/10.1039/c0cp00151a>.
- [26] I. Massova, P.A. Kollman, Combined molecularmechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding, *Perspect. Drug Discov. Des.* 18 (2000) 113–135, <https://doi.org/10.1023/A:1008763014207>.
- [27] T. Hou, J. Wang, Y. Li, W. Wang, Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations, *J. Chem. Inf. Model.* 51 (2011) 69–82, <https://doi.org/10.1021/ci100275a>.
- [28] P.A. Greenidge, C. Kramer, J. Mozziconacci, R.M. Wolf, MM/GBSA binding energy prediction on the PDBbind data set: successes, failures, and directions for further improvement, *J. Chem. Inf. Model.* 53 (2013) 201–209, <https://doi.org/10.1021/ci300425v>.
- [29] F. Adasme-Carreño, C. Muñoz-Gutierrez, J. Caballero, J.H. Alzate-Morales, Performance of the MM/GBSA scoring using a binding site hydrogen bond network-based frame selection: the protein kinase case, *Phys. Chem. Chem. Phys.* 16 (2014) 14047–14058, <https://doi.org/10.1039/C4CP01378F>.
- [30] S.R. Zia, Z. Ul-Haq, Molecular dynamics simulation of interleukin-2 and its complex and determination of the binding free energy, *Mol. Simulat.* 44 (2018) 1411–1425, <https://doi.org/10.1080/08927022.2018.1513651>.
- [31] J. Wang, W. Wang, P.A. Kollman, D.A. Case, Automatic atom type and bond type perception in molecular mechanical calculations, *J. Mol. Graph. Model.* 25 (2006) 247–260, <https://doi.org/10.1016/j.jmgm.2005.12.005>.
- [32] Y. D.M., P.A.K.D.A. Case, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, D. Ghoreishi, M.K. Gilson, H. Gohlke, A.W. Goetz, D. Greene, R. Harris, N. Homeyer, S. Izadi, A. Kovalevko, T. Kurtzman, T.S. Lee, S. LeGra, Amber 2018 (2018).
- [33] J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, D.A. Case, Development and testing of a general amber force field, *J. Comput. Chem.* 25 (2004) 1157–1174, <https://doi.org/10.1002/jcc.20035>.
- [34] Schrödinger Release 2016-3, Maestro, Schrödinger, LLC, New York, NY, 2016.
- [35] G.M. Sastry, S.L. Dixon, W. Sherman, Rapid shape-based ligand alignment and virtual screening method based on atom/feature-pair similarities and volume overlap scoring, *J. Chem. Inf. Model.* 51 (2011) 2455–2466, <https://doi.org/10.1021/ci2002704>.
- [36] A.C. Lee, G.M. Crippen, Predicting p K a, *J. Chem. Inf. Model.* 49 (2009) 2013–2033, <https://doi.org/10.1021/ci900209w>.
- [37] E. Rossini, R.R. Netz, E.-W. Knapp, Computing p K a values in different solvents by electrostatic transformation, *J. Chem. Theor. Comput.* 12 (2016) 3360–3369, <https://doi.org/10.1021/acs.jctc.6b00446>.
- [38] M.H.M. Olsson, C.R. Søndergaard, M. Rostkowski, J.H. Jensen, PROPKA3: consistent treatment of internal and surface residues in empirical p K a predictions, *J. Chem. Theor. Comput.* 7 (2011) 525–537, <https://doi.org/10.1021/ct100578z>.
- [39] J.C. Gordon, J.B. Myers, T. Folta, V. Shoja, L.S. Heath, A. Onufriev, H++: a server for estimating pKas and adding missing hydrogens to macromolecules, *Nucleic Acids Res.* 33 (2005) W368–W371, <https://doi.org/10.1093/nar/gki464>.
- [40] J.A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K.E. Hauser, C. Simmerling, ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB, *J. Chem. Theor. Comput.* 11 (2015) 3696–3713, <https://doi.org/10.1021/acs.jctc.5b00255>.
- [41] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, Comparison of simple potential functions for simulating liquid water, *J. Chem. Phys.* 79 (1983) 926, <https://doi.org/10.1063/1.445869>.
- [42] J.-P. Ryckaert, G. Ciccotti, H.J. Berendsen, Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes, *J. Comput. Phys.* 23 (1977) 327–341, [https://doi.org/10.1016/0021-9991\(77\)90098-5](https://doi.org/10.1016/0021-9991(77)90098-5).
- [43] T. Darden, D. York, L. Pedersen, Particle mesh Ewald: an N -log(N) method for Ewald sums in large systems, *J. Chem. Phys.* 98 (1993) 10089–10092, <https://doi.org/10.1063/1.464397>.
- [44] U. Essmann, L. Perera, M.L. Berkowitz, T. Darden, H. Lee, L.G. Pedersen, A smooth particle mesh Ewald method, *J. Chem. Phys.* 103 (1995) 8577–8593, <https://doi.org/10.1063/1.470117>.
- [45] S. Le Grand, A.W. Götz, R.C. Walker, SPFP: speed without compromise - A

- mixed precision model for GPU accelerated molecular dynamics simulations, *Comput. Phys. Commun.* 184 (2013) 374–380, <https://doi.org/10.1016/j.cpc.2012.09.022>.
- [46] R. Salomon-Ferrer, A.W. Götz, D. Poole, S. Le Grand, R.C. Walker, Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh ewald, *J. Chem. Theor. Comput.* 9 (2013) 3878–3888, <https://doi.org/10.1021/ct400314y>.
- [47] M. Misini Ignjatović, O. Caldararu, G. Dong, C. Muñoz-Gutierrez, F. Adasme-Carreño, U. Ryde, Binding-affinity predictions of HSP90 in the D3R Grand Challenge 2015 with docking, MM/GBSA, QM/MM, and free-energy simulations, *J. Comput. Aided Mol. Des.* 30 (2016) 707–730, <https://doi.org/10.1007/s10822-016-9942-z>.
- [48] A. Jadhav, R. Dash, R. Hirwani, M. Abdin, Sequence and structure insights of kazal type thrombin inhibitor protein: studied with phylogeny, homology modeling and dynamic MM/GBSA studies, *Int. J. Biol. Macromol.* 108 (2018) 1045–1052, <https://doi.org/10.1016/j.ijbiomac.2017.11.020>.
- [49] S.R. Peddi, S.K. Sivan, V. Manga, Molecular dynamics and MM/GBSA-integrated protocol probing the correlation between biological activities and binding free energies of HIV-1 TAR RNA inhibitors, *J. Biomol. Struct. Dyn.* 36 (2018) 486–503, <https://doi.org/10.1080/07391102.2017.1281762>.
- [50] B.K. Shukla, U. Yadava, Effect of polarization of ligand charges and estimation of MM/GBSA binding free energies of some Pyrazolo[3,4-d]pyrimidine inhibitors of Mycobacterium tuberculosis in view of experimental results, *Curr. Enzym. Inhib.* 13 (2017), <https://doi.org/10.2174/1573408013666170113120426>.
- [51] X. Zhang, H. Perez-Sanchez, F.C. Lightstone, A comprehensive docking and MM/GBSA rescoring study of ligand recognition upon binding antithrombin, *Curr. Top. Med. Chem.* 17 (2017) 1631–1639, <https://doi.org/10.2174/1568026616666161117112604>.
- [52] D.A. Pearlman, Evaluating the molecular mechanics Poisson–Boltzmann surface area free energy method using a congeneric series of ligands to p38 MAP kinase, *J. Med. Chem.* 48 (2005) 7796–7807, <https://doi.org/10.1021/jm050306m>.
- [53] G. Rastelli, A. Del Rio, G. Degliesposti, M. Sgobba, Fast and accurate predictions of binding free energies using MM-PBSA and MM-GBSA, *J. Comput. Chem.* (2009), <https://doi.org/10.1002/jcc.21372>. NA-NA.
- [54] B.R. Miller, T.D. McGee, J.M. Swails, N. Homeyer, H. Gohlke, A. Roitberg, MMPBSA.py: an efficient program for end-state free energy calculations, *J. Chem. Theor. Comput.* 8 (2012) 3314–3321, <https://doi.org/10.1021/ct300418h>.
- [55] S. Patel, P. Modi, M. Chhabria, Rational approach to identify newer caspase-1 inhibitors using pharmacophore based virtual screening, docking and molecular dynamic simulation studies, *J. Mol. Graph. Model.* 81 (2018) 106–115, <https://doi.org/10.1016/j.jmgm.2018.02.017>.
- [56] S. Patel, P. Modi, V. Ranjan, M. Chhabria, Structure-based design, synthesis and evaluation of 2,4-diaminopyrimidine derivatives as novel caspase-1 inhibitors, *Bioorg. Chem.* 78 (2018) 258–268, <https://doi.org/10.1016/j.bioorg.2018.03.019>.
- [57] R. Ganesan, S. Jelakovic, P.R.E. Mittl, A. Cafilisch, M.G. Grütter, In silico identification and crystal structure validation of caspase-3 inhibitors without a P1 aspartic acid moiety, *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* 67 (2011) 842–850, <https://doi.org/10.1107/S1744309111018604>.

Supplementary Material

Modeling caspase-1 inhibition: implications for catalytic mechanism and drug design

Carlos A. Ramos-Guzmán, Kirill Zinovjev and Iñaki Tuñón*

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

Corresponding author: Ignacio.tunon@uv.es

Table S1. Binding Free Energies obtained from MM/GBSA calculations (in kcal·mol⁻¹) for two protonation states of His237 and experimental inhibition constants from ref [8]

INHIBITOR			δ -Tauomer		ϵ -Tautomer	
	lnKi	Ki	ΔG	Std. Dev.	ΔG	Std. Dev.
5	4.78	119	-48.75	2.99	-45.62	4.50
7i	4.43	84	-46.78	3.26	-41.78	2.69
7g	4.20	66.5	-35.71	3.15	-51.88	2.56
7h	3.31	27.5	-38.53	3.44	-48.19	4.00
6	2,71	15	-50.05	3.19	-46.93	3.76
7b	2.22	9.2	-58.24	4.50	-54.94	3.18
4a	2.12	8.3	-44.76	3.39	-47.85	3.38
7d	1.91	6.77	-39.84	2.69	-47.55	3.19
4b	1.50	4.5	-48.61	3.74	-54.68	3.26
4p	-0.22	0.8	-41.46	4.42	-56.89	3.15
4i	-0.69	0.5	-53.56	3.29	-49.38	4.42
13	-2.01	0.134	-53.51	3.18	-60.98	3.29
1	-4.51	0.011	-51.24	3.33	-75.32	3.74
12a	-4.96	0.007	-79.12	4.53	-75.27	3.73
11a	-5.81	0.003	-77.94	4.00	-72.88	4.53
12b	-6.21	0.002	-79.85	3.76	-72.57	2.99
12c	-6.91	0.001	-71.33	3.38	-69.80	3.33
11b	-6.91	0.001	-58.78	3.77	-76.31	3.77

Table S2 Contribution to the binding energy per-residues of the active site. Results in kcal·mol⁻¹.

	11b	12c	12b	11a	12a	1	13	4i	4p	4b	7d	4a	7b	6	7h	7g	7i	5	Mean
ARG179	-18.00	-18.02	-18.79	-17.48	-18.93	-19.65	-18.11	-14.65	-17.30	-16.81	-18.09	-16.23	-17.60	-17.34	-17.44	-17.91	-16.64	-15.49	-17.47
ARG341	-20.95	-18.30	-25.20	-22.07	-19.82	-22.38	-18.63	-11.61	-11.19	-10.44	-14.07	-11.40	-11.45	-13.33	-15.43	-11.20	-3.73	-10.31	-15.08
GLN283	-8.03	-8.39	-8.29	-8.02	-8.59	-8.50	-8.08	-7.78	-7.61	-7.71	-7.28	-8.15	-8.14	-4.85	-4.75	-7.96	-7.81	-4.57	-7.47
HIS237	-5.78	-6.17	-4.50	-5.50	-4.86	-4.51	-6.68	-5.06	-6.18	-5.72	-5.04	-6.72	-5.73	-8.28	-4.54	-5.82	-4.11	-3.23	-5.47
TRP340	-8.06	-8.08	-8.12	-7.82	-8.06	-7.85	-5.80	-4.01	-3.50	-3.51	-4.05	-3.70	-3.76	-4.54	-4.28	-3.50	-3.55	-4.67	-5.38
SER339	-6.42	-5.96	-5.72	-5.67	-5.14	-4.67	-5.09	-4.76	-4.22	-5.31	-3.63	-4.67	-4.34	-4.25	-5.60	-4.53	-4.25	-3.85	-4.89
CYS285	-4.02	-4.07	-3.62	-4.70	-5.52	-4.38	-4.29	-6.50	-5.95	-5.38	-3.80	-4.39	-5.81	-3.85	-2.84	-4.45	-5.02	-2.49	-4.50
GLY238	-1.94	-1.68	-2.74	-2.48	-3.32	-2.23	-2.40	-2.21	-3.01	-2.54	-2.11	-2.05	-2.63	-2.16	-1.11	-2.80	-2.09	-0.48	-2.22
ALA284	-1.91	-2.32	-1.80	-1.81	-1.96	-1.96	-2.10	-2.60	-2.11	-2.71	-0.68	-2.41	-2.73	-0.74	-0.73	-2.06	-1.97	-0.38	-1.83
ILE176	-1.34	-1.27	-2.07	-1.54	-1.82	-1.74	-2.68	-0.12	-1.27	-1.24	-1.42	-1.27	-1.95	-1.68	-1.36	-2.53	-1.23	-2.58	-1.62
VAL338	-1.43	-0.72	-0.61	-0.46	-0.85	-0.90	-0.84	-2.84	-1.99	-2.71	-1.89	-0.21	-1.36	-1.93	-1.45	-1.80	-1.55	-1.98	-1.42
ARG383	-4.76	-4.93	-0.24	-2.94	-4.40	-4.17	-0.46	-0.13	-0.09	-0.11	-0.13	-0.10	-0.11	-0.19	-1.03	-0.06	-0.19	-0.15	-1.34
HIS342	-3.43	-3.93	-2.72	-3.93	-3.60	-4.59	-0.39	-0.02	-0.03	-0.03	-0.04	-0.02	0.00	-0.02	-0.59	-0.01	-0.01	-0.02	-1.30
ILE239	-1.12	-0.05	-1.05	-1.33	-1.73	-1.69	-1.00	-1.19	-1.54	-0.77	-0.62	-0.79	-1.37	-0.57	-0.90	-1.09	-2.35	-0.79	-1.11
PRO343	-3.29	-2.95	-2.27	-3.54	-3.35	-2.18	-1.16	-0.01	-0.01	-0.01	-0.02	-0.01	0.00	-0.01	-0.60	0.00	-0.01	-0.01	-1.08
HIS248	-0.91	-0.13	-0.75	-1.24	-0.91	-2.19	-1.21	-0.06	-0.98	-1.05	-0.07	-0.63	-1.24	-0.63	-0.86	-0.53	-1.00	-1.26	-0.87
PRO177	-0.79	-1.58	-0.79	-0.37	-0.32	-0.44	-1.42	-0.07	-0.44	-0.36	-0.58	-1.52	-0.19	-0.84	-1.00	-1.64	-0.23	-1.86	-0.80
ASP288	-0.19	-0.34	-0.09	-0.80	-2.19	-0.07	-0.03	-1.41	-2.78	-1.28	-0.12	-0.06	-0.48	-0.51	-0.11	-0.31	-2.10	-1.15	-0.78
SER236	-0.91	-0.53	-1.13	-0.69	-0.58	-0.69	-0.61	-0.92	-0.85	-1.23	-0.53	-1.13	-0.43	-0.58	-0.13	-0.43	-0.60	-0.40	-0.69
MET235	-0.57	-0.65	-0.85	-0.71	-0.69	-0.71	-0.63	-0.48	-0.61	-0.42	-0.54	-0.62	-0.69	-0.45	-0.46	-0.65	-0.78	-0.44	-0.61
ΔG	-76.31	-69.80	-72.57	-72.88	-75.27	-75.32	-60.98	-49.38	-56.89	-54.68	-47.55	-47.85	-54.94	-46.93	-48.19	-51.88	-41.78	-45.62	-58.27

Figure S1. RMSD for the complex between caspase-1 and inhibitor **1** along 50 ns MD simulation assuming that both Cys285 and His237 are ionized. X-axis represents the simulation time and the Y-axis the RMSD with respect to the first frame (in Å).

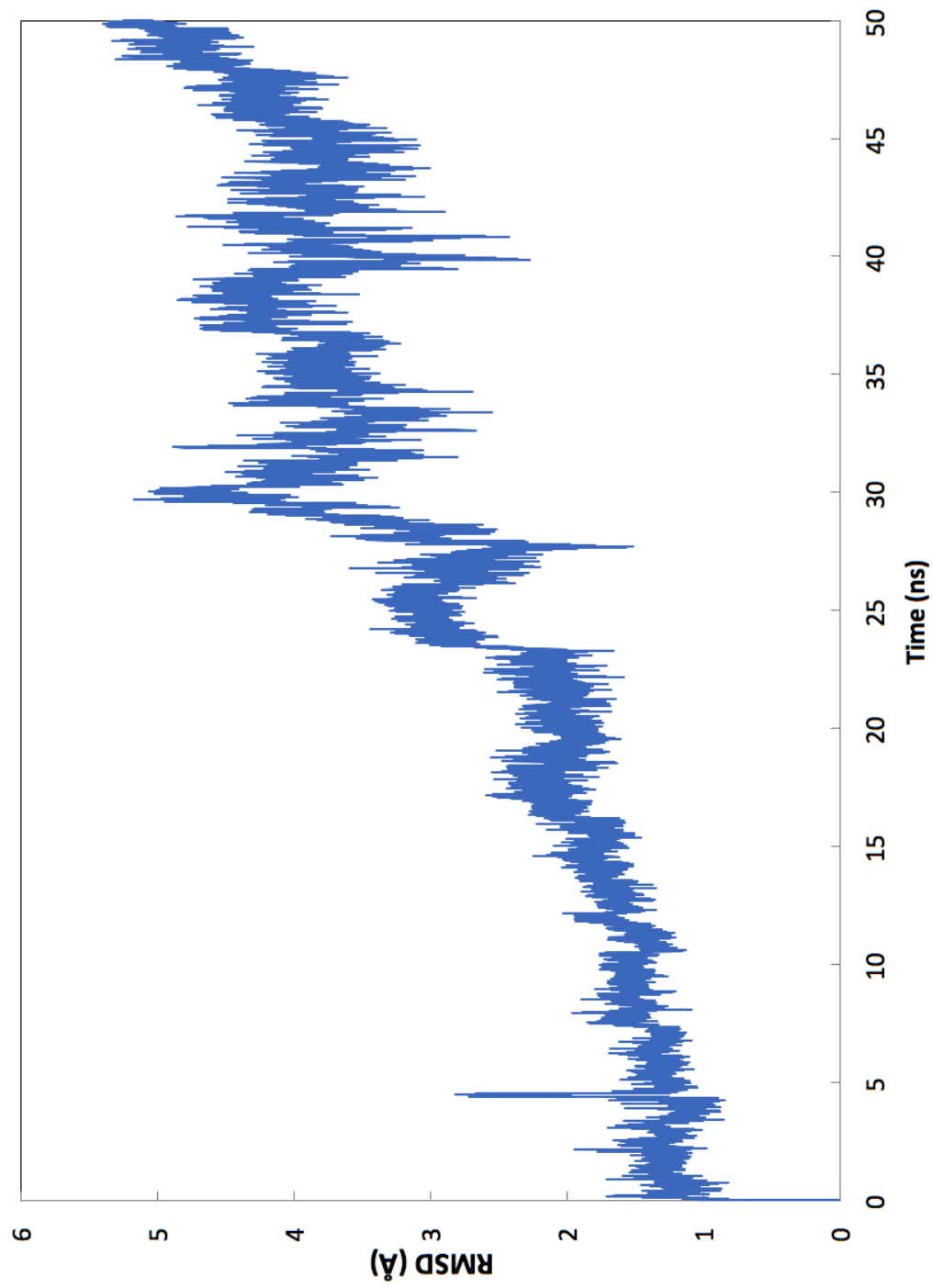


Figure S2. Dihedral angle formed by $C\alpha$ - $C\beta$ - $C\gamma$ - $N\delta$ atoms in His237 in the complex formed by caspase-1 with the reference inhibitor **1** along 50 ns MD simulation, with the two tautomeric states (ϵ and δ).

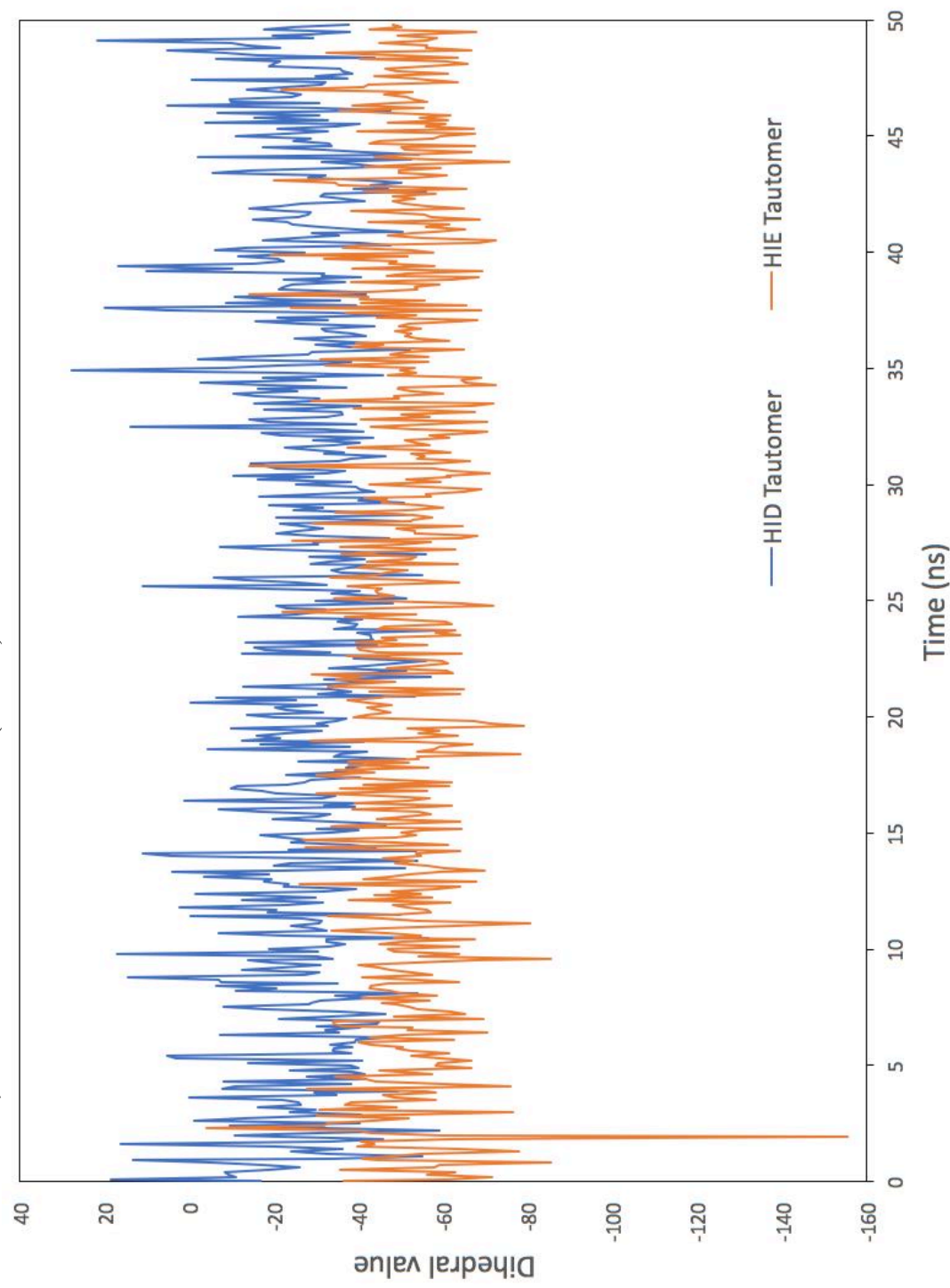


Figure S3. RMSD per inhibitor along 50 ns MD simulation, X-axis shows the frame number (1 frame each 2 ps), Y-axis RMSD with respect to the first frame of the MD simulation (in Å). The results correspond to the simulation with His237 in the ϵ -tautomeric state.

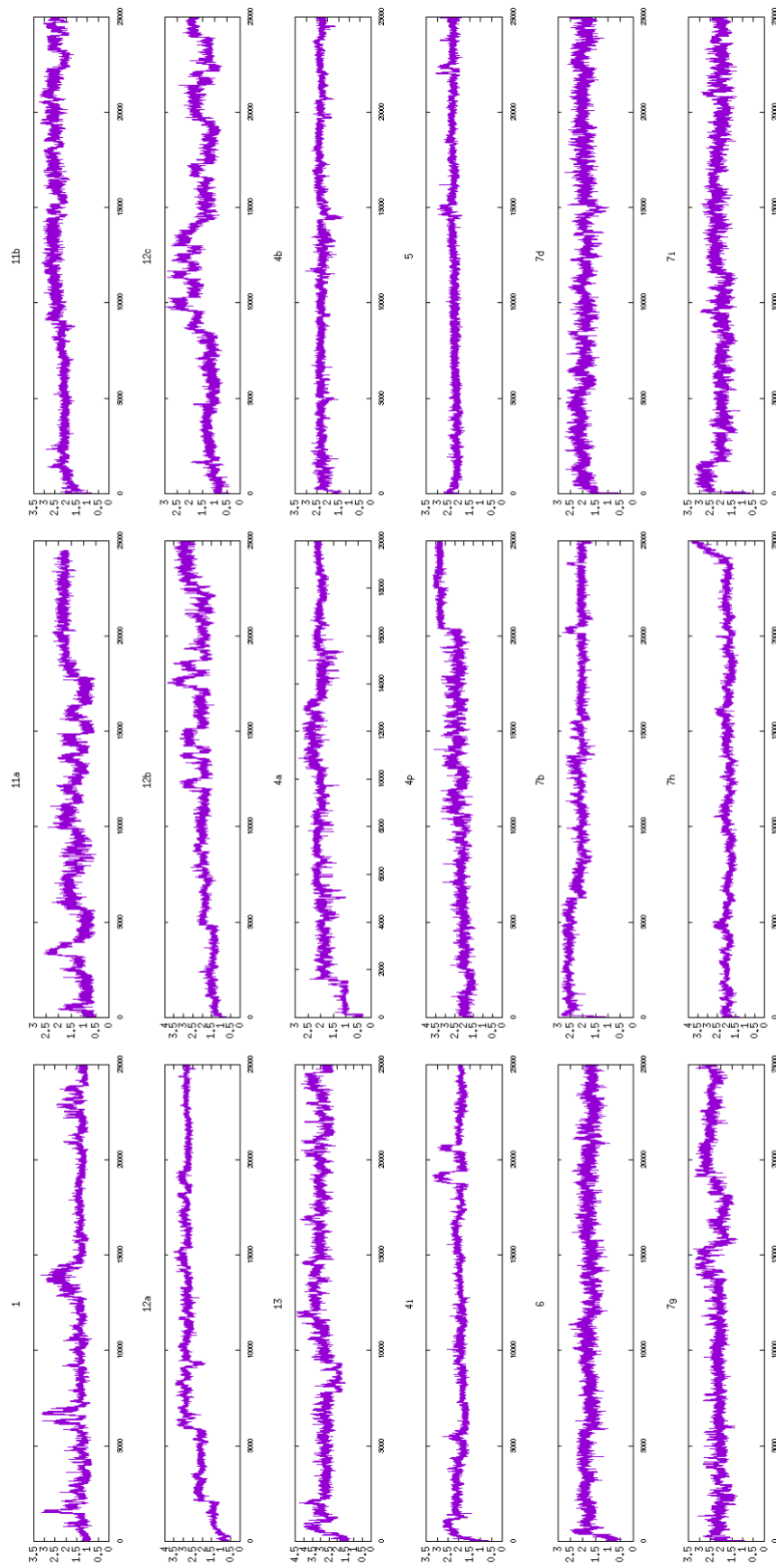


Figure S4. RMSD per inhibitor along 50 ns MD simulation, X-axis shows the frame number (1 frame each 2 ps), Y-axis RMSD with respect to the first frame of the MD simulation (in Å). The results correspond to the simulation with His237 in the δ -tautomeric state.

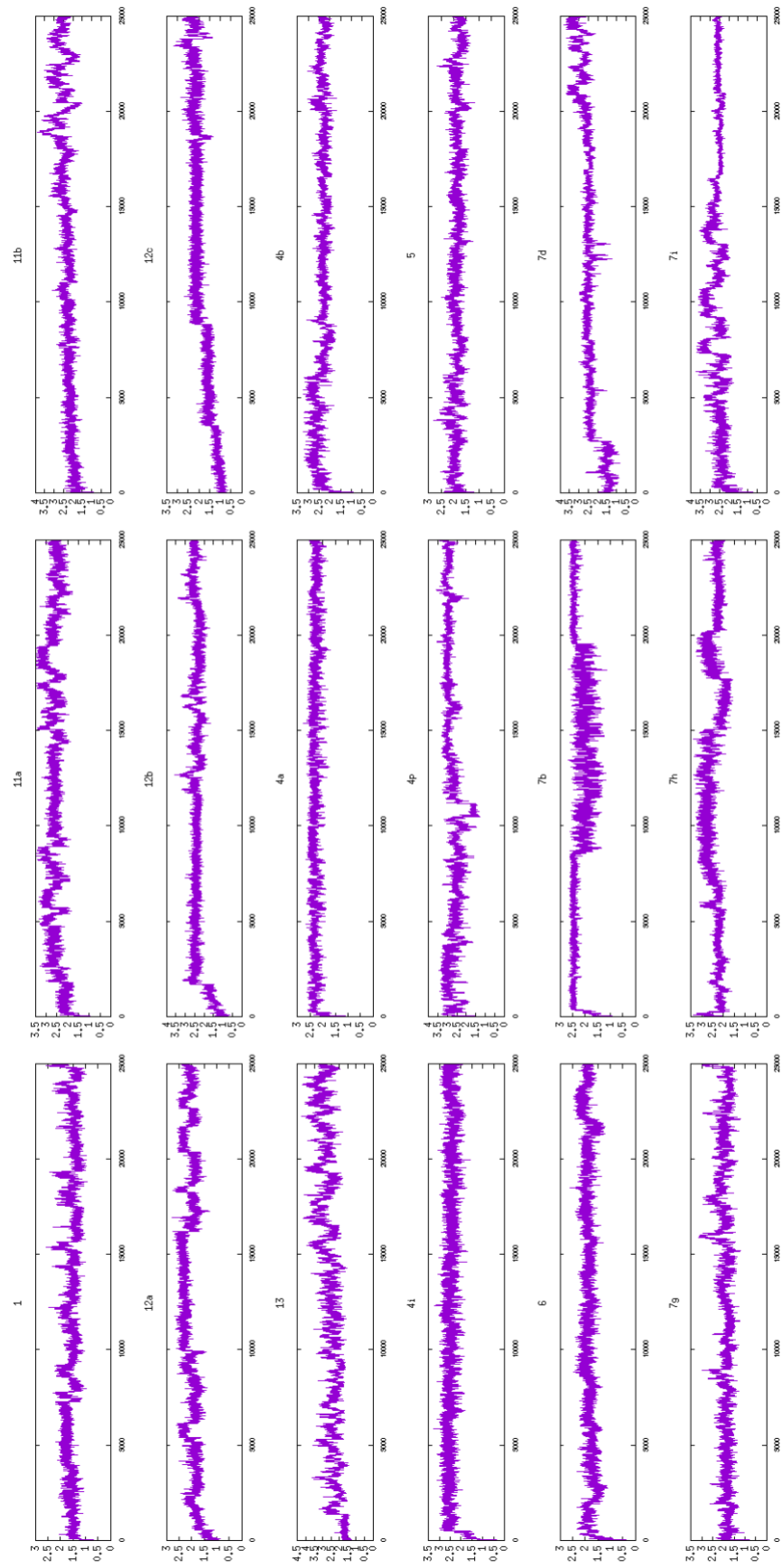


Figure S5. Protein backbone RMSD per complex along 50 ns MD simulation, X-axis shows the frame number (1 frame each 2 ps), Y-axis RMSD with respect to the first frame of the MD simulation (in Å). The results correspond to the simulation with His237 in the ϵ -tautomeric state.

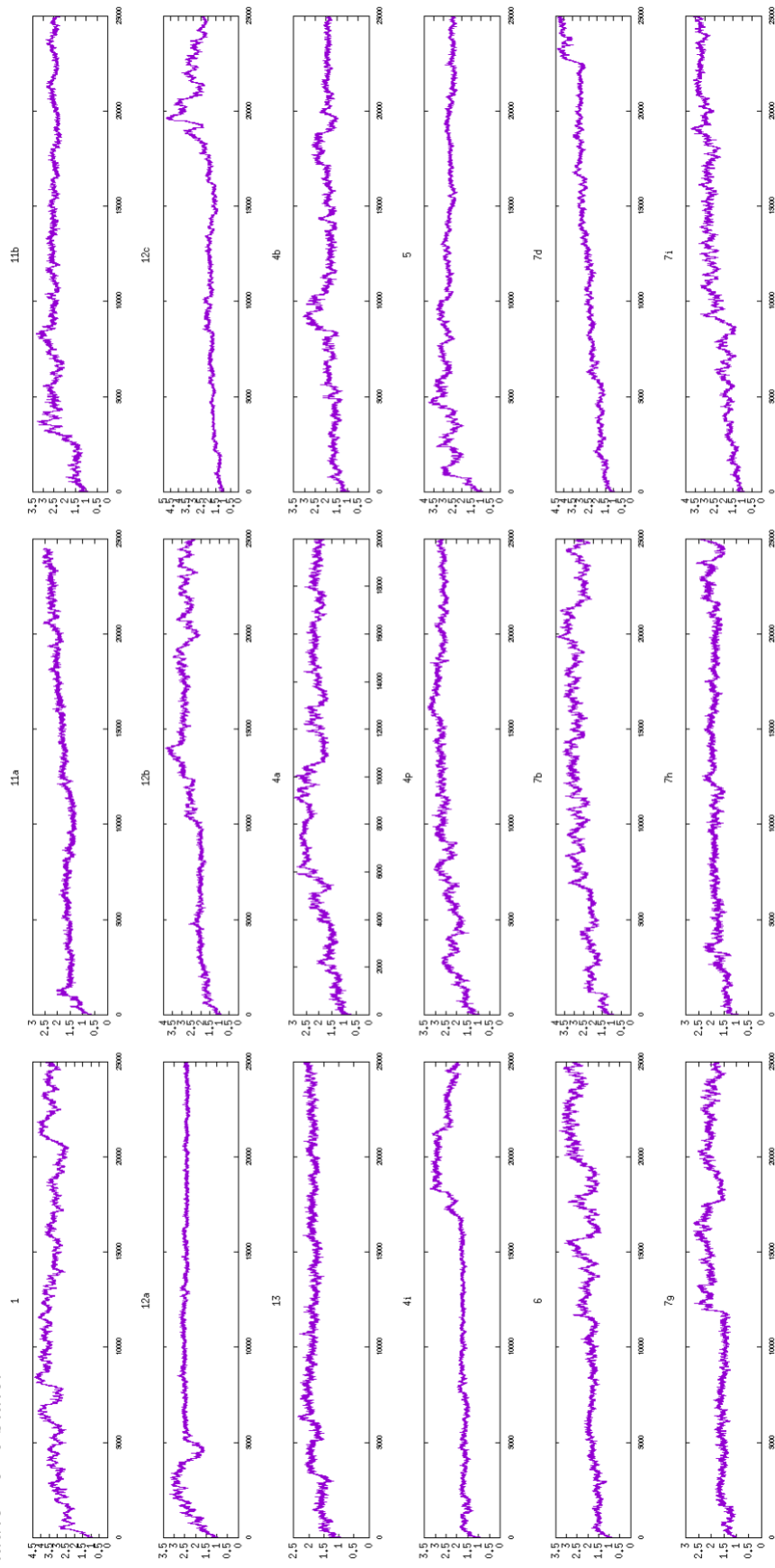


Figure S6. Protein backbone RMSD per complex along 50 ns MD simulation, X-axis shows the frame number (1 frame each 2 ps), Y-axis RMSD with respect to the first frame of the MD simulation (in Å). The results correspond to the simulation with His237 in the δ -tautomeric state.

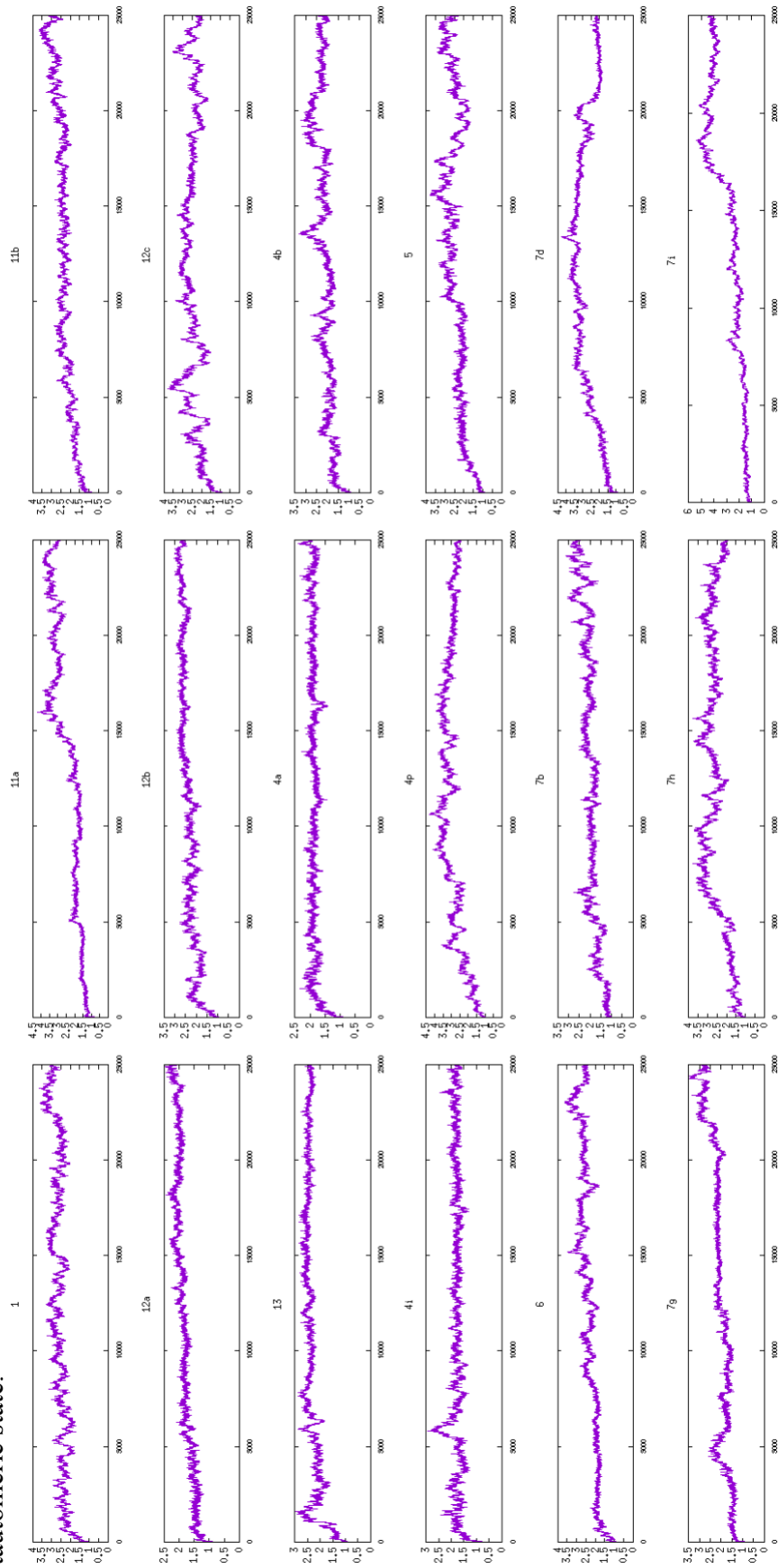


Figure S7. Binding free energy contribution per-residue histograms X-axis number of residue, Y-axis energy (kcal·mol⁻¹).



Appendix 2

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



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

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



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



Read Online

ACCESS |



Metrics & More



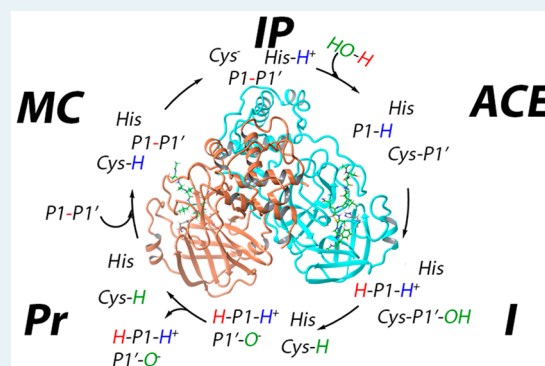
Article Recommendations



Supporting Information

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

KEYWORDS: 3CL^{pro} protease, SARS-CoV-2, minimum free energy path, QM/MM, acylation, deacylation, molecular dynamics



1. INTRODUCTION

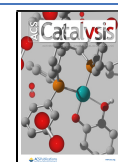
The recent outbreak of COVID-19, a pneumonia-like illness caused by a coronavirus named SARS-CoV-2,¹ has rapidly evolved into a pandemic as recognized by the World Health Organization. SARS-CoV-2 has been shown to be highly contagious, causing a large number of infections around the world. The absence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

The 3CL^{pro} enzyme of SARS-CoV-2 exists as a functional homodimer with two active sites in charge of cleaving the translated polyproteins into individual fragments to be used by the coronavirus.⁵ As with other cysteine proteases, each of the active sites contains a Cys-His catalytic dyad in charge of the hydrolysis of the peptide bond at specific sites of a polypeptide chain. Several structures of this protease have been already resolved by means of X-ray crystallography and deposited in the Protein Data Bank (PDB), including the free protease (PDB codes 6Y2E² and 6Y84⁶) and inhibitor-bound proteases (PDB codes 6LU7,⁶ 6Y2F,⁷ and 6LZE⁸). The SARS-CoV-2 main protease has a structure that is virtually identical to the ortholog from SARS-CoV (96% identity). Even more, the main residues involved in catalysis, binding, and dimerization processes are fully conserved.⁹ Consequently, these two ortholog enzymes display highly similar substrate preferences.¹⁰ The substrate cleavage by the 3CL^{pro} takes place between Gln at the P1 position and a Gly/Ala/Ser at the P1' position (P and P' identify the residues placed before and after the scissile bond, respectively), making the presence of Gln an essential requirement.¹¹

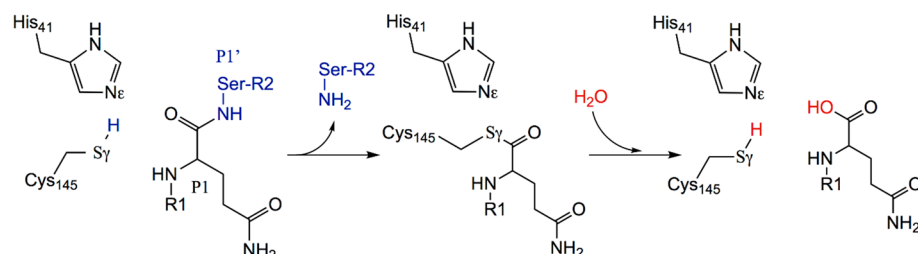
Received: August 5, 2020

Revised: September 26, 2020

Published: September 28, 2020



Scheme 1. Reaction Mechanism in Cysteine Proteases



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

2. METHODS

2.1. Classical Molecular Dynamics Simulations. The crystal structures with PDB codes 6Y2F⁷ and 3AW0¹⁴ were used as starting points to build the Michaelis complex. The former corresponds to the holoprotease of SARS-CoV-2, and the latter is the crystallographic structure of the SARS-CoV ortholog cocrystallized with the peptidic aldehyde inhibitor Ac-Ser-Ala-Val-Leu-His-aldehyde. To build the Michaelis complex corresponding to the 3CLP^{ro} protease of SARS-CoV-2, the two protein structures were aligned and then the cocrystallized ligand in 6Y2F was replaced with the crystallized ligand in 3AW0 in the two active sites of the homodimer (protomers A and B). The peptidomimetic Ac-Ser-Ala-Val-Leu-His-aldehyde inhibitor was elongated using the Maestro tool¹⁵ until we built the substrate-like sequence Ac-Ser-Ala-Val-Leu-Gln-Ser-Gly-Phe-NMe. We placed the substrate in the two active sites, in agreement with the observation that the X-ray structures (6LZE, 6M0K, 6WNP, 6LU7, and 7BQY) contain inhibitor molecules in both of them. The absent hydrogen atoms were added using the Protein Preparation Wizard tool of Maestro, and PROPKA3.0¹⁶ was used to calculate the protonation states of titratable residues at pH 7.4.

The tleap tool from AmberTools18¹⁷ was used to prepare the simulation systems. The Michaelis complexes, described with the ff14SB force field,¹⁸ were solvated into a box with a buffer region of at least 12 Å from any protein/substrate atom to the limits of the simulation box. TIP3P water molecules¹⁹ were used. Na⁺ atoms were added to neutralize the charge. The resulting system was minimized using 500 steps of the steepest descent method followed by the conjugate-gradient method, until the root-mean-square of the gradient was below 10⁻³ kcal mol⁻¹ Å⁻¹. The system was then heated from 0 to 300 K using a heating rate of 1.7 K-ps⁻¹. The backbone heavy atoms were

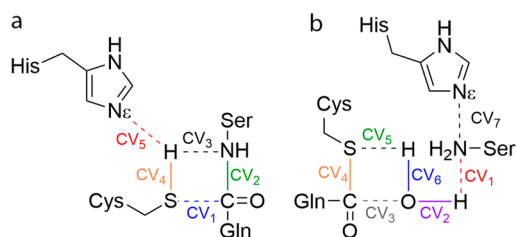
restrained in a Cartesian space using a harmonic potential with a force constant of 20 kcal mol⁻¹ Å⁻². Along the equilibration step, the positional restraint force constant was changed from 15 to 3 kcal mol⁻¹ Å⁻², decreasing by 3 units every 1.25 ns, and after 6.25 ns the positional restraints were removed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

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

electronic description in agreement with higher-level methods.²⁹ Finally, a recent QM/MM analysis of the electronic properties of enzyme–substrate complexes in SARS-CoV-2 protease emphasizes the importance of choosing an adequate functional for a proper description of the process.³⁰ All calculations were run with a modified version of Amber18^{17,31} coupled to Gaussian16³² for DFT calculations. A cutoff radius of 15 Å was used for all QM/MM interactions, and the temperature was 300 K.

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

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



Sy(Cys145)-C(P1), H_{w2}-O_w, N(P1')-N ϵ (His41), and H_{w2}-Sy(Cys145) for the deacylation step (see Scheme 2b). Different initial guesses, corresponding to different mechanistic proposals, were explored at the B3LYPD3/MM level using first the 6-31G* basis set, and then the best candidates were recalculated using the 6-31+G* basis set. Each of the strings was composed of at least 96 nodes (see Table S1) that were propagated with a time step of 1 fs until the RMSD of the string fell below 0.1 amu^{1/2}·Å for at least 2 ps. The converged MFEPs are then averaged to define the s path-CV corresponding to each string. The free energy profiles along the path-CVs were obtained using an umbrella sampling algorithm³⁴ running simulations for at least 10 ps and were integrated using the WHAM technique.³⁵ The values of the force constants employed to bias the ASM simulations were determined on-the-fly to ensure a probability density distribution of the reaction coordinate that was as homogeneous as possible.³³ Replica exchange between neighboring string nodes was attempted every 50 steps to improve convergence.

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

the free energy profile using umbrella sampling³⁴ along a simple proton-transfer coordinate defined as the antisymmetric combination of the distances of the proton to the donor and the acceptor atoms ($d(\text{N}\epsilon\text{-H}) - d(\text{S}\gamma\text{-H})$). For this profile only the side chains of the two involved residues were included in the QM region (using the B3LYPD3/6-31+G* level of theory). A total of 40 windows were used, corresponding to an increment of the reaction coordinate of 0.06 Å, with each of them composed of 10 ps of equilibration and 20 ps of data collection. The force constant employed to drive the reaction coordinate change was 600 kcal·mol⁻¹·Å⁻². All the rest of the details of the simulations were as described earlier. We also used umbrella sampling along distinguished coordinates to explore the free energy landscape associated with the separation between the first peptide fragment and the acyl-enzyme complex and between the two peptide fragments at the end of the deacylation process. Details of all the free energy simulations performed in this work are given in Table S1.

3. RESULTS

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

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

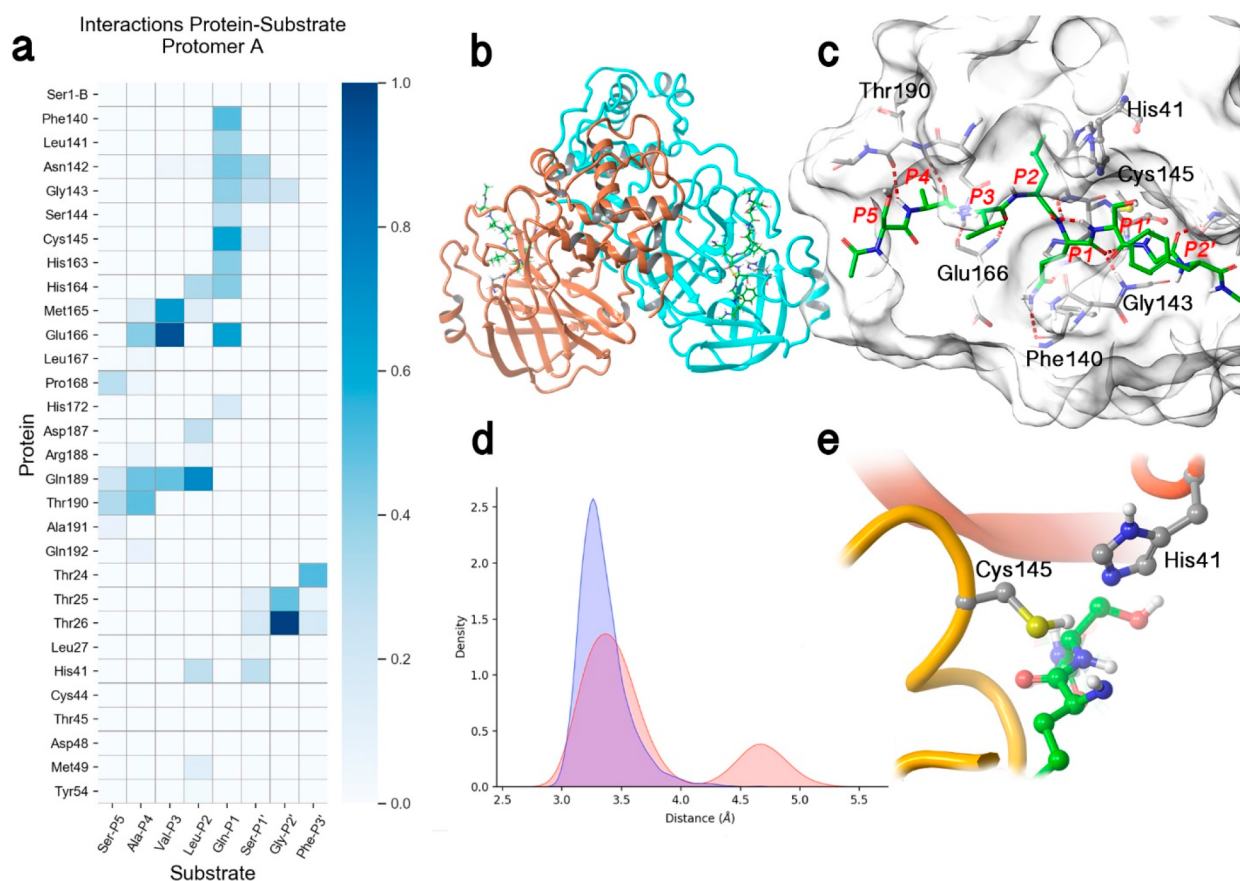


Figure 1. Results of the molecular dynamics simulation of 3CL^{pro} of SARS-CoV-2 in complex with the substrate with sequence Ac-Ser-Ala-Val-Leu-Gln-Ser-Gly-Phe-NMe. (a) Fraction of hydrogen-bond contacts between the residues of the substrate and those of the protease found during the trajectory of the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

this residue is placed between the side chains of Met165, Leu167, and Pro168. The S5 subsite is formed by the side chain of Pro168 and by main-chain atoms of Thr190. This description of the S1–S5 interaction subsites observed in our simulations agrees with the characteristics found in the X-ray structures of SARS-CoV-2 3CL protease with inhibitors bound in the active site.^{6,8} Our MD simulations of the substrate–enzyme complex offer, in addition, a detailed description of the S' subsites, those that accommodate the P' residues placed after the scissile peptide bond. The main-chain O atom of Ser-P1' establishes a hydrogen bond with the amide group of Gly143 and the side chain of Asn142. The hydroxyl group of the P1' side chain can contact with the catalytic dyad (Cys145 and His41), while the CH₂ group is packed between Thr25, Thr26, and Leu27. Gly-P2' is stabilized through main-chain contacts with Thr25 and Thr26. Finally, the side chain of the Phe-P3' residue is packed against the side chain of Thr24. This structural information can be useful in order to improve the binding and specificity of potential inhibitors of the protease activity because these structural findings are lost in the X-ray structures obtained from those inhibitors in which the

fragment corresponding to P' residues either is released during the formation of the acyl–enzyme complex¹³ or is smaller than in our substrate.^{6,8}

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

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

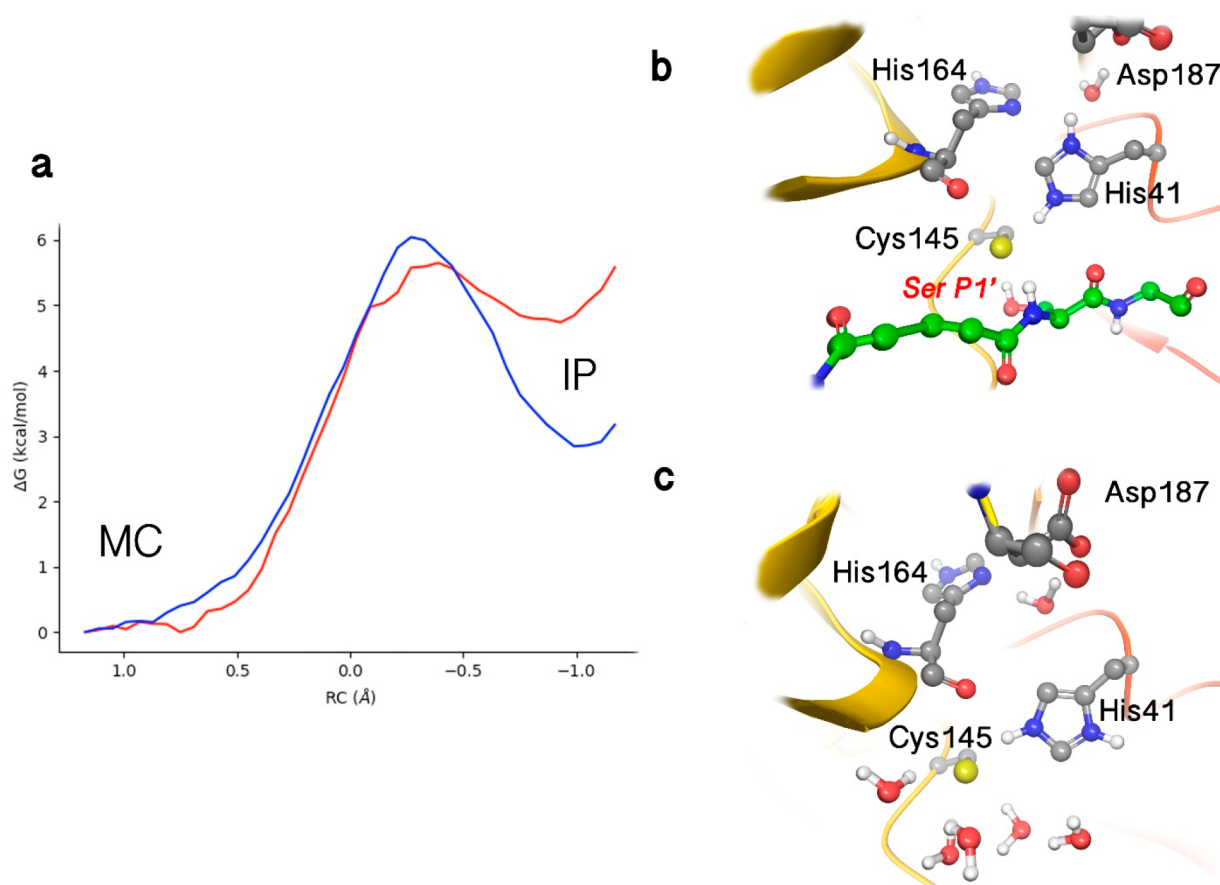


Figure 2. Analysis of the formation of an ion pair (IP) catalytic dyad ($\text{Cys145}^- \cdots \text{His41H}^+$) from the neutral form ($\text{Cys145H} \cdots \text{His41}$) found in the Michaelis complex (MC). (a) B3LYPD3/6-31+G*/MM free energy profile associated with the proton-transfer coordinate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

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

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

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

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

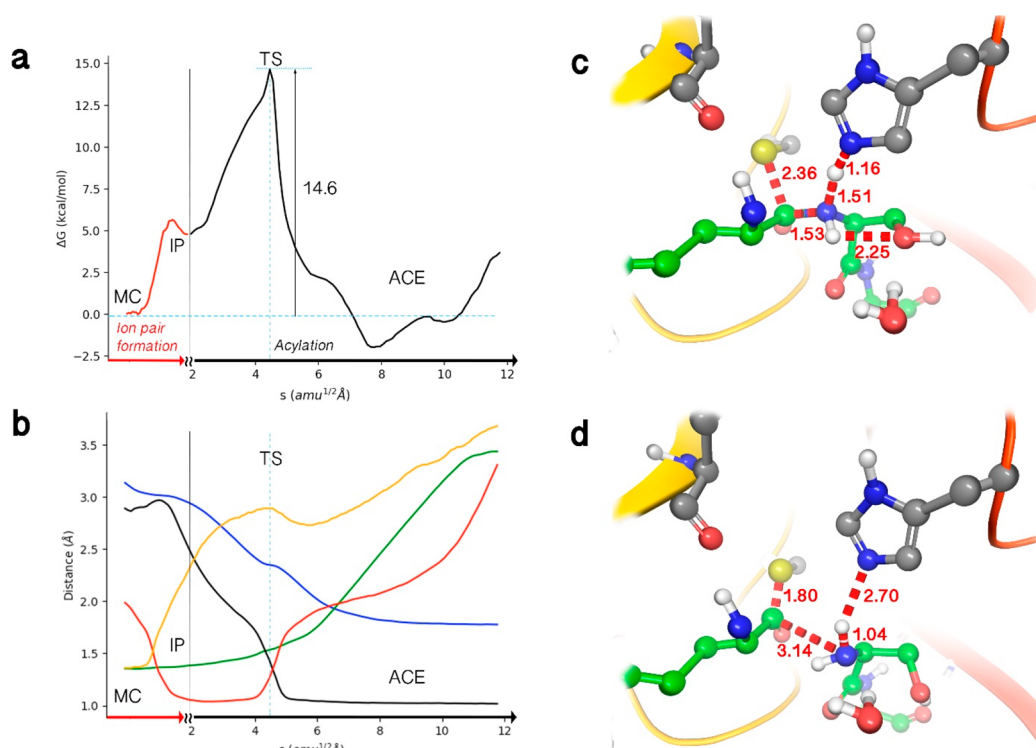


Figure 3. Simulation of the acylation reaction taking place through the formation of the ion pair. (a) B3LYPD3/6-31+G*/MM free energy profile along the path-CV for the acylation reaction after the formation of the ion pair (IP) from the Michaelis complex (MC). The reaction takes place with a single transition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

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

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

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

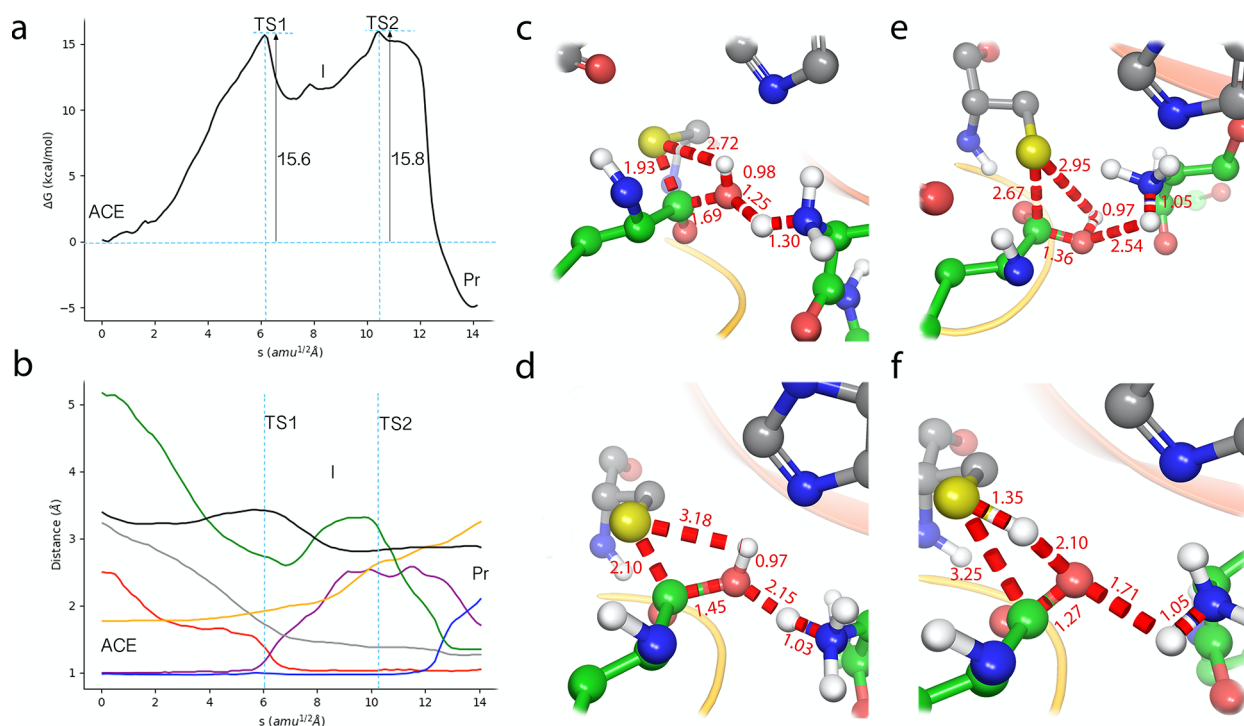


Figure 4. Simulation of the deacylation reaction with the N-terminal group of the P' fragment acting as a general base in charge of water activation. (a) B3LYPD3/6-31+G*/MM free energy profile along the path-CV for the deacylation reaction. The process takes place in two steps. In the first one the water activated by the N-terminal group attacks the acyl–enzyme complex (ACE) to form a thiodiolate intermediate (I) through the first TS (TS1). In the second one the reaction product (Pr) is obtained after breaking the acyl–enzyme bond in the second TS (TS2). (b) Evolution of the distances selected as collective variables along the minimum free energy path. $Sy-H_{w2}$ is in green, $Ne-N(P1')$ is in black, $O_w-C(P1)$ is in gray, $N(P1')-H_{w1}$ is in red, $Sy-C(P1)$ is in yellow, O_w-H_{w1} is in purple, and O_w-H_{w2} is in blue. (c) Representation of TS1 where the water molecule is transferring a proton to the N-terminal group of the P' fragment and the resulting hydroxyl anion attacks the carbonyl carbon atom of the P fragment (C(P)). The values of the distances correspond to the coordinates of the MFEP at TS1. (d) Representation of the thiodiolate intermediate (I). (e) Representation of TS2 corresponding to the breaking of the $Sy-C(P1)$ bond and the proton transfer from the carboxylic terminal group to the leaving sulfur atom. (f) Representation of the reaction products (Pr) with the $P-COO^-$ and $P'-NH_3^+$ peptide fragments in the active site of the protease.

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

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

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

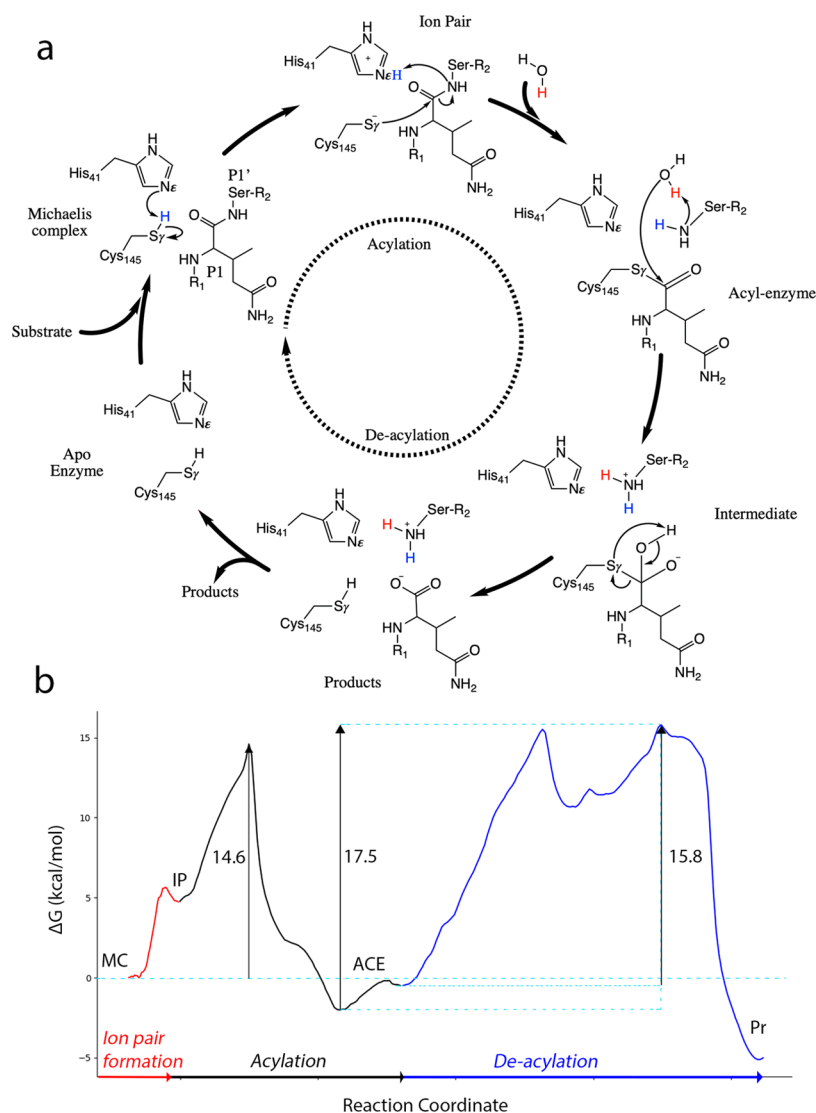


Figure 5. Proteolysis mechanism in 3CL^{PRO} of SARS-CoV-2 as deduced from our simulations. (a) Schematic representation of the reaction mechanism. (b) Free energy profile associated with the reaction mechanism.

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

3.4. Deacylation Step. Regarding the deacylation step, the standard mechanistic proposal suggested for related enzymes⁴⁷ assumes that, once the neutral P'-NH₂ peptide fragment has left the active site, a water molecule activated by His41 attacks the C(P1)-S γ bond, releasing the P-COOH peptide (also with a neutral terminal group) and, finally, regenerating the enzyme after a proton transfer from His41 to Cys145 (see Scheme 1). In our simulations of the acylation product, we found that a water molecule can be placed in between the P'-NH₂ leaving fragment and the acyl-enzyme complex, being correctly oriented to perform the hydrolysis of the acyl-enzyme (see Figures 3d and S5). This configuration suggests an alternative reaction mechanism that can yield the two peptide fragments with correct protonation states in the terminal groups and regenerate the enzymatic active site in its most stable state (the neutral catalytic dyad). An additional advantage of this novel mechanism is that it involves water activation by means of the N-terminus of the P' fragment, which is known to be a better base than histidine side chains (the average pK_a values are

about 7.7 and 6.6, respectively).⁴⁸ Finally, the proposed mechanism can be at the origin of the mechanistic differences observed when the scissile bond is an amide instead of an ester (in that case a basic N-terminal group is not formed after the acylation).⁴¹ We obtained the MFEP corresponding to such a mechanism at the B3LYPD3/MM level (see Figure 4). This process is stepwise, presenting two TSs (see Figure 4a). The first TS (TS1 in Figure 4c) corresponds to the proton transfer from the water molecule to the N(P1') atom, resulting in the formation of the P'-NH₃⁺ peptide fragment. The free energy barrier associated with this step is 15.6 kcal·mol⁻¹. This proton transfer is concomitant to the attack of the hydroxyl group on the C(P1) carbonyl carbon atom, resulting in the formation of an intermediate thiodiolate (Figure 4d). After rotation of the hydroxyl group to orient the proton toward the sulfur atom, the reaction proceeds with the cleavage of the C(P1)-S γ bond. The second TS observed during the deacylation (TS2, Figure 4e) corresponds to the separation of the S γ atom (with the C(P1)-S γ distance being 2.67 Å). The free energy barrier associated with this second step from the acyl-enzyme complex is very close to the first one, 15.8 kcal·mol⁻¹ (17.5 kcal·mol⁻¹ from the most stable acyl-enzyme conformer; see

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

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

4. CONCLUSIONS

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

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

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

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

■ ASSOCIATED CONTENT

Supporting Information

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

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

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

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

■ AUTHOR INFORMATION

Corresponding Authors

J. Javier Ruiz-Pernía – Departamento de Química Física, Universidad de Valencia, 46100 Burjassot, Spain; orcid.org/0000-0002-4640-0419; Email: jjavier.ruiz@uv.es

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

Author

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

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

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

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

REFERENCES

- (1) Zhou, P.; Yang, X.-L.; Wang, X.-G.; Hu, B.; Zhang, L.; Zhang, W.; Si, H.-R.; Zhu, Y.; Li, B.; Huang, C.-L.; Chen, H.-D.; Chen, J.; Luo, Y.; Guo, H.; Jiang, R.-D.; Liu, M.-Q.; Chen, Y.; Shen, X.-R.; Wang, X.; Zheng, X.-S.; Zhao, K.; Chen, Q.-J.; Deng, F.; Liu, L.-L.; Yan, B.; Zhan, F.-X.; Wang, Y.-Y.; Xiao, G.-F.; Shi, Z.-L. A Pneumonia Outbreak Associated with a New Coronavirus of Probable Bat Origin. *Nature* **2020**, *579*, 270–273.
- (2) Lau, S. K. P.; Woo, P. C. Y.; Li, K. S. M.; Huang, Y.; Tsoi, H.-W.; Wong, B. H. L.; Wong, S. S. Y.; Leung, S.-Y.; Chan, K.-H.; Yuen, K.-Y. Severe Acute Respiratory Syndrome Coronavirus-like Virus in Chinese Horseshoe Bats. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 14040–14045.
- (3) Cheng, V. C. C.; Lau, S. K. P.; Woo, P. C. Y.; Yuen, K. Y. Severe Acute Respiratory Syndrome Coronavirus as an Agent of Emerging and Reemerging Infection. *Clin. Microbiol. Rev.* **2007**, *20*, 660–694.
- (4) Pillaiyar, T.; Manickam, M.; Namasivayam, V.; Hayashi, Y.; Jung, S.-H. An Overview of Severe Acute Respiratory Syndrome–Coronavirus (SARS-CoV) 3CL Protease Inhibitors: Peptidomimetics and Small Molecule Chemotherapy. *J. Med. Chem.* **2016**, *59*, 6595–6628.
- (5) Bangham, C. R. M. The Immune Control and Cell-to-Cell Spread of Human T-Lymphotropic Virus Type 1. *J. Gen. Virol.* **2003**, *84*, 3177–3189.
- (6) Jin, Z.; Du, X.; Xu, Y.; Deng, Y.; Liu, M.; Zhao, Y.; Zhang, B.; Li, X.; Zhang, L.; Peng, C.; Duan, Y.; Yu, J.; Wang, L.; Yang, K.; Liu, F.; Jiang, R.; Yang, X. X.; You, T.; Liu, X. X.; Yang, X. X.; Bai, F.; Liu, H.; Liu, X. X.; Guddat, L. W.; Xu, W.; Xiao, G.; Qin, C.; Shi, Z.; Jiang, H.; Rao, Z.; Yang, H. Structure of Mpro from COVID-19 Virus and Discovery of Its Inhibitors. *Nature* **2020**, *582*, 289–293.
- (7) Zhang, L.; Lin, D.; Sun, X.; Curth, U.; Drosten, C.; Sauerhering, L.; Becker, S.; Rox, K.; Hilgenfeld, R. Crystal Structure of SARS-CoV-2 Main Protease Provides a Basis for Design of Improved α -Ketoamide Inhibitors. *Science* **2020**, *368*, 409–412.
- (8) Dai, W.; Zhang, B.; Jiang, X.-M.; Su, H.; Li, J.; Zhao, Y.; Xie, X.; Jin, Z.; Peng, J.; Liu, F.; Li, C.; Li, Y.; Bai, F.; Wang, H.; Cheng, X.; Cen, X.; Hu, S.; Yang, X.; Wang, J.; Liu, X.; Xiao, G.; Jiang, H.; Rao, Z.; Zhang, L.-K.; Xu, Y.; Yang, H.; Liu, H. Structure-Based Design of Antiviral Drug Candidates Targeting the SARS-CoV-2 Main Protease. *Science* **2020**, *368*, 1331–1335.
- (9) Chen, Y. W.; Yiu, C.-P. B.; Wong, K.-Y. Prediction of the SARS-CoV-2 (2019-nCoV) 3C-like Protease (3CLpro) Structure: Virtual Screening Reveals Velpatasvir, Ledipasvir, and Other Drug Repurposing Candidates. *FI1000Research* **2020**, *9*, 129.
- (10) Rut, W.; Groborz, K.; Zhang, L.; Sun, X.; Zmudzinski, M.; Pawlik, B.; Mlynarski, W.; Hilgenfeld, R.; Drag, M. Substrate Specificity Profiling of SARS-CoV-2 Main Protease Enables Design of Activity-Based Probes for Patient-Sample Imaging. *bioRxiv* **2020**, DOI: 10.1101/2020.03.07.981928.
- (11) Hilgenfeld, R. From SARS to MERS: Crystallographic Studies on Coronaviral Proteases Enable Antiviral Drug Design. *FEBS J.* **2014**, *281*, 4085–4096.
- (12) Brocklehurst, K.; Willenbrock, F.; Sauh, E. Cysteine Proteinases. In *Hydrolytic Enzymes*; Elsevier: New York, 1987; pp 39–158; DOI: 10.1016/S0167-7306(09)60016-8.
- (13) Jin, Z.; Zhao, Y.; Sun, Y.; Zhang, B.; Wang, H.; Wu, Y.; Zhu, Y.; Zhu, C.; Hu, T.; Du, X.; Duan, Y.; Yu, J.; Yang, X.; Yang, X.; Yang, K.; Liu, X.; Guddat, L. W.; Xiao, G.; Zhang, L.; Yang, H.; Rao, Z. Structural Basis for the Inhibition of SARS-CoV-2 Main Protease by Antineoplastic Drug Carmofur. *Nat. Struct. Mol. Biol.* **2020**, *27*, 529–532.
- (14) Akaji, K.; Konno, H.; Mitsui, H.; Teruya, K.; Shimamoto, Y.; Hattori, Y.; Ozaki, T.; Kusunoki, M.; Sanjoh, A. Structure-Based Design, Synthesis, and Evaluation of Peptide-Mimetic SARS 3CL Protease Inhibitors. *J. Med. Chem.* **2011**, *54*, 7962–7973.
- (15) *Schrödinger Release 2016-3: Maestro*; Schrödinger, LLC: New York, 2016.
- (16) Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical PKa Predictions. *J. Chem. Theory Comput.* **2011**, *7*, 525–537.
- (17) Case, D. A.; Cerutti, D. S.; Cheatham, T. E. I.; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Greene, D.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D.; Merz, K. M.; Monard, G.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Simmerling, C. L.; Bötolo-Smith, W. M.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Xiao, L.; York, D. M.; Kollman, P. A. *Amber 2018*; University of California: San Francisco, 2018.
- (18) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.
- (19) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926.
- (20) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.
- (21) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N-log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (22) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.
- (23) Le Grand, S.; Götz, A. W.; Walker, R. C. SPFP: Speed without Compromise - A Mixed Precision Model for GPU Accelerated Molecular Dynamics Simulations. *Comput. Phys. Commun.* **2013**, *184*, 374–380.
- (24) Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9*, 3878–3888.
- (25) Becke, A. D. Density-functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, S648–S652.
- (26) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785–789.
- (27) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion

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

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

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

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

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

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

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

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

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

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

(37) Yang, H.; Yang, M.; Ding, Y.; Liu, Y.; Lou, Z.; Zhou, Z.; Sun, L.; Mo, L.; Ye, S.; Pang, H.; Gao, G. F.; Anand, K.; Bartlam, M.; Hilgenfeld, R.; Rao, Z. The Crystal Structures of Severe Acute Respiratory Syndrome Virus Main Protease and Its Complex with an Inhibitor. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 13190.

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

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

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

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

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

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

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

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

(45) Hsu, W.-C.; Chang, H.-C.; Chou, C.-Y.; Tsai, P.-J.; Lin, P.-I.; Chang, G.-G. Critical Assessment of Important Regions in the Subunit Association and Catalytic Action of the Severe Acute Respiratory Syndrome Coronavirus Main Protease. *J. Biol. Chem.* **2005**, *280*, 22741–22748.

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

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

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

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

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

Supporting Information

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

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

Figure S1. Example of string convergence	S2
Figure S2. Plot of the RMSD values.	S3
Figure S3. Comparison of substrate-enzyme interactions in both protomers	S4
Figure S4. Distribution of S-C distances in the dimer with one and two substrates	S5
Figure S5. Analysis of conformations of the acyl-enzyme (ACE) complex	S6
Figure S6. Simulation of the acylation reaction with different QM descriptions	S7
Figure S7. Simulation of the acylation reaction without formation of an ion pair	S8
Figure S8. Deacylation reaction with His41 as general base	S9
Table S1. Details of free energy calculations	S10
Section S1. Evaluation of the Energy Difference between neutral and ion-pair dyad	S11
Table S2. Absolute enthalpies for the gas phase model at B3LYPD3/6-31+G(d) level	S11
References.	S12

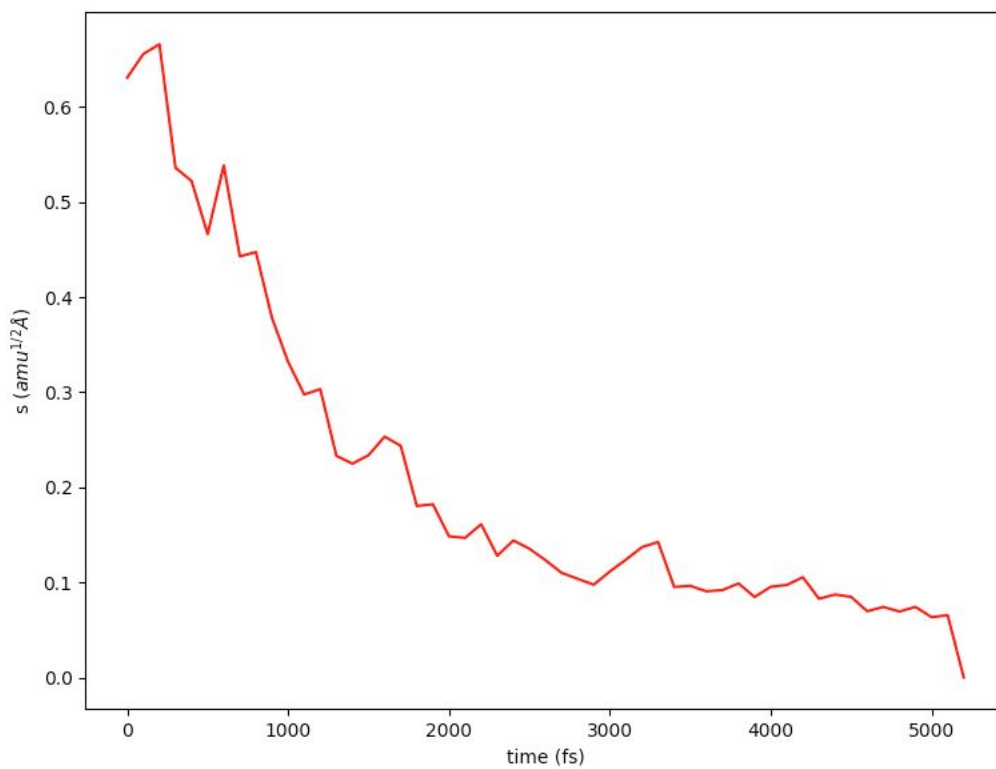


Figure S1. Example of string convergence showing the RMSD with respect to the final string nodes along the equilibration of the minimum free energy path.

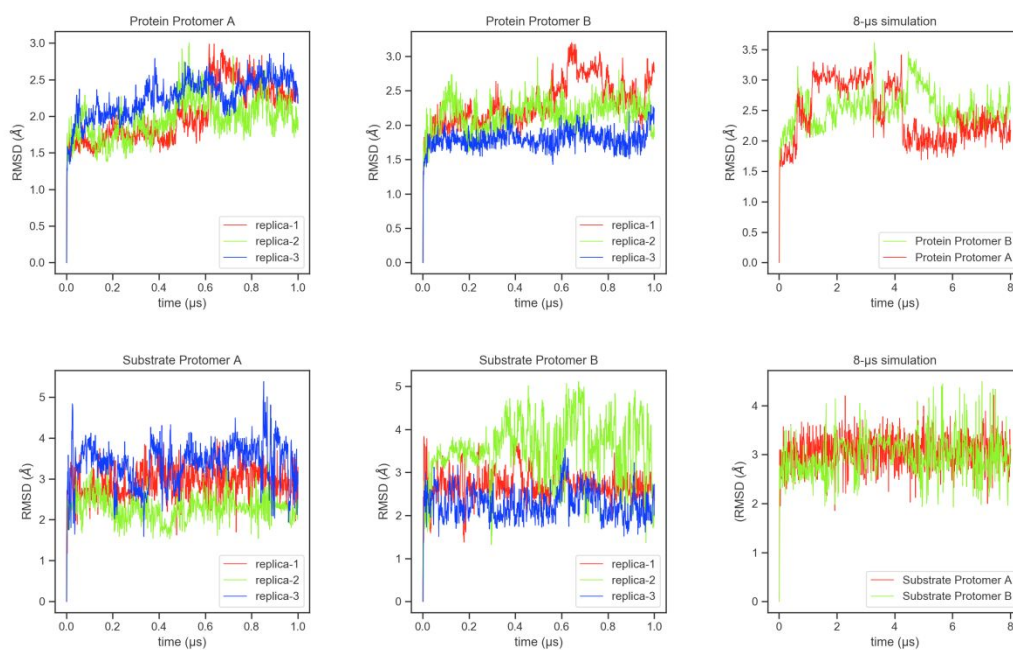


Figure S2. Plot of the RMSD values obtained for the protein (up) and the substrate (down) during the 3 replicas (2 of 1 μs and one of 8 μs) of the Michaelis Complex of the main protease with a peptidic substrate. From left to right, RMSD of protomer A, protomer B and RMSD obtained during the 8 μs -long simulation. The RMSD has been computed using all non-hydrogen atoms taking as reference the initial structure prepared for the simulations.

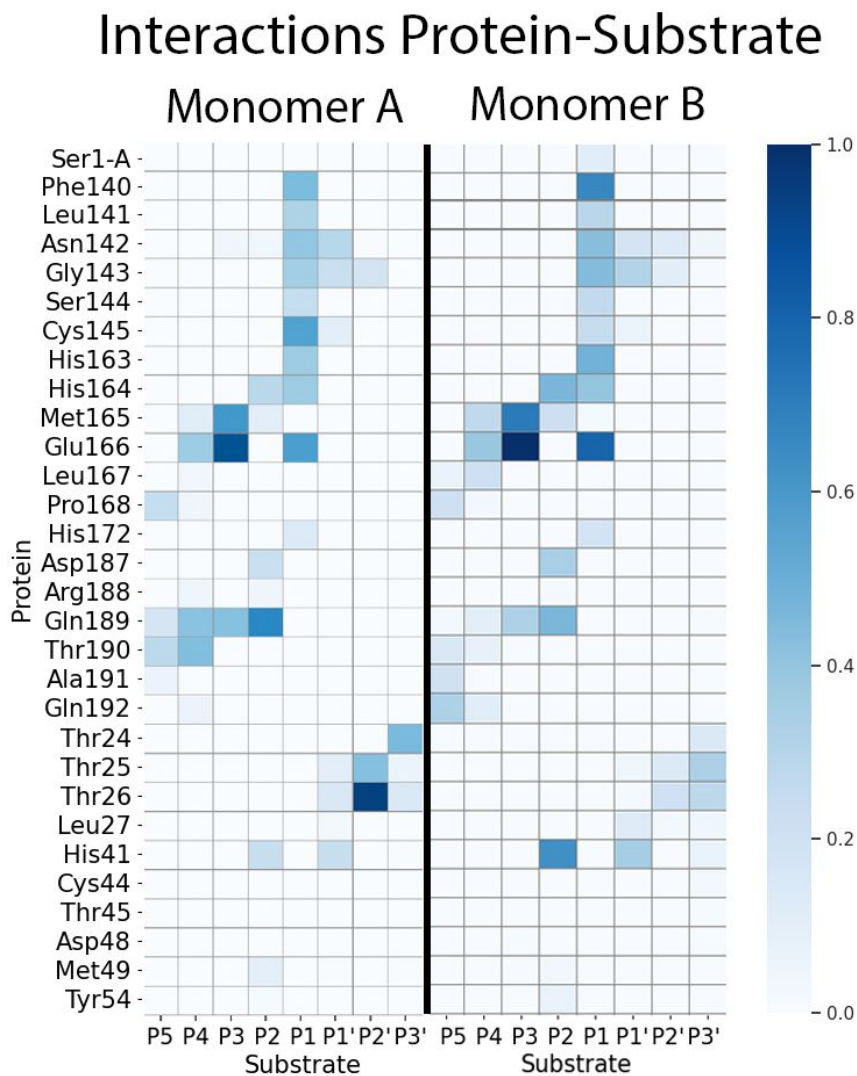


Figure S3. Fraction of hydrogen bond contacts between the residues of the substrate and those of the protease found during the μ s trajectory of the Michaelis Complex in protomers A and B. A hydrogen bond contact is counted when the donor-acceptor distance is $< 3.8 \text{ \AA}$ and the hydrogen bond angle is $> 120^\circ$.

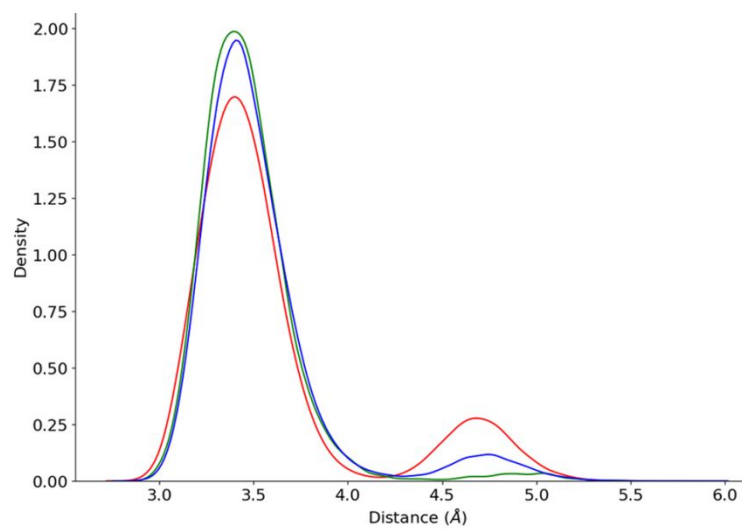


Figure S4. Probability densities of the distances from the Cys145-S γ atom to the carbonyl carbon atom of the substrate (C(P1)), in monomer A (red), monomer B (blue) and in a 1 μ s simulation with only monomer A occupied (green).

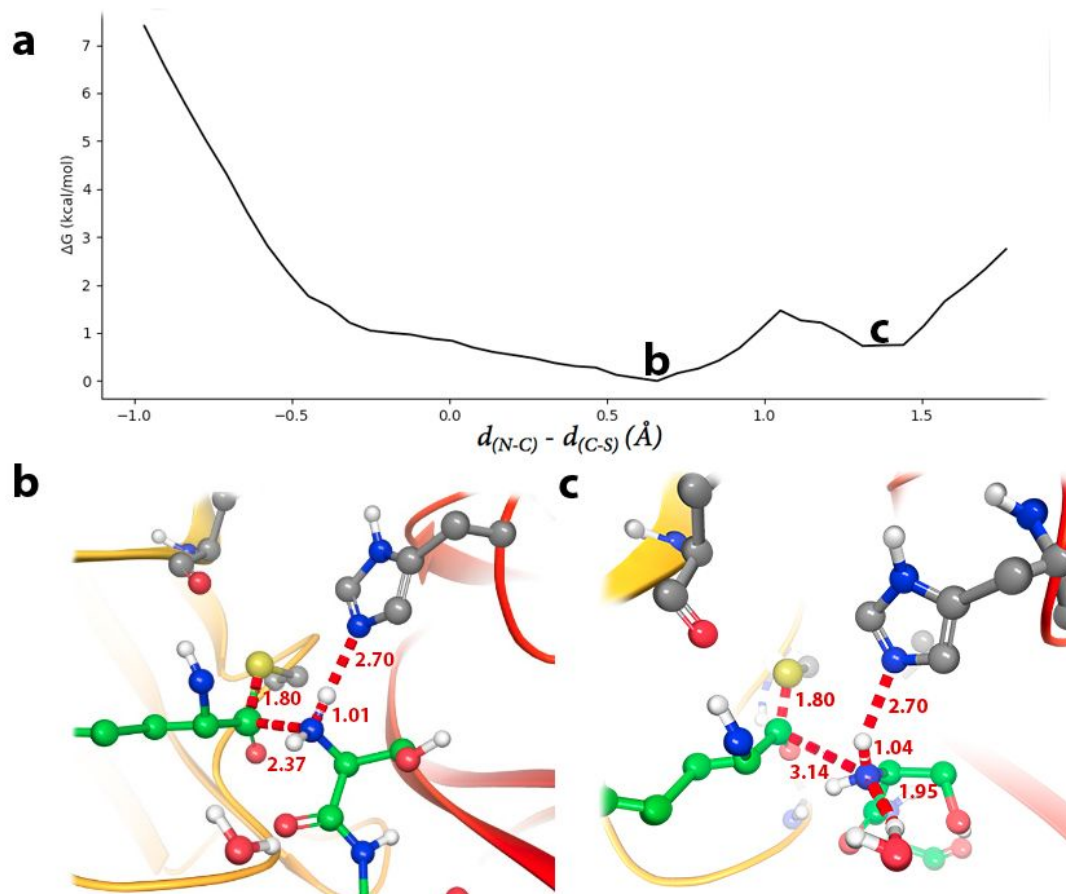


Figure S5. Analysis of conformations of the acyl-enzyme (ACE) complex as a function of the distance to the peptide fragment formed after acylation. **(S5a)** B3LYP/6-31+G*/MM free energy profile along the distinguished coordinate $d(\text{C}(\text{P1})-\text{N}(\text{P1}') - d(\text{C}(\text{P1})-\text{S}_\gamma)$. This profile matches the final part of the profile showed in Figure 3a, with two minima (presented in 1b and 1c) that correspond to the acyl-enzyme complex with increased distance to the peptide fragment formed during the acylation step. This free energy profile also confirms that once the nitrogen atom of the scissile bond is protonated by His41 the nucleophilic attack of the S_γ atom to the carbonyl carbon atom is barrierless. The small barrier appearing between the two minima is due to the inversion of the N-terminal group of the P' peptide. **(S5b)** Representation of the first minimum appearing in the free energy profile. The lone pair of the N-terminal group is pointing toward the carbonyl carbon atom of the acyl-enzyme ($\text{C}(\text{P1})$) and the distance between these two atoms is about 2.4 Å. **(S5c)** Representation of the second minimum appearing in the free energy profile. The lone pair of the N-terminal group is now pointing toward the solvent and the distance between $\text{N}(\text{P1}')$ and $\text{C}(\text{P1})$ atoms is about 3.1 Å. A water molecule hydrogen bonded to the N-terminal group perform the hydrolysis during the de-acylation reaction.

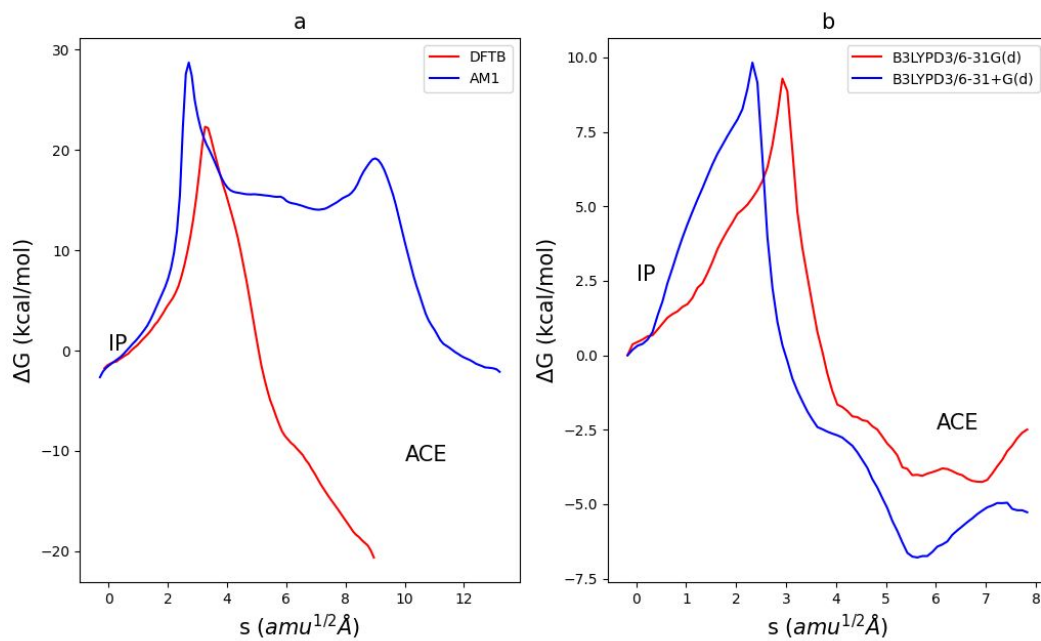


Figure S6. Simulation of the acylation reaction from the ion pair at different theoretical levels. (**S6a**) Free energy profile obtained at AM1/MM and DFTB/MM levels. Note the differences in the concerted/stepwise nature of the process. (**S6b**) B3LYPD3MM free energy profiles using two different basis sets, 6-31G* and 6-31+G*.

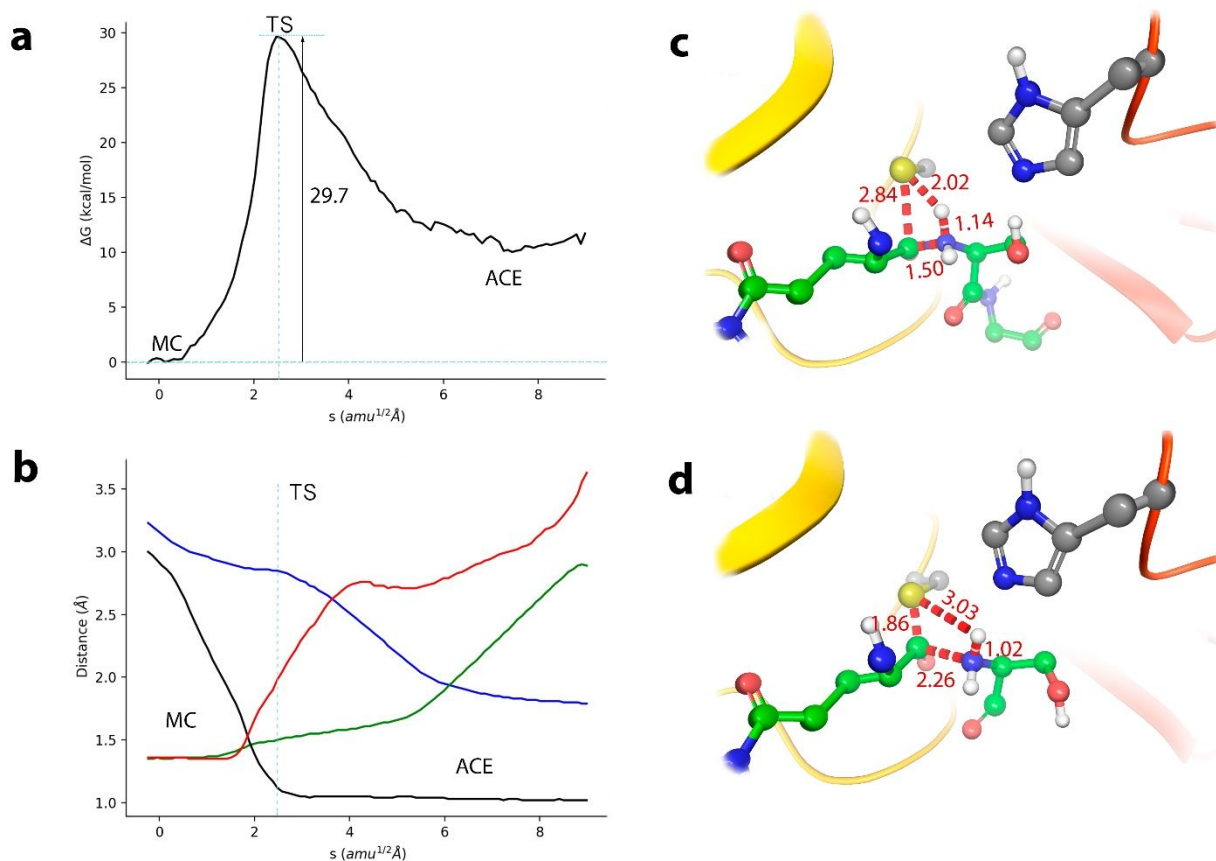


Figure S7. Simulation of the acylation reaction taking place without formation of an ion pair. (**S7a**) B3LYPD3/6-31G*/MM free energy profile along the path-CV for the acylation reaction from the Michaelis complex (MC). The reaction takes place with a single transition state (TS) that yield the acyl-enzyme (ACE). (**S7b**) Evolution of the distances selected as collective variables (CV) along the Minimum Free Energy Path (MFEP). S_{γ} -H in red, C(P1)-N(P1') (the scissile peptide bond) in green, H-N(P1') in black, S S_{γ} -C(P1) in blue. (**S7c**) Representation of the TS for the acylation process through this mechanism. This TS corresponds to the proton transfer from Cys145 to the nitrogen atom of peptide bond (N(P1')) with the approach of the S_{γ} atom to the carbonyl carbon atom (C(P1)) and the lengthening of the peptide bond. The values of the distances correspond to the coordinates of the MFEP at the TS. (**S7d**) Representation of the acyl-enzyme complex (ACE) obtained with indication of the main distances obtained along the MFEP.

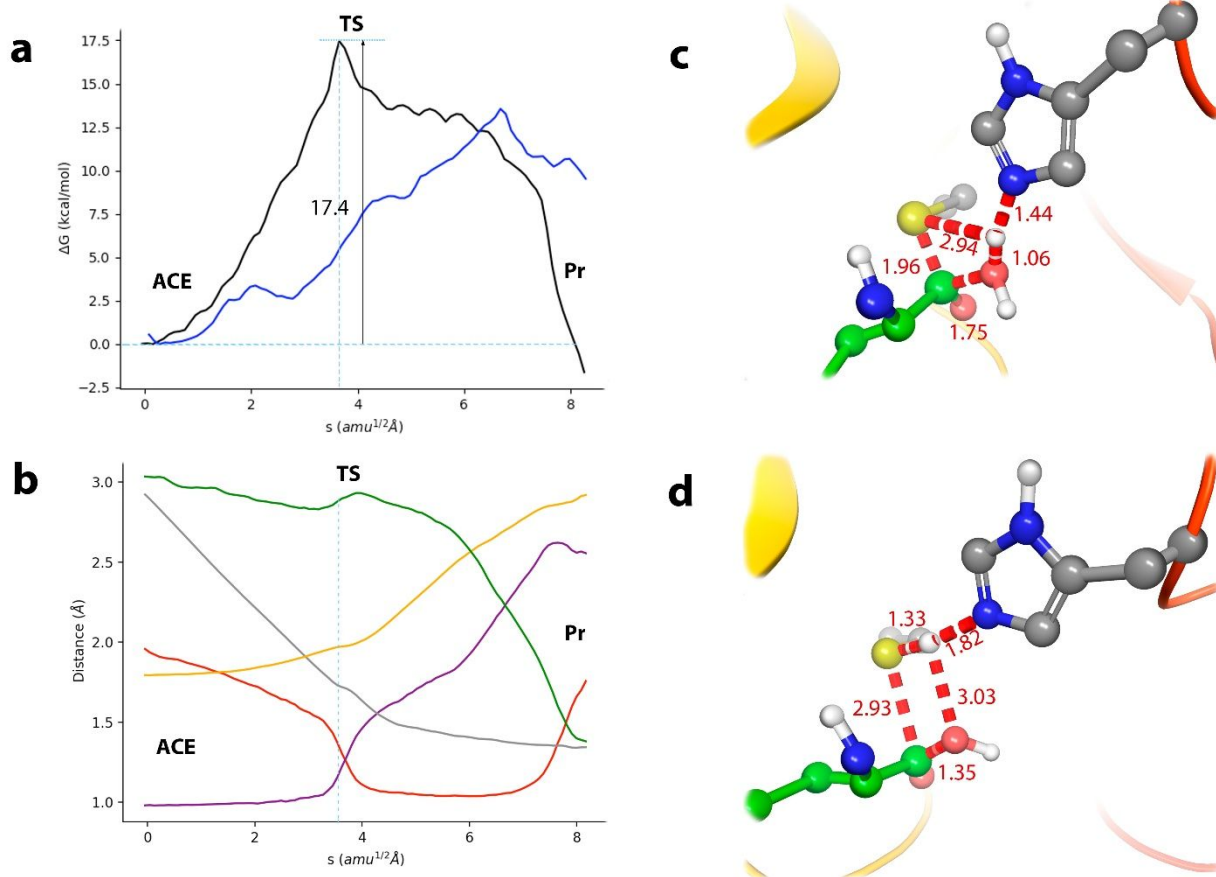


Figure S8. Simulation of the hydrolysis reaction when the general base activating the water molecule is His41. **(S8a)** B3LYPD3/6-31G*/MM free energy profiles along the path-CV for the de-acylation reaction when the general base is His41 (black line) or the N-terminal group of the P' peptide fragment (blue line). Note that in the second case the activation free energy is smaller, in agreement with the trend expected from the pK_a of the respective bases (see main text). **(S8b)** Evolution of the distances selected as collective variables along the Minimum Free Energy Path (MFEP) for the hydrolysis reaction when His41 acts as the base. O_w-H_w in purple, $H_w-N\epsilon(\text{His41})$ in red, $S_\gamma-C(\text{P1})$ in yellow, $O_w-C(\text{P1})$ in grey and $S_\gamma-H_w$ in green. **(S8c)** Representation of the transition state (TS) for the de-acylation process through this mechanism. This TS corresponds to the proton transfer from water to the nitrogen atom of His41 with the approach of the hydroxyl group to the carbonyl carbon atom of the acyl-enzyme. The values of the distances correspond to the coordinates of the MFEP at the TS. **(S8d)** Representation of the final product obtained through the proposed mechanism. Note that in this mechanism the peptide fragment appears with the C-terminal group protonated.

Table S1. Details about the Free Energy calculations performed in this work

Description	Type of Simulation*	Coordinate	QM region [‡]	QM level	# String Nodes or US Windows	Eq. time (ps)	Prod. time (ps)
Cys145-His41 Ion Pair formation	US	d(N ϵ -H)- d(S γ -H)	Cys145 sc + His41 sc	B3LYPD3/6-31+G*	40	10	20
Acylation through Ion Pair	ASM + US	Path-CV (6 CVs)	Cys145 sc + His41 sc + P1 + P1' + npb	B3LYPD3/6-31G*	96	6	15
Acylation through Ion Pair	ASM + US	Path-CV (6 CVs)	Cys145 sc + His41 sc + P1 + P1' + npb	B3LYPD3/6-31+G*	128	3 (from 6-31G* path)	11.5
Acylation through Ion Pair	US	d(N ϵ -H)- d(N(P1')-H)	Cys145 sc + His41 sc + P1 + P1' + npb	B3LYPD3/6-31+G*	80	5	10
Acyl-enzyme Conformations	US	d(S γ -C(P1))- d(N(P1')-C(P1))	Cys145 sc + His41 sc + P1 + P1' + npb	B3LYPD3/6-31+G*	60	10	10
Acylation without Ion Pair	ASM + US	Path-CV (6 CVs)	Cys145 sc + His41 sc + P1 + P1' + npb	B3LYPD3/6-31G*	96	6	15
De-acylation (general base: N-terminus)	ASM+US	Path-CV (7 CVs)	Cys145 sc + His41 sc + P1 + P1' + npb + H ₂ O	B3LYPD3/6-31G*	96	6	11
De-acylation (general base: N-terminus)	ASM+US	Path-CV (7 CVs)	Cys145 sc + His41 sc + P1 + P1' + npb + H ₂ O	B3LYPD3/6-31+G*	96	3 (from 6-31G* path)	10
De-acylation (general base :His41)	ASM+US	Path-CV (6 CVs)	Cys145 sc + His41 sc + P1 + npb + H ₂ O	B3LYPD3/6-31G*	96	5	10

*US stands for Umbrella Sampling and ASM for Adaptive String Methods

[‡]sc stands for side chain and npb for next peptide bonds (to P1 and P1')

Section S1. Evaluation of the Energy Difference between the neutral and ion-pair catalytic dyad.

We used Gaussian16 to calibrate the computational level selected to evaluate the proton transfer between Cys145 and His41. With this purpose we calculated the reaction enthalpy for the reaction between Imidazol (Im) and Methanethiol to yield Imidazloium ion and Methanethiolate:

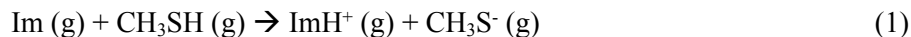


Table S2 shows the absolute enthalpies obtained for each of the species appearing in (1) calculated at 25° C and 1 atm using the standard rigid-rotor and harmonic oscillator approximations calculated with the B3LYP functional and the 6-31+G(d) basis set and including also D3 dispersion corrections.

Table S2. Absolute enthalpies (in Hartrees) for chemical species appearing in (1) and reaction enthalpy (in kcal·mol⁻¹) calculated at 25° C and 1 atm at the B3LYPD3/6-31+G(d) level.

Im	CH ₃ SH	ImH ⁺	CH ₃ S ⁻	ΔH _{r,calc}	ΔH _{r,exp}
-226.153903	-438.651641	-226.509775	-438.090464	128.8	131.5 ± 2.2

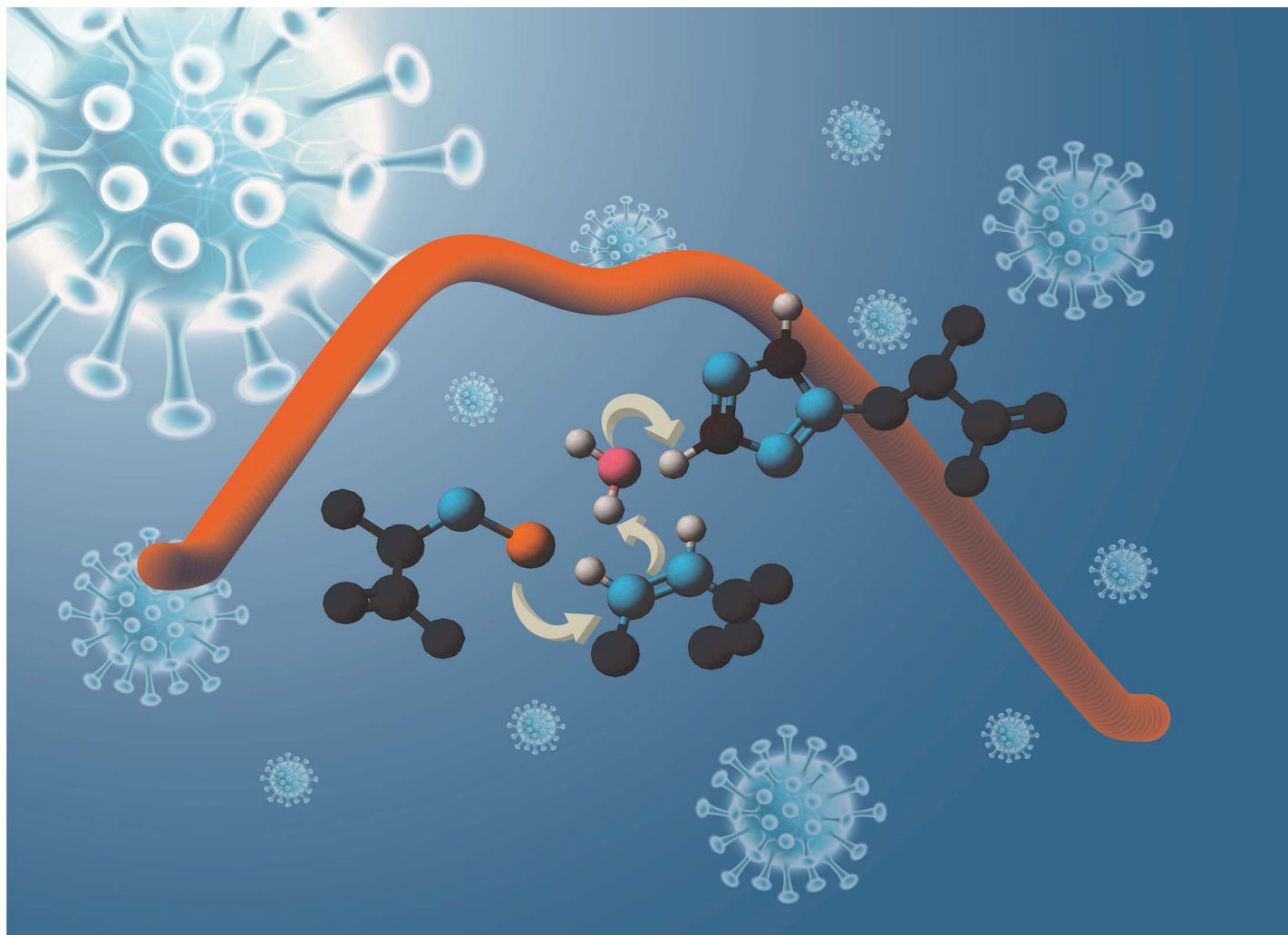
The calculated reaction enthalpy (128.8 kcal·mol⁻¹) is in very good agreement with the experimental estimation (131.5 ± 2.2 kcal·mol⁻¹). This value was obtained from the proton affinity of imidazole (225.3 kcal·mol⁻¹)¹ and the deprotonation enthalpy of methanethiol (356.8 ± 2.2 kcal·mol⁻¹)² as reported by the National Institute of Standards and Technology.³ The uncertainty given for the experimental reaction enthalpy corresponds to that of the deprotonation enthalpy of methanethiol because the value for the proton enthalpy of imidazole is not reported.

References

1. Hunter, E. P. L. & Lias S. G. Evaluated Gas Phase Basicities and Proton Affinities of Molecules: An Update. *J. Phys. Chem. Ref. Data* **27**, 413-656 (1998).
2. Bartmess, J. E., Scott, J. A. & McIver Jr, R. T. Scale of acidities in the gas phase from methanol to phenol. *J. Am. Chem. Soc.* **101**, 6046-6056 (1979).
3. <https://webbook.nist.gov/> accessed May 16th 2020.

Appendix 3

A Microscopic Description of SARS-CoV-2 Main Protease Inhibition with Michael Acceptors. Strategies for Improving Inhibitor Design.



Showcasing research from Professor Tuñón's laboratory, Department of Physical Chemistry, University of Valencia, Spain.

A microscopic description of SARS-CoV-2 main protease inhibition with Michael acceptors. Strategies for improving inhibitor design

Classical and QM/MM Molecular Dynamics simulations are used to unveil the reaction mechanism for the SARS-CoV-2 3CL protease inhibition with a Michael acceptor. The on-the-fly string method is used to explore the multidimensional free energy landscape associated with the reaction showing that the process requires the activation of the catalytic dyad (Cys-His), followed by the nucleophilic attack and a water-mediated proton transfer. These findings could guide the design of new inhibitors of the protease to be employed as therapeutic agents against COVID-19.

As featured in:



See J. Javier Ruiz-Pernía, Iñaki Tuñón *et al.*, *Chem. Sci.*, 2021, 12, 3489.

Cite this: *Chem. Sci.*, 2021, 12, 3489

All publication charges for this article have been paid for by the Royal Society of Chemistry

A microscopic description of SARS-CoV-2 main protease inhibition with Michael acceptors. Strategies for improving inhibitor design†

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

The irreversible inhibition of the main protease of SARS-CoV-2 by a Michael acceptor known as N3 has been investigated using multiscale methods. The noncovalent enzyme–inhibitor complex was simulated using classical molecular dynamics techniques and the pose of the inhibitor in the active site was compared to that of the natural substrate, a peptide containing the Gln–Ser scissile bond. The formation of the covalent enzyme–inhibitor complex was then simulated using hybrid QM/MM free energy methods. After binding, the reaction mechanism was found to be composed of two steps: (i) the activation of the catalytic dyad (Cys145 and His41) to form an ion pair and (ii) a Michael addition where the attack of the S γ atom of Cys145 to the C β atom of the inhibitor precedes the water-mediated proton transfer from His41 to the C α atom. The microscopic description of protease inhibition by N3 obtained from our simulations is strongly supported by the excellent agreement between the estimated activation free energy and the value derived from kinetic experiments. Comparison with the acylation reaction of a peptide substrate suggests that N3-based inhibitors could be improved by adding chemical modifications that could facilitate the formation of the catalytic dyad ion pair.

Received 8th September 2020

Accepted 28th January 2021

DOI: 10.1039/d0sc04978f

rsc.li/chemical-science

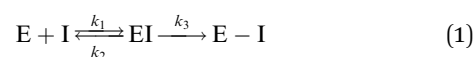
Introduction

A powerful strategy to fight against infectious diseases is the development of drugs to inhibit the activity of one of those enzymes that are crucial in the life cycle of the pathogenic agents. This is the case of the main protease, or 3CL protease (3CL^{PRO}), of coronaviruses in general and of SARS-CoV-2 in particular. The 3CL^{PRO} cleaves the polyproteins translated into the infected cells to produce functional proteins for the coronavirus.¹ As in other cysteine proteases, the proteolysis is performed in the active site of 3CL^{PRO} by a Cys/His catalytic dyad, the substrate cleavage taking place between Gln at the P1 position of the peptide chain and a Gly/Ala/Ser at the P1' one.² This enzyme plays an essential role during the replication of the virus and has no closely related homologues in human cells, making it in an attractive drug target.³

Several lead compounds have already been demonstrated to be effective at inhibiting the activity of SARS-CoV-2 3CL^{PRO}, including Michael acceptors,⁴ α -ketoamides,⁵ carbamoyl derivatives⁶ and aldehydes.⁷ N3 is a Michael acceptor, an α,β -unsaturated carbonyl compound, that was designed as an inhibitor of the 3CL^{PRO} of several coronaviruses, including SARS-CoV and MERS-CoV⁸ and that has been demonstrated to have inhibitory

activity against the ortholog enzyme of SARS-CoV-2.⁴ This compound has a chemical structure similar to that of a peptide, the natural substrate of the enzyme (see Fig. 1). However, the microscopic details of 3CL^{PRO} inhibition by N3 are still unclear.

Kinetic experiments showed that N3 is a potent time-dependent irreversible inhibitor of SARS-CoV-2 3CL^{PRO} that follows the next kinetic scheme:⁴



In a first stage, the inhibitor reversibly binds into the active site of the enzyme forming a noncovalent complex (EI) with a dissociation constant ($K_1 = k_2/k_1$). Afterwards, the inhibitor irreversibly reacts with the enzyme, with a rate constant k_3 , to give a stable acylenzyme (E–I). This acylenzyme is characterized by the formation of a covalent bond between the S γ atom of Cys145 and the C β atom of the inhibitor, as observed in the X-ray structure of the inhibited enzyme.⁴

N3, or any of the other 3CL^{PRO} inhibitors characterized until now, can be used as a starting point for the development of an efficient drug for the treatment of COVID-19. One of the steps in this development is the optimization of the thermodynamic (binding) and kinetic properties of the inhibitor. This improvement should be based in the microscopic knowledge of the inhibition process, which in part relies on the details provided by simulations of the enzyme and the complex formed with the inhibitor. The analysis of the reaction step requires of

Departamento de Química Física, Universidad de Valencia, 46100 Burjassot, Spain.

E-mail: ignacio.tunon@uv.es; j.javier.ruiz@uv.es

† Electronic supplementary information (ESI) available. See DOI: 10.1039/d0sc04978f



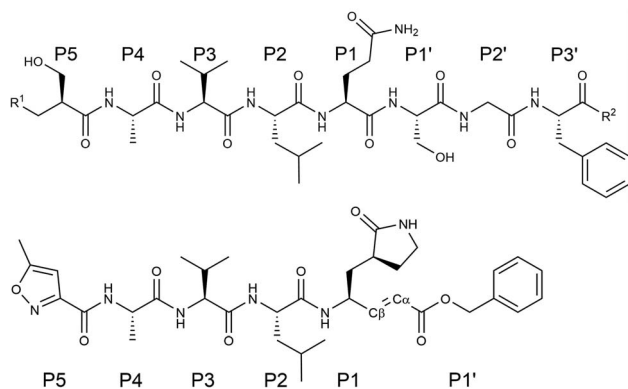


Fig. 1 Chemical structure of a peptide substrate of SARS-CoV-2 3CL^{Pro} (top) and of the N3 inhibitor (bottom). The scissile bond of the peptide is placed between Gln-P1 and Ser-P1'.

the use of QM/MM potentials, which are adequate to describe bond forming and breaking processes. QM/MM techniques have been already employed to study, at microscopic level, the SARS-CoV-2 3CL^{Pro} hydrolysis mechanism with a natural peptide⁹ and a modified peptide having a fluorescent tag as leaving group¹⁰ as substrates. We here use these methods to investigate the inhibition process of this enzyme by N3. The atomistic details provide here could be applied to improve the design of future drugs based in this compound.

Results and discussion

As detailed in the Methods section (see ESI[†]), we carried out classical Molecular Dynamics (MD) simulations of the non-covalent enzyme-inhibitor (EI) complex built from the 7BQY PDB structure.⁴ A total of 4.0 μ s (2 replicas) of classical MD simulation were run using the AMBER19 GPU version of pmemd.^{11,12} We then explored the reaction mechanism for the formation of the covalent acylenzyme complex (E-I, see eqn (1)) using QM/MM simulation methods at the hybrid B3LYP/MM level,^{13,14} including D3 dispersion corrections,¹⁵ with the 6-31+G* basis set, as explained in Methods section. The string-method^{16,17} was employed to find the reaction minimum free energy paths (MFEP) on multidimensional free energy surfaces defined by a set of Collective Variables (CVs) in which we included those geometrical parameters (bond lengths) suffering noticeable changes during the process. A path-CV (s) that measures the advance along the MFEP was defined to trace the corresponding free energy profiles. Umbrella sampling¹⁸ along a distinguished coordinate was used to obtain the free energy difference between the neutral and ion pair (IP) forms of the catalytic dyad (Cys145/His41) within the same QM/MM approach. This methodological combination was previously used in the study of the acylation and de-acylation steps of a natural peptide substrate by SARS-CoV-2 3CL^{Pro} with results in excellent agreement with experiments.⁹ In that work we used the string method to explore the reaction mechanism, with different starting points and initial guesses for the path. In the string method, the initial and final nodes are allowed to evolve

until they reach free energy minima while the rest of nodes trace a MFEP between them. After all our attempts, we found two kinds of mechanisms: either the reaction goes through a metastable IP intermediate, where the Cys-His proton transfer precedes the dyad attack on the substrate or the reaction proceeds without formation of the ion pair by means of a direct proton transfer from the catalytic cysteine to the substrate; presenting this last mechanism a significantly higher activation free energy.⁹ The existence of an IP dyad is compatible with the experimental observations made during the kinetic characterization of the homologue protease of SARS-CoV (96% identical), in which a proteolysis mechanism involving the IP formation was proposed on the basis of the pH-inactivation profile and the analysis of solvent isotope effects.¹⁹ We have then here explored similar reaction mechanisms for the SARS-CoV-2 3CL^{Pro} protease inactivation with N3. It must be also noticed that, in a recent work published during the revision of this manuscript, Moliner and coworkers explored SARS-CoV-2 inhibition by N3 and related inhibitors finding also a reaction mechanism that involves first the formation of the IP followed by the formation of the enzyme-inhibitor covalent bond.²⁰

The noncovalent EI complex

N3 has a chemical structure that resembles that of a peptide substrate (see Fig. 1). As such, the pose found for N3 in the active site of SARS-CoV-2 3CL^{Pro} during our MD simulations is also quite similar to that described for the peptide.⁹ Fig. 2a shows the N3 inhibitor in the active site of one of the protomers of the dimeric enzyme. Analysis of root-mean-square-deviation (RMSD) of the protein and the inhibitor shows that this configuration of the system was stable during the simulated time (see Fig. S1[†]). In this complex, the S_{γ} atom of Cys145 remains close to the C_{β} atom of N3. According to the probability distribution shown in Fig. 2b, the most probable distance between these two atoms is 3.3 Å (there is a small fraction of configurations with larger distances, about 5 Å, that corresponds to a configurational change of the side chain of Cys145 from trans to gauche conformation). In the enzyme-inhibitor complex the catalytic dyad remains hydrogen bonded with the sulfhydryl proton pointing towards the N_{ϵ} atom of His41 (the most probable donor-acceptor distance is 3.3 Å, see Fig. 2b). This configuration suggests a mechanism for the formation of the acylenzyme (E-I) in which the catalytic dyad could be activated by means of a proton transfer from Cys145 to His41 to form an ion pair (IP). This activation mechanism of the catalytic dyad was found to be the first step in the acylation of the natural substrate⁹ and it is compatible with experimental kinetic observations on the ortholog enzyme of SARS-CoV, which is highly similar to the 3CL^{Pro} of SARS-CoV-2.¹⁹

Fig. 2c displays the fraction of hydrogen bond interactions established between protein residues and the different groups of the inhibitor and the peptide substrate. 3CL^{Pro} presents an absolute requirement for Gln at P1 position.²¹ As seen in Fig. 2c the P1 residue of the peptide substrate is the one establishing more hydrogen bond interactions with the enzyme.⁹ In particular, main chain atoms of Gln-P1 form hydrogen bonds with



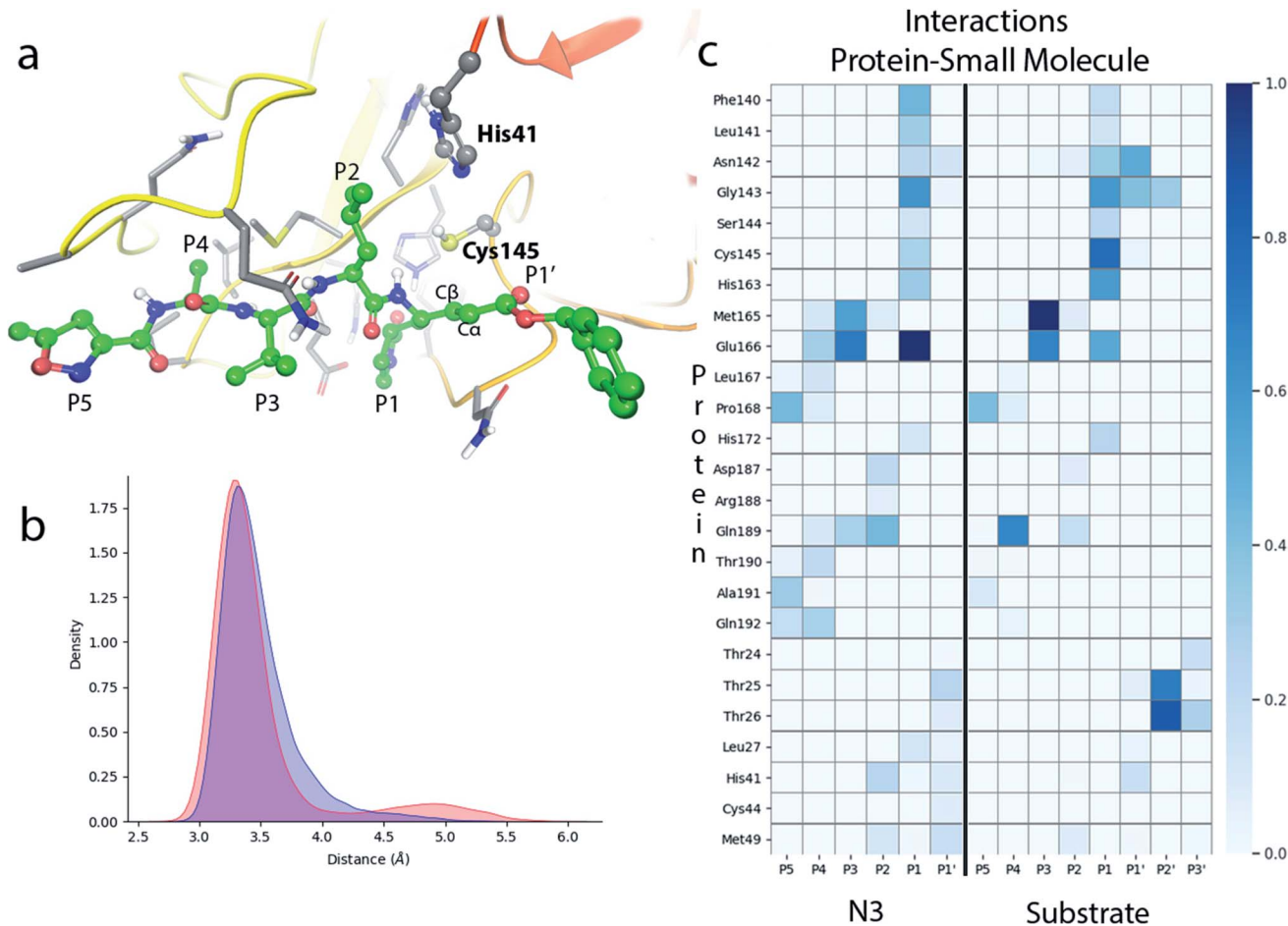


Fig. 2 Molecular dynamics simulation of the noncovalent EI complex between N3 and SARS-CoV-2 3CL^{Pro}. (a) N3 in the active site of the enzyme, showing the location of the catalytic dyad. (b) Probability densities of the distances from the Cys145-S γ atom to the C β carbon atom of the substrate, in red, and to the N ϵ atom of His41, in blue. (c) Fraction of hydrogen bond contacts between residues of N3 and a peptide substrate⁹ 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°.

residues Gly143, Ser144 and His164, while the side chain is accommodated through hydrogen bond contacts with Phe140, Leu141, His163 and Glu166. In N3, the Gln residue is substituted by a γ -lactam ring at P1 position, which essentially reproduces the same hydrogen bond interactions. The interaction pattern of the P2–P5 groups is also quite similar in the inhibitor and the peptide substrate, which could explain the affinity between the enzyme and the inhibitor. Important differences appear at the P1' site, where the serine residue of the peptide substrate is substituted in the inhibitor by a benzyl ester group. While the main chain O atom of Ser-P1' forms hydrogen bonds with the amide group of Gly143 and the side chain of Asn142, the hydroxyl group of its side chain can contact the catalytic dyad (Cys145 and His41). In the inhibitor, the carbonyl O atom of the P1' group can form a hydrogen bond contact with the main chain NH of Gly143. The terminal benzyl group can establish a CH $\cdots\pi$ interaction with the methyl group of Asn142 side chain, while the other side of the ring remains solvent exposed. Because of its mobility, this benzyl ester group can also establish interactions with threonine residues placed

at positions 24–26 and nearby residues. It is also noticeable to remark that the strong interactions established by Gly-P2' of the peptide substrate with Thr25 and Thr26 are significantly weakened or absent in the inhibitor, opening a way to improve the binding affinity between the protease and N3.

Formation of the covalent E-I complex

According to the analysis performed on the noncovalent EI complex, we investigated a possible activation of the catalytic dyad *via* IP formation. With this purpose, we obtained the B3LYPD3/6-31+G*/MM free energy profile associated to the proton transfer from the S γ atom of Cys145 to the N ϵ atom of His41 using an antisymmetric transfer coordinate ($d(\text{S}\gamma\text{-H})-d(\text{N}\epsilon\text{-H})$). Fig. 3a shows the profile obtained with N3 present in the active site; as well as the free energy profiles corresponding to the same proton transfer in the apo enzyme and when a peptide substrate is present in the active site.⁹ The free energy cost to form the IP from the neutral catalytic dyad is the lowest (2.9 kcal mol⁻¹) for the apo enzyme, because the charged



residues (CysS⁻ and HisH⁺) can be stabilized by solvent molecules (see Fig. 3c). The free energy cost is increased up to 4.8 kcal mol⁻¹ when the peptide substrate is present in the active site. In this case the hydroxyl group of Ser-P1' can contribute to stabilize the negative charge on Cys145, but the accessibility of water molecules to the ion pair dyad is significantly reduced when compared to the apo form. In the case of the N3 inhibitor the free energy cost of forming the IP is increased up to 10.7 kcal mol⁻¹, a value obtained as the average between the forward (11.1 kcal mol⁻¹) and backward (10.3 kcal mol⁻¹) profiles. The ionized catalytic dyad is stabilized by a single water molecule that enters into the active site and is placed in between the inhibitor and His41, being hydrogen bonded to the two residues of the dyad (see Fig. 3b). It must be noticed that the barrier for the proton transfer back from His41 to Cys145 from the IP is very small, which suggest that this protonation state is not very stable and that could only appear as a transient species during the acylation process.⁹ Recently, Warshel and coworkers, using an Empirical Valence Bond method, reported an identical value, 2.9 kcal mol⁻¹, for the formation of the IP in the apo enzyme and a similar increase when an α -ketoamide inhibitor is bound in the active site, 7.3 kcal mol⁻¹.²² Altogether, these values indicate that desolvation of the active site upon ligand binding can destabilize the IP form. In fact, formation of the IP, where two charged residues are found at short distance, results in a large dipole

moment (about 14 D at the B3LYP/6-31+G* level) that can be stabilized by solvent molecules. This result also suggests that binding of the ligand into the active site after IP formation could have associated a large energy penalty due to the need to remove water molecules from the active site. Classical MD simulations of the noncovalent complex with the catalytic dyad in the IP form show a trend for the ligand to slowly depart from the active site after several hundred nanoseconds (see Fig. S2[†]), being the active site then occupied by water molecules. In the recent work of Moliner and coworkers, in which a combination of AM1 dynamics and M06-2X energies are used to describe the QM region and the Amber ff03 force field for the MM region, the IP was found to be only 1.3 kcal mol⁻¹ above the neutral dyad when N3 is present in the active site.²⁰ As discussed below, this value leads to an estimated activation energy for the covalent inhibition of 3CL^{Pro} by N3 which seems to be too small when compared to the reference values derived from the inhibition rate constant of closely related enzymes.⁸

Because of the comparatively larger free energy cost of forming the IP from the noncovalent EI complex with the N3 inhibitor found in our study, we investigated first the possibility of a reaction mechanism for the formation of the covalent E-I complex that does not involve a proton transfer from Cys145 to His41. In such a mechanism the sulfhydryl proton is directly transferred to the C α atom of N3 while the S γ atom attacks on the C β one (see Fig. S3[†]). However, the activation free energy

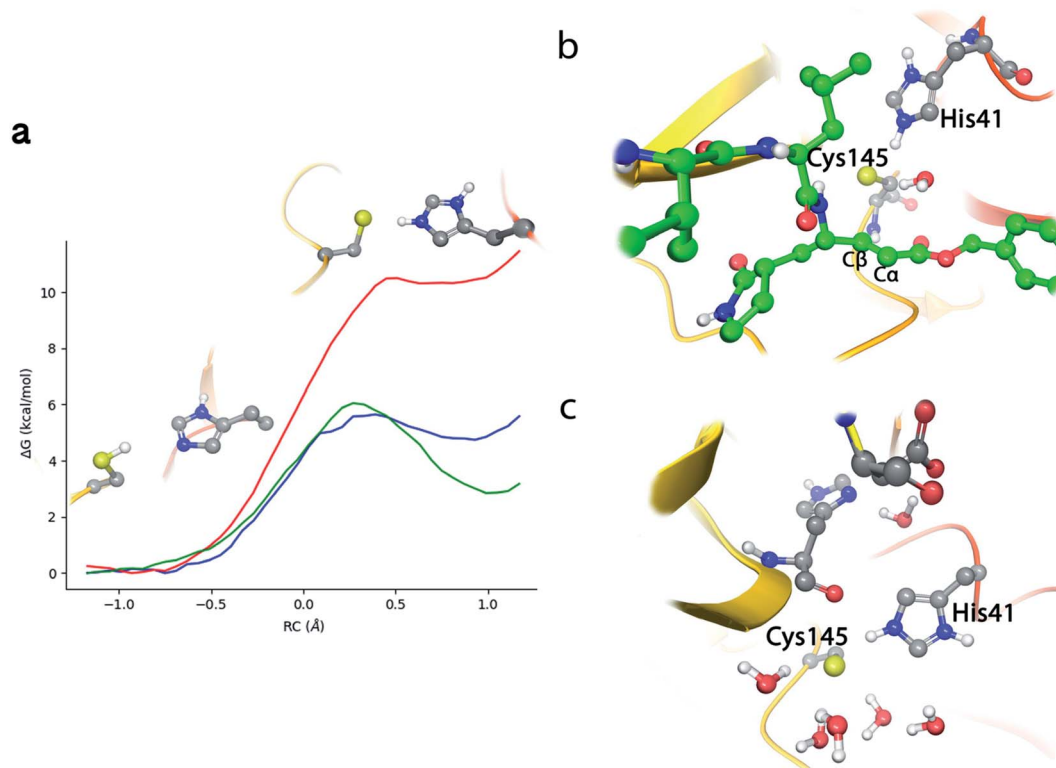


Fig. 3 Proton transfer from Cys145 to His41 in SARS-CoV-2 3CL^{Pro}. (a) Free energy profile for the transformation from the neutral catalytic dyad (left) to the ion pair (IP, right) in the EI complex with N3 (red line); in the complex with the peptide substrate⁹ (blue line) and in the apo enzyme⁹ (green line). (b) Snapshot of the IP configuration in the EI complex showing a water molecule placed between His41 and the inhibitor. (c) Snapshot of the IP configuration in the apo enzyme.



found for this mechanism (about 50 kcal mol^{-1}) is too high and incompatible with the observed inhibition rates (see below).

We thus explored a reaction mechanism for the formation of the E-I acylenzyme from the IP form. This mechanism implies the proton transfer from the $N\epsilon$ atom of His41 to the $C\alpha$ atom of the inhibitor, mediated by the water molecule placed in between, and the nucleophilic attack of the $S\gamma$ atom of Cys145 to the $C\beta$ atom of N3 (see Fig. 4a). The results obtained for the MFEP corresponding to this mechanism at the B3LYPD3/6-31+G*/MM level are shown in Fig. 4b and c. According to the free energy profile the reaction proceeds *via* two Transition States (TS1 and TS2) separated by a shallow intermediate (see Fig. 4b). TS1 is the rate-limiting one with a free energy

$10.6 \text{ kcal mol}^{-1}$ higher than the IP, while the free energy difference corresponding to TS2 is $9.4 \text{ kcal mol}^{-1}$. The evolution of the CVs used to define the multidimensional free energy surface (Fig. 4c) shows that TS1 is associated to the nucleophilic attack of the $S\gamma$ atom to the $C\beta$ atom and the change of the bond between $C\alpha$ and $C\beta$ atoms from double to single. The $S\gamma$ - $C\beta$ distance at TS1 has been reduced from 3.3 to 2.33 Å, while the $C\beta$ - $C\alpha$ distance has been slightly increased from 1.34 to 1.41 Å (see Fig. 4d). TS2 corresponds to the proton transfer from His41 to the neighbor water molecule and from this to the $C\alpha$ atom, being the first proton transfer more advanced than the second one (see Fig. 4e). At TS2 the $S\gamma$ - $C\beta$ bond is significantly shorter (1.91 Å) while the $C\beta$ - $C\alpha$ distance has been elongated up to

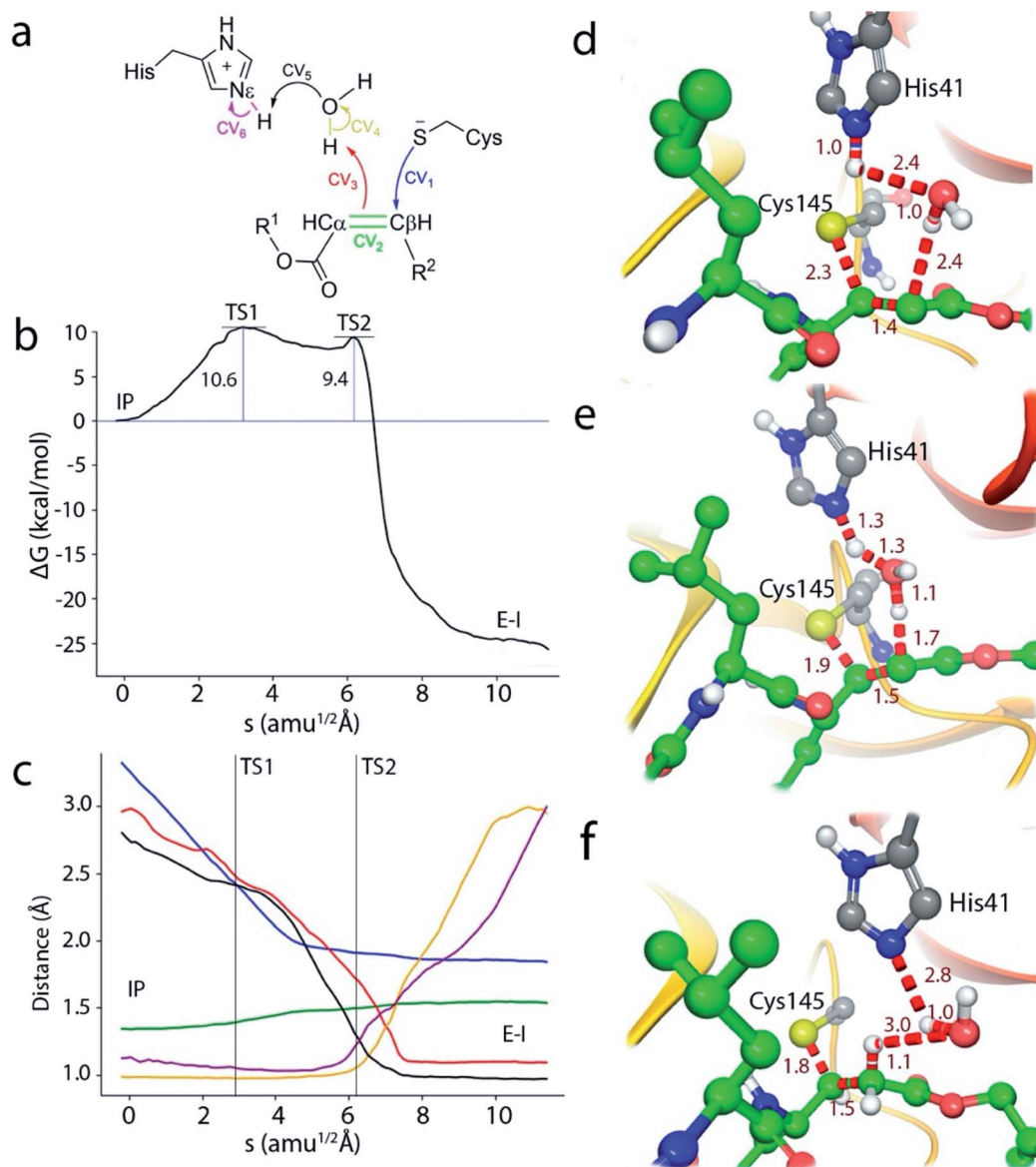


Fig. 4 Simulation of the reaction from the ion pair (IP) to the covalent complex E-I. (a) Collective Variables (CVs) employed to explore the MFEP. (b) B3LYPD3/6-31+G*/MM free energy profile along the path-CV for the formation of the covalent E-I complex from the IP. (c) Evolution of the distances selected as CVs along the MFEP. The color code corresponds to those shown in (a). (d) Representation of TS1. The values of the distances correspond (in Å) to the coordinates of the MFEP where TS1 is located. (e) Representation of TS2 and values of key distances. (f) Representation of the reaction product, the acylenzyme E-I.



a value close to that of a single bond (1.50 Å). Note that in this mechanism the sequence of nucleophilic attack and proton transfer is just the reverse of that observed for the acylation mechanism of the peptide substrate, where the proton transfer to the N atom of the scissile bond precedes the nucleophilic attack on the carbonyl carbon atom.⁹ Finally, the reaction product, where a proton has been transferred to the C α atom, is shown in Fig. 4f. The S γ -C β bond distance found at the E-I complex (1.85 Å) is close to the value found in the X-ray structure of the inhibited enzyme (1.77 Å).⁴ The overlap between the QM/MM and X-ray structures is shown in Fig. S4.†

In order to check the robustness of our mechanistic proposal, the string calculation was repeated using the M06-2X functional with the same basis set and D3 corrections. The resulting free energy profile was almost identical to the B3LYP one, both from the energetic and structural points of view: the geometries and energies of the transition states were very similar at both theoretical levels, as can be seen in Fig. S5.† This result confirms the adequacy of the B3LYP functional for the present Michael addition, in spite of the reported limitations of this functional to correctly describe some enolate or carbanion intermediates.^{23,24} Note that these species are not strictly found in the proposed mechanism because of the proton transfer from His41 to the substrate. It must be also stressed that at both theoretical levels, the string converges to a mechanism evolving from the IP to the covalent product, confirming that, in agreement with our previous work,¹⁰ the IP is a metastable species from which the most favorable mechanism may proceed.

A complete representation of the free energy path from the noncovalent complex (EI) to the covalent one (E-I) is provided in Fig. 5. According to this free energy profile, resulting from the combination of those presented in Fig. 3 and 4, the transformation from the noncovalent EI complex to the covalent one (E-I) is an exothermic process. The free energy difference between E-I and IP (Fig. 4b) is -25.7 kcal mol⁻¹, while the free energy difference between IP and EI is 10.7 kcal mol⁻¹ (Fig. 3a). Combining these two values, our simulations predict that the covalently bonded E-I complex is -15.0 kcal mol⁻¹ more stable than the noncovalent EI complex, which agrees with the observed irreversibility of the inhibition process of 3CL^{Pro} by N3.⁴

Regarding the inactivation rate (k_3 in eqn (1)), our simulations predict that the associated activation free energy results also from the sum of two contributions: the free energy cost of creating the IP form from EI (10.7 kcal mol⁻¹, Fig. 3a) plus the activation free energy of TS1 relative to IP (10.6 kcal mol⁻¹, Fig. 4b). This gives in a total activation free energy of 21.3 kcal mol⁻¹, which according to Transition State Theory (see eqn (S1) in SI†) corresponds to a rate constant of 1.9×10^{-3} s⁻¹ at 300 K. Unfortunately, only the second-order rate constant (k_3/K_1) and not the inactivation rate constant of SARS-CoV-2 3CL^{Pro} by N3 has been estimated.⁴ However, the inactivation rate constant by N3 (k_3) was determined for the highly similar ortholog protease of SARS-CoV.⁸ In this case the activation free energy derived from the rate constant measured at 303 K (3.1×10^{-3} s⁻¹) is 21.2 kcal mol⁻¹. Comparison between SARS-CoV and SARS-CoV-2 main proteases seems appropriate

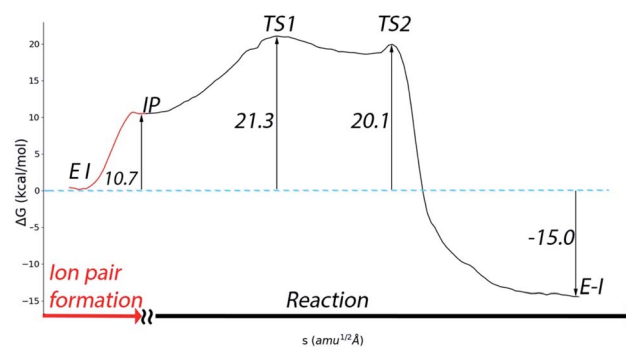


Fig. 5 Free energy profile for the whole transformation of the non-covalent complex (EI) into the covalent one (E-I) through formation of the IP.

considering that they present identical active sites, the same substrate specificity and very similar reaction rate constants for the hydrolysis of peptides.²⁵ The order of magnitude predicted for the k_3 rate constant seems also correct when compared to the values determined for the main proteases of other coronaviruses.⁸ Even if the rate constant for SARS-CoV-2 inhibition would be one order of magnitude faster, which could account for the rapid inhibition reported experimentally,⁴ the resulting activation free energy for the SARS-CoV-2 enzyme would only be 1.4 kcal mol⁻¹ smaller than the reported value for the SARS-CoV one. Lastly, our prediction for the rate constant is compatible with the experimental estimation of k_3/K_1 for the SARS-CoV-2 enzyme ($11\,300$ M⁻¹ s⁻¹).⁴ Combination of our predicted k_3 with the reported estimation for k_3/K_1 gives a K_1 of ~ 0.2 μ M, a value similar to those determined for the inhibition of other coronaviruses's proteases with N3.⁸ The excellent agreement between the experimental values and our theoretical estimation strongly supports our mechanistic proposal for the inhibition of SARS-CoV-2 3CL^{Pro} by a Michael acceptor.

The QM/MM study of Moliner and coworkers found an activation free energy of 11.2 kcal mol⁻¹ and a reaction free energy of -17.9 kcal mol⁻¹.²⁰ While the latter value is close to our findings (-15.0 kcal mol⁻¹), the former departs significantly from our estimation (21.3 kcal mol⁻¹). Their activation free energy provides a rate constant of $\sim 10^4$ s⁻¹, significantly larger than the aforementioned value measured for the homologous SARS-CoV protease. In their mechanistic proposal the proton transfer from His41 to the inhibitor is direct and not water-mediated. However, this mechanistic difference does not explain the gap between their and ours calculated activation free energies. In their simulations the rate-limiting TS corresponds, as in our case, to the S γ -C β bond formation and the free energy difference with the IP is 9.9 kcal mol⁻¹, very close to our value of 10.6 kcal mol⁻¹ (see Fig. 4b). The main difference between our results and those reported by Moliner and coworkers is found in the first part of the process, the free energy cost of forming the IP from the EI complex, 10.7 and 1.3 kcal mol⁻¹, respectively. Differences between the two works may arise from the different QM levels of theory, MM forcefields or to the sampling of different enzymatic configurations (their exploration of the mechanism started from the E-I complex



while we started from the noncovalent EI complex); factors that could affect the relative stability of the neutral and ionic forms of the catalytic dyad. According to our previous discussion, we think that our simulations provides a general picture (see Fig. 5) in better agreement with current experimental results.

Roughly speaking, in our simulations the two steps presented in Fig. 3 and 4, IP formation and Michael addition, contribute similarly to the activation free energy of the inhibition process (about 10 kcal mol⁻¹ each of them). This suggests that the kinetic properties of inhibitors can be improved also by stabilizing the ligand-bound ion pair state. It has been already suggested for other related 3CL proteases (from MERS and SARS-CoV) that stabilization of a charged catalytic dyad could promote catalysis.²⁶ For the SARS-CoV-2 protease it has been shown that inhibitors can shift the protonation state of some residues, not only the catalytic dyad but also other residues found in the vicinity of the active site.²⁷ In principle, a possible strategy is the introduction of chemical groups in the inhibitor structure that imitate the role played by Ser-P1' in the natural substrate. The hydroxyl group of this residue can make contacts with the catalytic dyad that, together with the presence of solvent molecules, can contribute to stabilize the IP.⁹ Interestingly, the position of the water molecule in the rate limiting TS structure found in this work (TS1, see Fig. 4d) could be useful to assist in the design of inhibitors that favor this stabilization process. In this sense, a recently reported potent inhibitor of SARS-CoV-2 3CL^{Pro} (PF-00835231)²⁸ presents a hydroxyl group that matches the position of the water molecule in TS1 (see Fig. S6† for an overlap of the X-ray structure of the enzyme inhibited by PF-00835231 and TS1). This observation illustrates the insights offered by mechanistic studies for the design of new inhibitors.

Conclusions

We have here presented the results of microscopic simulations of SARS-CoV-2 3CL protease inhibition by N3, a Michael acceptor. Classical and hybrid QM/MM simulations were performed to investigate the noncovalent and the covalent enzyme-inhibitor complexes, respectively.

Molecular dynamics simulations of the noncovalent EI complex show that the inhibitor mimics the interactions established by the P1–P5 residues of the natural substrate. Our analysis also shows that an interesting strategy to improve a potential inhibitor based in N3 could be the introduction of chemical changes in the benzyl ester group in such a way that could restore the interactions that the P2' group of the peptide substrate establishes with Thr25, Thr26 and Gly143. This change could increase the affinity between the inhibitor and the protein, reducing the dissociation constant K_i .

Regarding the formation process of the covalent E–I complex, our simulations show that the inhibition mechanism of SARS-CoV-2 3CL^{Pro} by a Michael acceptor involves two steps after binding the inhibitor: (i) the activation of the catalytic dyad by means of the formation of an ion pair and (ii) a Michael addition process where Cys145 attacks to the C β atom of the Michael acceptor and a proton is transferred, water mediated, from His41

to the C α atom of the inhibitor. The contribution of each of the two steps to the activation free energy of the inhibition process is roughly the same (about 10 kcal mol⁻¹) and thus inhibition kinetics can be favored by reducing either of the two contributions. The free energy cost to form the IP is substantially smaller in the enzymatic complex with the peptide substrate than when N3 is present in the active site (about 5.9 kcal mol⁻¹, according to Fig. 3a) and the activation free energy for the acylation of a peptide substrate is also significantly smaller than for the N3 inhibitor (by about the same quantity reflected in Fig. 3a).⁹ This clearly suggests that, in order to improve the kinetic behavior of newly designed inhibitors (increasing k_3), attention must be paid to the formation of the IP. In this sense, the structures found along our reaction path, in particular the rate-limiting transition state, could be useful to guide that design.

Conflicts of interest

The are no conflicts to declare.

Acknowledgements

The authors acknowledge financial support from Feder funds and the Ministerio de Ciencia, Innovación y Universidades (project PGC2018-094852-B-C22) and Generalitat Valenciana (GVCOV19/Decreto180/2020). We also acknowledge Barcelona Supercomputing Center (BSC) for awarding us access to *Mar-*e*Nostrum* and Servei d'Informàtica de la Universitat de València for the use of *Tirant*. The support of Alejandro Soriano from Servei d'Informàtica de la Universitat de València, and the staff from BSC is gratefully acknowledged. Authors also deeply acknowledge Dr Kirill Zinovjev for assistance in the adaptation of the code and helpful discussions during about the use of the string method.

References

- 1 C. R. M. Bangham, *J. Gen. Virol.*, 2003, **84**, 3177–3189.
- 2 R. Hilgenfeld, *FEBS J.*, 2014, **281**, 4085–4096.
- 3 T. Pillaiyar, M. Manickam, V. Namasivayam, Y. Hayashi and S.-H. Jung, *J. Med. Chem.*, 2016, **59**, 6595–6628.
- 4 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. X. X. Yang, T. You, X. X. X. X. Liu, X. X. X. Yang, F. Bai, H. Liu, X. X. X. 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.
- 5 L. Zhang, D. Lin, X. Sun, U. Curth, C. Drosten, L. Sauerhering, S. Becker, K. Rox and R. Hilgenfeld, *Science*, 2020, **368**, 409–412.
- 6 Z. Jin, Y. Zhao, Y. Sun, B. Zhang, H. Wang, Y. Wu, Y. Zhu, C. Zhu, T. Hu, X. Du, Y. Duan, J. Yu, X. Yang, X. Yang, K. Yang, X. Liu, L. W. Guddat, G. Xiao, L. Zhang, H. Yang and Z. Rao, *Nat. Struct. Mol. Biol.*, 2020, **27**, 529–532.
- 7 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.
- 8 H. Yang, W. Xie, X. Xue, K. Yang, J. Ma, W. Liang, Q. Zhao, Z. Zhou, D. Pei, J. Ziebuhr, R. Hilgenfeld, Y. Y. Kwok, L. Wong, G. Gao, S. Chen, Z. Chen, D. Ma, M. Bartlam and Z. Rao, *PLoS Biol.*, 2005, **3**, 1742–1752.
- 9 C. A. Ramos-Guzmán, J. J. Ruiz-Pernía and I. Tuñón, *ACS Catal.*, 2020, **10**, 12544–12554.
- 10 K. Świderek and V. Moliner, *Chem. Sci.*, 2020, **11**, 10626–10630.
- 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 A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5652.
- 14 C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1988, **37**, 785–789.
- 15 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.
- 16 K. Zinovjev and I. Tuñón, *Proc. Natl. Acad. Sci. U. S. A.*, 2017, **114**, 12390–12395.
- 17 L. Maragliano and E. Vanden-Eijnden, *Chem. Phys. Lett.*, 2007, **446**, 182–190.
- 18 G. M. Torrie and J. P. Valleau, *J. Comput. Phys.*, 1977, **23**, 187–199.
- 19 J. Solowiej, J. A. Thomson, K. Ryan, C. Luo, M. He, J. Lou and B. W. Murray, *Biochemistry*, 2008, **47**, 2617–2630.
- 20 K. Arafet, N. Serrano-Aparicio, A. Lodola, A. J. Mulholland, F. V. González, K. Świderek and V. Moliner, *Chem. Sci.*, 2021, DOI: 10.1039/D0SC06195F.
- 21 K. Fan, P. Wei, Q. Feng, S. Chen, C. Huang, L. Ma, B. Lai, J. Pei, Y. Liu, J. Chen and L. Lai, *J. Biol. Chem.*, 2004, **279**, 1637–1642.
- 22 D. Mondal and A. Warshel, *Biochemistry*, 2020, **59**, 4601–4608.
- 23 E. Awoonor-Williams, W. C. Isley, S. G. Dale, E. R. Johnson, H. Yu, A. D. Becke, B. Roux and C. N. Rowley, *J. Comput. Chem.*, 2020, **41**, 427–438.
- 24 J. M. Smith, Y. Jami Alahmadi and C. N. Rowley, *J. Chem. Theory Comput.*, 2013, **9**, 4860–4865.
- 25 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.
- 26 H. Wang, S. He, W. Deng, Y. Zhang, G. Li, J. Sun, W. Zhao, Y. Guo, Z. Yin, D. Li and L. Shang, *ACS Catal.*, 2020, **10**, 5871–5890.
- 27 A. Pavlova, D. L. Lynch, I. Daidone, L. Zanetti-Polzi, M. D. Smith, C. Chipot, D. W. Kneller, A. Kovalevsky, L. Coates, A. A. Golosov, C. J. Dickson, C. Velez-Vega, J. S. Duca, J. V. Vermaas, Y. T. Pang, A. Acharya, J. M. Parks, J. C. Smith and J. C. Gumbart, *Chem. Sci.*, 2021, DOI: 10.1039/D0SC04942E.
- 28 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.



Supplementary Information

A Microscopic Description of SARS-CoV-2 Main Protease Inhibition with Michael Acceptors. Strategies for Improving Inhibitors Design

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

Methodology	S2
Figure S1. Plot of the RMSD values for the EI complex.	S6
Figure S2. Plot of RMSD values for the ligand in the noncovalent complex with the dyad in the IP form.	S7
Figure S3. Simulation of the acylation reaction without formation of an ion pair	S8
Figure S4. Overlap between QM/MM and x-ray acylenzyme structures	S9
Figure S5. M062X free energy profile for the IP transformation into E-I	S10
Figure S6. Overlap between QM/MM TS structure and the x-ray structure of the E-I complex formed with PF-00835231.	S11
References	S12

Methodology

Classical Molecular Dynamics simulations

The Michaelis complex was built from the structure with PDB code 7BQY. This contains the crystallographic data for the 3CL^{pro} of SARS-CoV-2 forming a covalent complex with the N3 inhibitor (with a resolution of 1.7 angstrom).¹ Crystallographic water molecules were preserved because they can play an important role in modulating the properties of the catalytic dyad, in particular the water molecule bridging between His41-His164 and Asp187.^{2,3}

The N3 inhibitor was parameterized following the non-standard residue parameterization procedure implemented in Amber using the Antechamber program⁴ from the AmberTools18⁵ package. For this purpose, terminal residues of the inhibitor (named as 02J PJE and 010 in the PDB file) were capped using NME and ACE groups respectively. Atomic charges were then obtained using the Restrained Electrostatic Potential (RESP) method⁶ at the HF/6-31G* level.

The protonation states of the titratable groups were determined using PROPKA3.0⁷ at pH 7.4. For neutral histidine residues, the δ/ϵ protonation state was determined after visual inspection of the x-ray structure. In particular, the pair of histidine residues closer to the active site (His41 and His164) were modelled in their neutral states, protonated at δ and ϵ positions, respectively (HD41-HE164). This combination has been recently shown to be the most stable N3-bound state.² Then, the system was built using the tleap tool from AmberTools,⁵ with regular amino acids being described using the ff14SB forcefield.⁸ The charge of the enzyme-inhibitor complex was neutralized adding Na⁺ atoms. The system was solvated into a box of TIP3P water molecules,⁹ with a buffer region of at least 12 Å from any protein/substrate atom to the limits of the simulation box. The resulting system was then minimized 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. A linear heating ramp was used to rise the temperature from 0 to 300 K along 120 ps followed by 20 ps simulation at 300 K. During the heating process a harmonic potential was used to restrain the backbone heavy atoms with a force constant of 20 kcal·mol⁻¹·Å⁻². Afterwards, a set of equilibration simulations in the NPT ensemble (300K and 1 bar) were carried out where the force constants for the positional restraint were gently reduced from 15 to 0 kcal·mol⁻¹·Å⁻², decreasing 3 units every 1.25 ns. After 6.25 ns the positional restraint was completely removed and the system run restraint-free for another 1.25 ns. Finally, two replicas of NVT simulations at 300 K were performed with a 2 fs time step using SHAKE¹⁰ during 3.0 and 1.0

μs , respectively. The long-range electrostatic interactions were described using the Particle Mesh Ewald Method,^{11,12} while a 10 Å cut-off radius was used to evaluate the short-range interactions. Berendsen barostat and Langevin thermostat were used to control pressure and temperature, respectively, where required. For all classical molecular dynamic simulations AMBER19 GPU version of PMEMD^{13,14} was employed.

QM/MM calculations

The free energy surfaces associated to the reaction mechanism were explored using QM/MM simulations. The QM region includes the side chains of the catalytic dyad (Cys145 and His41) and a water molecule (in the mechanism presented in Figure 4). The QM region selected for the N3 inhibitor includes the backbone of P1 and P1' residues corresponding plus the carbonyl carbon atom of P2. The P1 side chain and the benzene ring in P1' were not included in the QM subsystem. The rest of the system was described at the MM level as previously explained. To describe the QM subsystem we used the B3LYP functional^{15,16} and D3 dispersion corrections.¹⁷ Mechanisms were explored using the 6-31G* and 6-31+G* basis sets (see below). As previously reported, this is a good choice to describe the acylation of a peptide substrate by the SARS-CoV-2 protease,³ providing results in excellent agreement with experiments. A systematic study on cysteine-histidine proton transfer found that the B3LYP functional was the most adequate to obtain an electronic description in agreement with higher level methods.¹⁸ For the reaction step we also used the M06-2X functional¹⁹ with D3 corrections to verify the robustness of our conclusions. As usual, Lennard-Jones parameters of QM atoms were imported from the ff14SB forcefield without additional optimization. All QM/MM calculations were performed using a modified version of Amber18^{5,20} coupled to Gaussian16²¹ for Density Functional Theory calculations. A cutoff-radius of 15 Å was used for all QM-MM interactions and the temperature was 300 K.

Our Adaptive String Method (ASM)²² implementation was used to explore the free energy landscape associated to the chemical reaction. N replicas of the system (the string nodes) were used to connect the reactants and product regions. These nodes are evolved according to the averaged forces and kept equidistant, converging into the minimum free energy path (MFEP) in a space of arbitrary dimensionality defined by the collective variables (CVs). The bond lengths selected as CVs coordinates to describe the two reaction mechanisms explored in this work can be seen in the panels a) of Figures 4 and S2. After convergence, a path Collective Variable (path-

CV) is defined to measure the advance of the system along the MFEP. This path-CV is the reaction coordinate employed to trace the reaction free energy.

Two mechanistic proposals were explored at the B3LYPD3/MM level. Each of the strings was composed of 96 nodes, which were propagated with a time step of 1 fs until the RMSD of the string fell below $0.1 \text{ amu}^{1/2} \cdot \text{\AA}$ for at least 2 ps. Umbrella sampling²³ technique was employed to obtain the free energy profiles along the path-CVs. At least 10 ps simulations were run for every node along the converged path and the Weighted Histogram Analysis Method (WHAM)²⁴ 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.²² In these calculations we firstly used the 6-31G* basis set and then the best path was recalculated using the 6-31+G* basis set. Inclusion of diffuse functions in the basis set resulted in a modest reduction of the activation free energy for the mechanism presented in Figure 4 of about $2 \text{ kcal} \cdot \text{mol}^{-1}$.

For the study of the proton transfer from Cys145 to His41, a simple coordinate, defined as the antisymmetric combination of the distances of the proton to the donor and acceptor atoms of the catalytic dyad ($d(\text{S}\gamma\text{-H}) - d(\text{N}\epsilon\text{-H})$), was employed. Then, the free energy profile associated to the formation of the catalytic dyad ion pair (IP) was obtained using Umbrella Sampling.²³ The integration was carried out using WHAM method.²² For this profile only the side chains of the two involved residues were included in the QM region (using the B3LYPD3/6-31+G* level of theory). 40 windows were used separated every 0.06 \AA along the reaction coordinate, and a force constant of $600 \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{\AA}^{-2}$ was used to drive the reaction coordinate change. For each simulation window the system was first minimized, heated and then production was run for 20 ps. The free energy change was calculated as the average between the forwards and backwards profiles. All the rest of details of the simulations were as described previously.

The calculated activation free energies are compared to experimentally derived values, using standard Transition State Theory expression to convert between rate constants (k_r) and activation free energies (ΔG^\ddagger):

$$k_r = \frac{k_B T}{h} e^{-\frac{\Delta G^\ddagger}{RT}} \quad (\text{eq. S1})$$

where k_B , h and R are the Boltzmann, Planck and gases constants and T the temperature.

Examples of MM and QM/MM (the TS) trajectories can be accessed through covid.molssi.org/simulations:

<https://covid.molssi.org/simulations/#695-us-simulation-of-sars-cov2-3clpro-in-complex-with-the-n3-inhibitor-all-atom-simulation>

<https://covid.molssi.org/simulations/#11-ps-qmmm-simulation-of-transition-state-of-sars-cov2-3clpro>

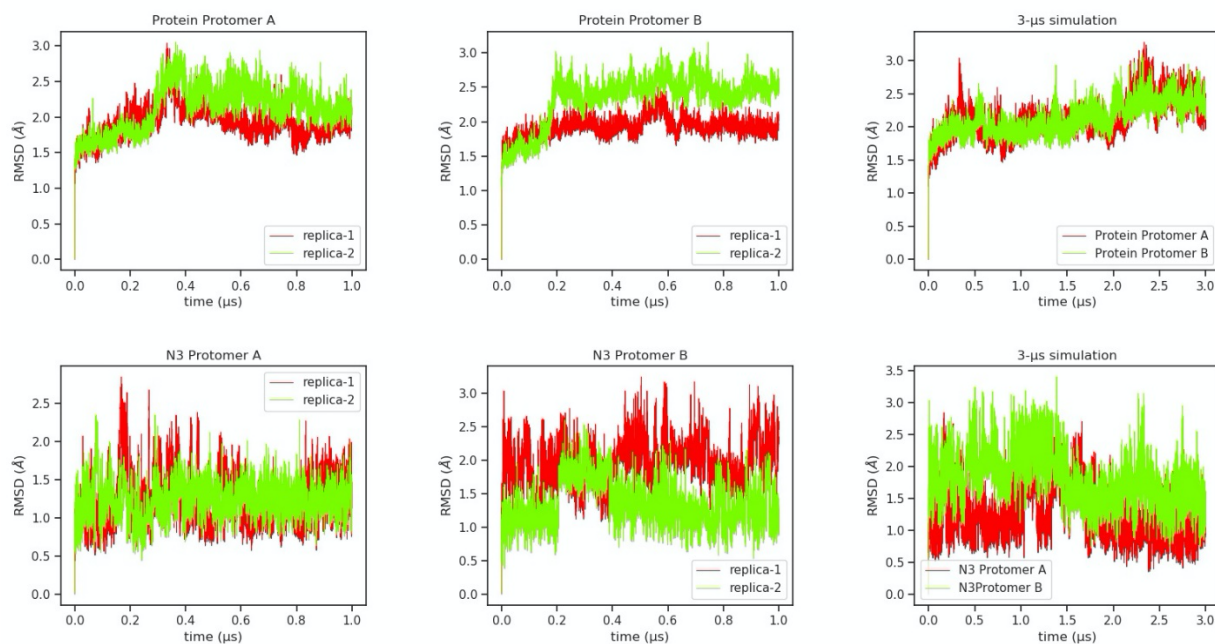


Figure S1. Plot of the RMSD values obtained for the protein (up) and the substrate (down) during the 2 replicas (simulation time of 3.0 μs and 1 μs , respectively) of the noncovalent enzyme-inhibitor complex of the SARS-CoV-2 protease with N3. From left to right, RMSD of protomer A, protomer B. The RMSD has been computed using all non-hydrogen atoms taking as reference the initial structure prepared for the simulations.

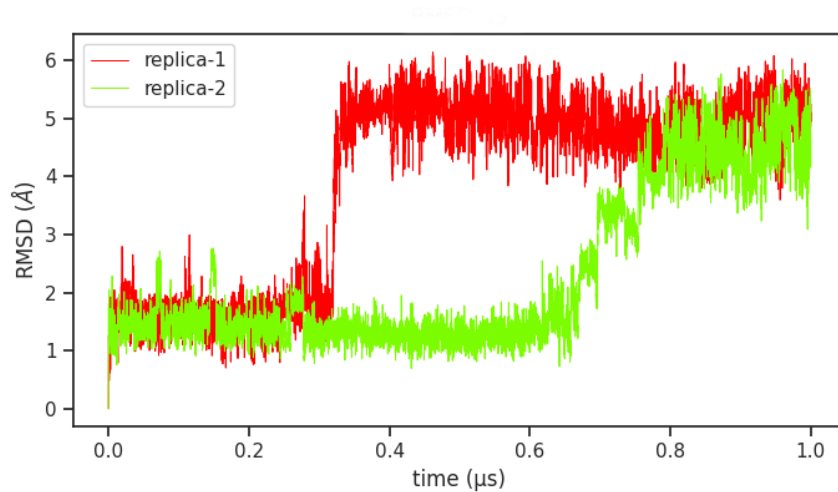


Figure S2. Plot of the RMSD values obtained for the substrate during the 2 replicas (simulation time of 1.0 μs each) of the noncovalent enzyme-inhibitor complex of the SARS-CoV-2 protease with N3 with the catalytic dyad in the IP form.

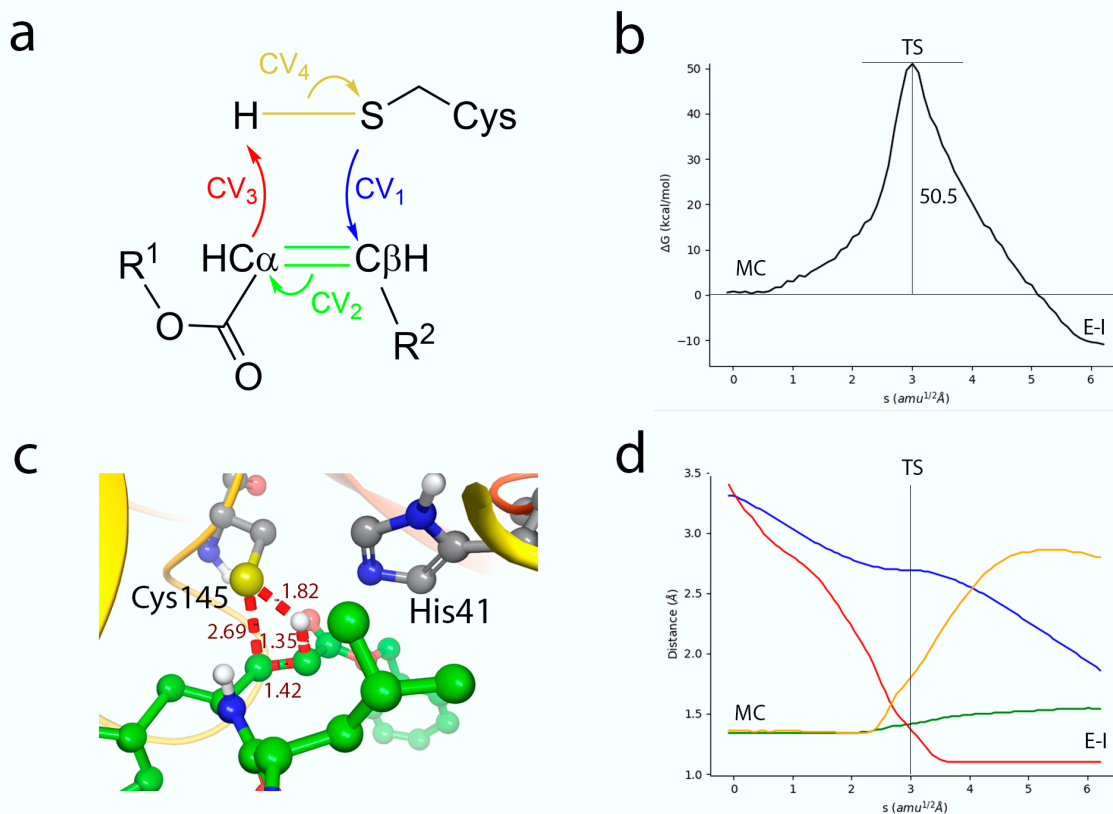


Figure S3. Simulation of the reaction from noncovalent (EI) to covalent complex (E-I) without a proton transfer from Cys145 to His41. **(a)** Collective Variables (CVs) employed to explore the MFEP. **(b)** B3LYPD3/6-31G*/MM free energy profile along the path-CVs for the formation of the covalent E-I complex from EI. **(c)** Representation of TS. The values of the distances (in \AA) correspond to the coordinates of the MFEP where the TS is located. Evolution of the distances selected as CVs along the MFEP. The color code corresponds to those shown in Figure S2a. **(d)** Evolution of the distances selected as CVs along the MFEP. The color code corresponds to those shown in Figure S2a.

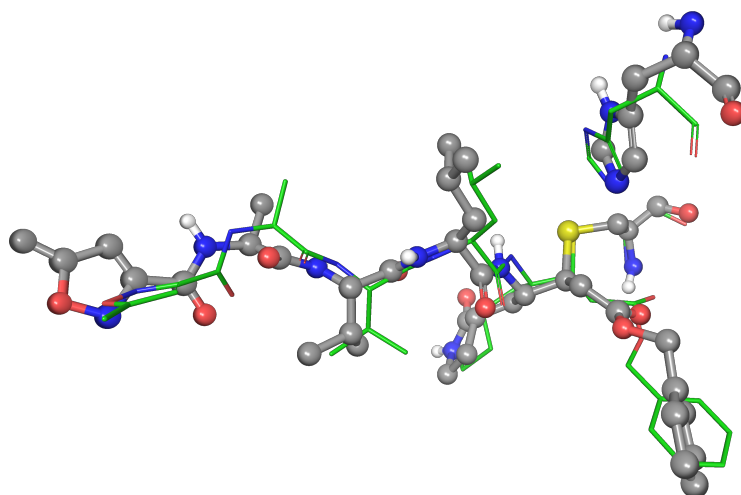


Figure S4. Overlap between the QM/MM (ball & sticks representation) and x-ray (licorice style) structures of the inhibitor in the acylenzyme complex. The x-ray structure corresponds to the PDB file with code 7BQY.

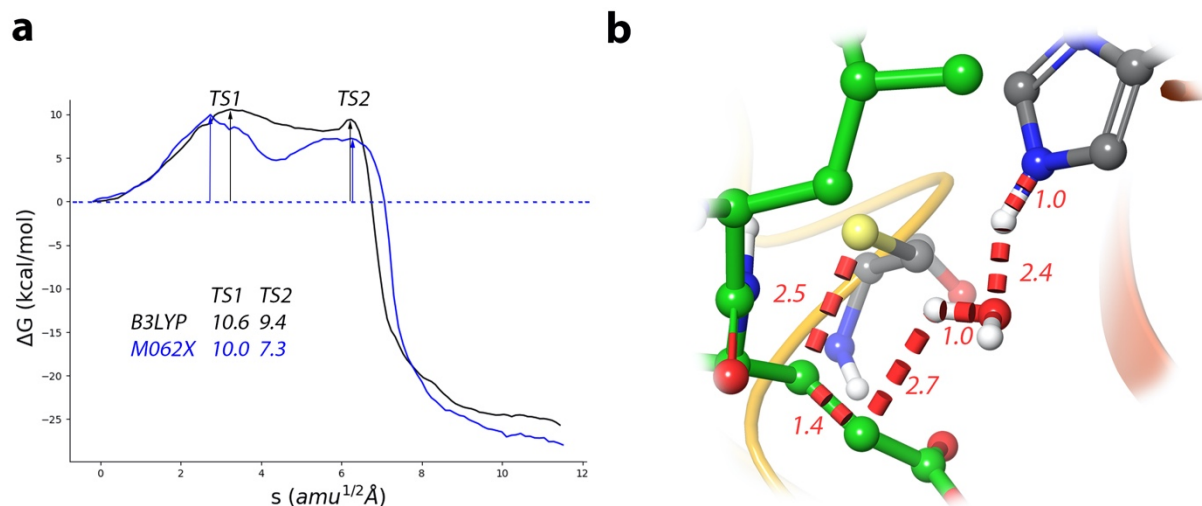


Figure S5. Simulation of the reaction from the IP to the covalent E-I complex. **(a)** B3LYPD3/6-31G*/MM and M062xD3/6-31+G*/MM free energy profiles along the path-CV for the formation of the covalent E-I complex. **(b)** Representation of the M062X structure for TS1, to be compared with the B3LYP one presented in Figure 4d. The values of the distances (in Å) correspond to the coordinates of the MFEP where TS1 is located.

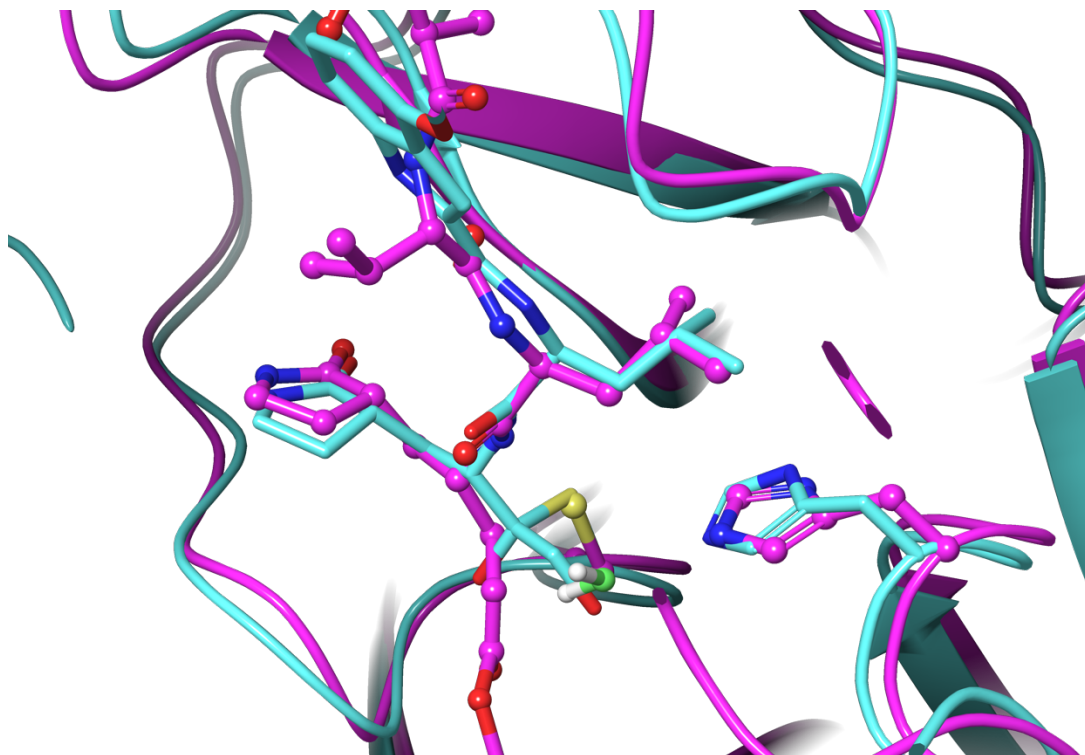


Figure S6. Overlap between the QM/MM TS1 structure (ball & sticks representation with carbon atoms in purple, the oxygen of the water molecule is highlighted in light green) and x-ray structure of the acylenzyme complex formed with PF-00835231 (licorice style with carbon atoms in light blue). The x-ray structure corresponds to the PDB file with code 6XHM.

References

- 1 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. X. X. Yang, T. You, X. X. X. X. Liu, X. X. X. Yang, F. Bai, H. Liu, X. X. X. 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.
- 2 A. Pavlova, D. L. Lynch, I. Daidone, L. Zanetti-Polzi, M. D. Smith, C. Chipot, D. W. Kneller, A. Kovalevsky, L. Coates, A. A. Golosov, C. J. Dickson, C. Velez-Vega, J. S. Duca, J. V. Vermaas, Y. T. Pang, A. Acharya, J. M. Parks, J. C. Smith and J. C. Gumbart, *Chem. Sci.*, , DOI:10.1039/D0SC04942E.
- 3 C. A. Ramos-Guzmán, J. J. Ruiz-Pernía and I. Tuñón, *ACS Catal.*, 2020, **10**, 12544–12554.
- 4 J. Wang, W. Wang, P. A. Kollman and D. A. Case, *J. Mol. Graph. Model.*, 2006, **25**, 247–260.
- 5 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.*
- 6 C. I. Bayly, P. Cieplak, W. Cornell and P. A. Kollman, *J. Phys. Chem.*, 1993, **97**, 10269–10280.
- 7 M. H. M. Olsson, C. R. Søndergaard, M. Rostkowski and J. H. Jensen, *J. Chem. Theory Comput.*, 2011, **7**, 525–537.
- 8 J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser and C. Simmerling, *J. Chem. Theory Comput.*, 2015, **11**, 3696–3713.
- 9 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.
- 10 J.-P. Ryckaert, G. Ciccotti and H. J. . Berendsen, *J. Comput. Phys.*, 1977, **23**, 327–341.
- 11 T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089–10092.
- 12 U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577–8593.
- 13 S. Le Grand, A. W. Götz and R. C. Walker, *Comput. Phys. Commun.*, 2013, **184**, 374–380.

- 14 R. Salomon-Ferrer, A. W. Götz, D. Poole, S. Le Grand and R. C. Walker, *J. Chem. Theory Comput.*, 2013, **9**, 3878–3888.
- 15 A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5652.
- 16 C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B*, 1988, **37**, 785–789.
- 17 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.
- 18 A. Paasche, T. Schirmeister and B. Engels, *J. Chem. Theory Comput.*, 2013, **9**, 1765–1777.
- 19 Y. Zhao and D. G. Truhlar, *Theor. Chem. Acc.*, 2008, **120**, 215–241.
- 20 K. Zinovjev, String-Amber, <https://github.com/kzinovjev/string-amber>, (accessed 24 June 2020).
- 21 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, 2016.
- 22 K. Zinovjev and I. Tuñón, *J. Phys. Chem. A*, 2017, **121**, 9764–9772.
- 23 G. M. Torrie and J. P. Valleau, *J. Comput. Phys.*, 1977, **23**, 187–199.
- 24 S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen and P. A. Kollman, *J. Comput. Chem.*, 1992, **13**, 1011–1021.

Appendix 4

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



ACS Catalysis

APRIL 2, 2021

VOLUME 11 NUMBER 7

PUBS.ACS.ORG/ACSCATALYSIS



ACS Publications
Most Trusted. Most Cited. Most Read.

www.acs.org

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

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



Cite This: *ACS Catal.* 2021, 11, 4157–4168



Read Online

ACCESS |



Metrics & More



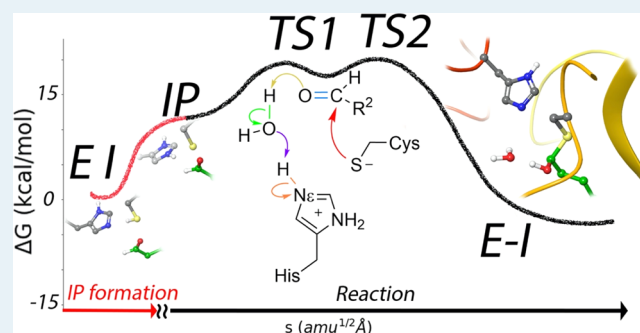
Article Recommendations



Supporting Information

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 a 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 γ atom of Cys145 to the aldehyde carbon atom and a water-mediated proton transfer from the N ϵ 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.

KEYWORDS: 3CL protease, SARS-CoV-2, minimum free energy path, QM/MM, inhibitors, aldehyde derivatives, conformational changes



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.

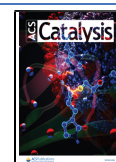
One of the strategies for the development of antiviral drugs is to target the inhibition of one of those enzymes that are essential for the vital cycle of the virus. When SARS-CoV-2 infects a cell, it utilizes the transcription machinery to translate the viral genome into two long polyproteins. These long chains

must be cleaved at specific sites to produce the nonstructural 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 the Gln-(Ser/Ala/Gly) peptide bond, a sequence preference that is not used by human proteases.⁴ This characteristic makes the 3CL protease an 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

Received: December 16, 2020

Revised: February 15, 2021

Published: March 25, 2021



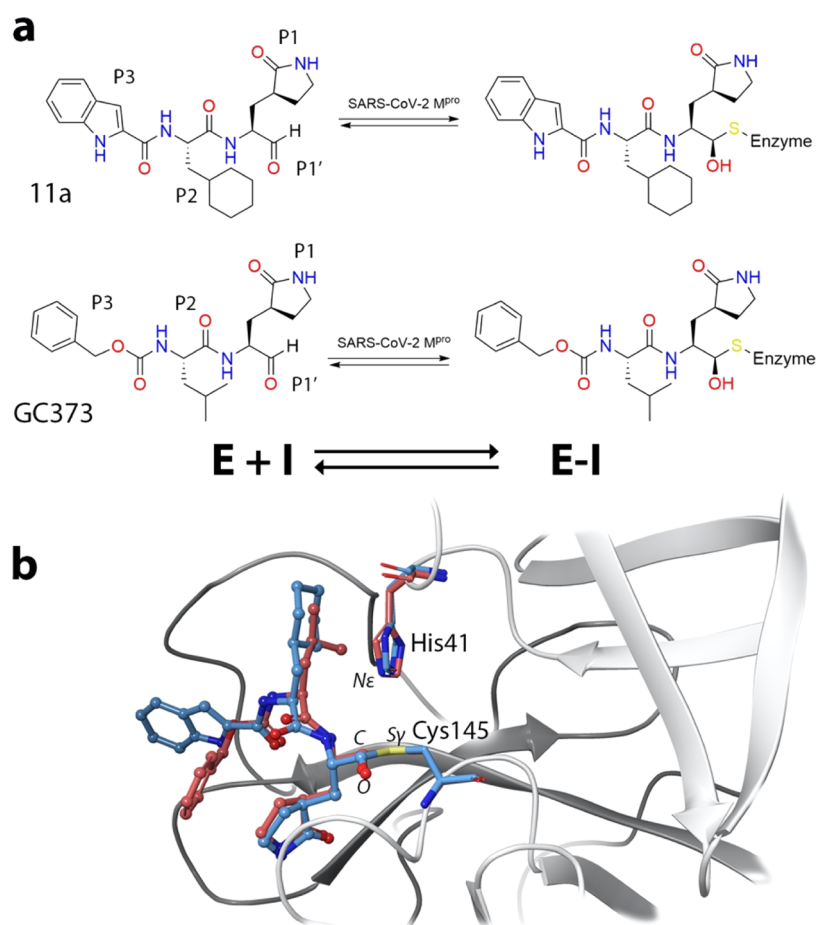


Figure 1. Inhibitors of the SARS-CoV-2 3CL protease presenting an aldehyde group as a warhead. **(1a)** Aldehyde derivatives bind to the active site forming the noncovalent E–I complex and then react with the thiol group of the catalytic cysteine to yield a hemithioacetal (the E–I covalent complex). **(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 carbon atoms in light blue. The PDB codes are 6WTK and 6LZE, respectively.

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 deacylation. 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* γ atom of the catalytic cysteine. During deacylation, the second fragment is released by the action of a water molecule, recovering the enzyme for a new catalytic cycle.

Until now, 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 to the active site of the protease forming a noncovalent complex (E–I) 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* assays.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 *S* γ atom of the catalytic cysteine and the aldehyde carbon atom is of about 1.8 Å, corresponding to the formation of an S–C covalent bond. The 6LZE and 6WTK PDB¹⁵ structures shown in Figure 1b correspond to the (*S*) configuration of the hemithioacetal, 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 *N* ϵ atom of His41 is pointing toward

the binding site, while in 6LZE, the N δ 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 N ϵ atom in the ϵ -rotamer is better oriented to serve as a proton donor to the aldehydic oxygen atom of the inhibitor.

Computational simulations of SARS-CoV-2 3CL protease have been devoted to study its reactivity with peptide substrates, including the acylation and deacylation steps.^{16,17} Regarding covalent inhibition, QM/MM methods have been also used to analyze the reaction with irreversible Michael acceptors^{18,19} and α -ketoamide inhibitors.²⁰ This last study provides a complete evaluation of the binding free energy of the covalent inhibitor as the sum of the noncovalent binding and the reaction steps. 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 description of the interactions established in the noncovalent complex (E-I) and the reaction mechanism leading to the covalent product (the hemithioacetal complex or E-I). Our simulations stress on the interplay between conformational changes of the protein and/or the substrate and the possible reaction mechanisms. The most stable configuration for His41 both in the apo form and in the noncovalent complex formed with the inhibitor is the ϵ -rotamer. 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 experimental⁷ and computational studies.^{16,17} The process continues with the nucleophilic attack of the S γ atoms on the aldehydic carbon atom and the proton transfer from the catalytic histidine to the aldehyde oxygen atom through a water molecule to form the (S)-hemithioacetal. This mechanism is confirmed by comparison with the structures of recently proposed inhibitor PF-00835231¹² that mimic the transition state of protease inhibition by aldehydes. Our simulations are also useful to rationalize the behavior of other proposed inhibitors, considering possible conformational changes of the inhibitor warhead in the active site and how the reaction mechanism can change depending on the conformation of the inhibitor. Other reaction mechanisms, involving different proton transfer routes and/or leading to the (R) form of the product, can be feasible if other protein and/or inhibitor conformations are populated. 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 derivatives, 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 Å).¹⁰ **11a** was parameterized following the nonstandard residue parameterization procedure implemented in Amber using the Antechamber program²¹ from the AmberTools18²² package. Atomic charges were obtained using the restrained electrostatic potential (RESP) method²³ at the HF/6-31G* level.

Tleap tool from abertools²² was used to build the system, with the ff14SB forcefield²⁴ to describe the canonical amino acids. PROPKA3.0²⁵ 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 the 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 (300 K 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 μ s of the noncovalent enzyme inhibitor complex with the catalytic dyad residues (Cys145 and His41) in the neutral state and with these residues charged forming an IP. In these simulations, the time step was 2 fs using SHAKE²⁶ to constraint bonds involving hydrogen atoms. Long-range electrostatic interactions were described using the particle mesh Ewald method,^{27,28} while the cut-off radius to evaluate the short-range interactions was 10 Å. Pressure and temperature were controlled using a Berendsen barostat and Langevin thermostat, respectively. For all classical molecular dynamic simulations, AMBER19 GPU version of PMEMD^{29,30} was employed.

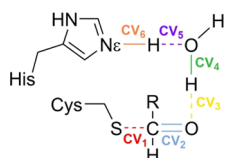
To study the free-energy profiles associated to conformational changes of the catalytic histidine and of the aldehyde inhibitor, classical potentials of mean force were traced using umbrella sampling.³¹ For the change in the rotameric state of His41, the dihedral angle formed by the C α -C β -C γ -N δ atoms of the residue was used as a distinguished coordinate. The dihedral angle was evolved using increments of 5° 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 the distinguished coordinate. In this case, the dihedral angle was evolved using increments of 10° 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 the 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. Afterward, a total of 52 ns of classical MD was performed for each window, the first 2 ns was run for relaxation followed by 50 ns of production. Therefore, the total simulation time for the production stage was longer than 2 μ s for each free-energy profile corresponding to His41 rotation and longer than 1 μ s for the rotation of the aldehyde group of the inhibitor. The free-energy profiles were integrated using the weighted histogram analysis method (WHAM).³²

QM/MM Calculations. The adaptative string method (ASM)³³ 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 in energy related to the variation of a couple of distances, angles, or dihedrals (as made in most of computational studies) but including all the needed degrees of freedom without implying an 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 MFEP in a space of arbitrary dimensionality defined by the collective variables (CVs) shown in Scheme 1. Using QM/MM MD simulations, nodes are

Scheme 1. Representation of the CVs Used in the Exploration of the Reaction Mechanism



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 MFEP displaying an RMSD below $0.1 \text{ amu}^{1/2} \cdot \text{\AA}$ or 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 advancement 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 WHAM³² 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.³³ The QM region was described using a B3LYP functional^{34,35} with a 6-31+G* basis set and D3 dispersion corrections.³⁶ This is a good choice to describe both the acylation of a peptide substrate¹⁶ and an inhibitor¹⁸ by the SARS-CoV-2 protease with activation energy results in excellent agreement with the experiment. A systematic study on the cysteine–histidine proton transfer also pointed out to the B3LYP functional as the most adequate.³⁷ All QM/MM calculations were performed using a modified version of Amber18^{22,38} coupled to Gaussian16³⁹ 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, and the backbone atoms of residues P1 and P1' in the 11a inhibitor. Any other atom was described as an MM level as explained in the classical molecular dynamic section. The integration time step in QM/MM simulations was of 1 fs. Because hydrogen atoms are involved in the reaction mechanism, we checked that the string method converged to the same MFEP when a time step of 0.5 fs was used (see Figure S1).

For the study of the proton transfer within the catalytic dyad, that is, from Cys145 to His41, the antisymmetric

combination of the distances of the proton to the donor and acceptor atoms [$d(\text{S}\gamma\text{--H})\text{--}d(\text{N}\epsilon\text{--H})$] was employed to trace the free-energy profile using umbrella sampling.³¹ The integration was carried out using the WHAM method. In this case, only the side chains of the two involved residues were included in the QM region (described at the B3LYP/6-31+G* level of theory with D3 corrections). A total of 40 windows evenly spaced each at 0.06 Å were used to cover the whole range of the reaction coordinate. A force constant of $600 \text{ kcal}\cdot\text{mol}^{-1}\text{\AA}^{-2}$ was employed to drive the reaction coordinate. Remaining details of the simulations were as described previously.

RESULTS AND DISCUSSION

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 a 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 S γ –C distance between the enzyme and the inhibitor. The catalytic dyad (Cys145 and His41) was modeled in the neutral state, which is the most stable form for the Michaelis and E–I complexes.^{16,18,37}

As explained before, in the 6LZE X-ray structure, the catalytic histidine is found in the δ -rotameric state. With the help of Pocketome,⁴⁰ we explored 91 different X-ray structures of the 3CL main protease in the apo form containing several covalent and noncovalent inhibitors (see Table S1). In 87.5% of the active sites, including all those corresponding to the 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 modeled the noncovalent complex of the 3CL protease with 11a with the two His41 rotamers (see Figure 2a,b) and ran 1 μs long MD simulations. In principle, both rotameric states could be productive as far as 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 N ϵ (His41)–S γ (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) corresponded 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 C β –C γ bond in the apo form and in the noncovalent complex (E–I) 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 2.2 and 3.2 $\text{kcal}\cdot\text{mol}^{-1}$, 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, in particular, for the apo form, which could explain that the δ -rotamer is found in 12.5% of the active sites of the SARS-CoV-2 3CL protease, as reported in Table S1.

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.

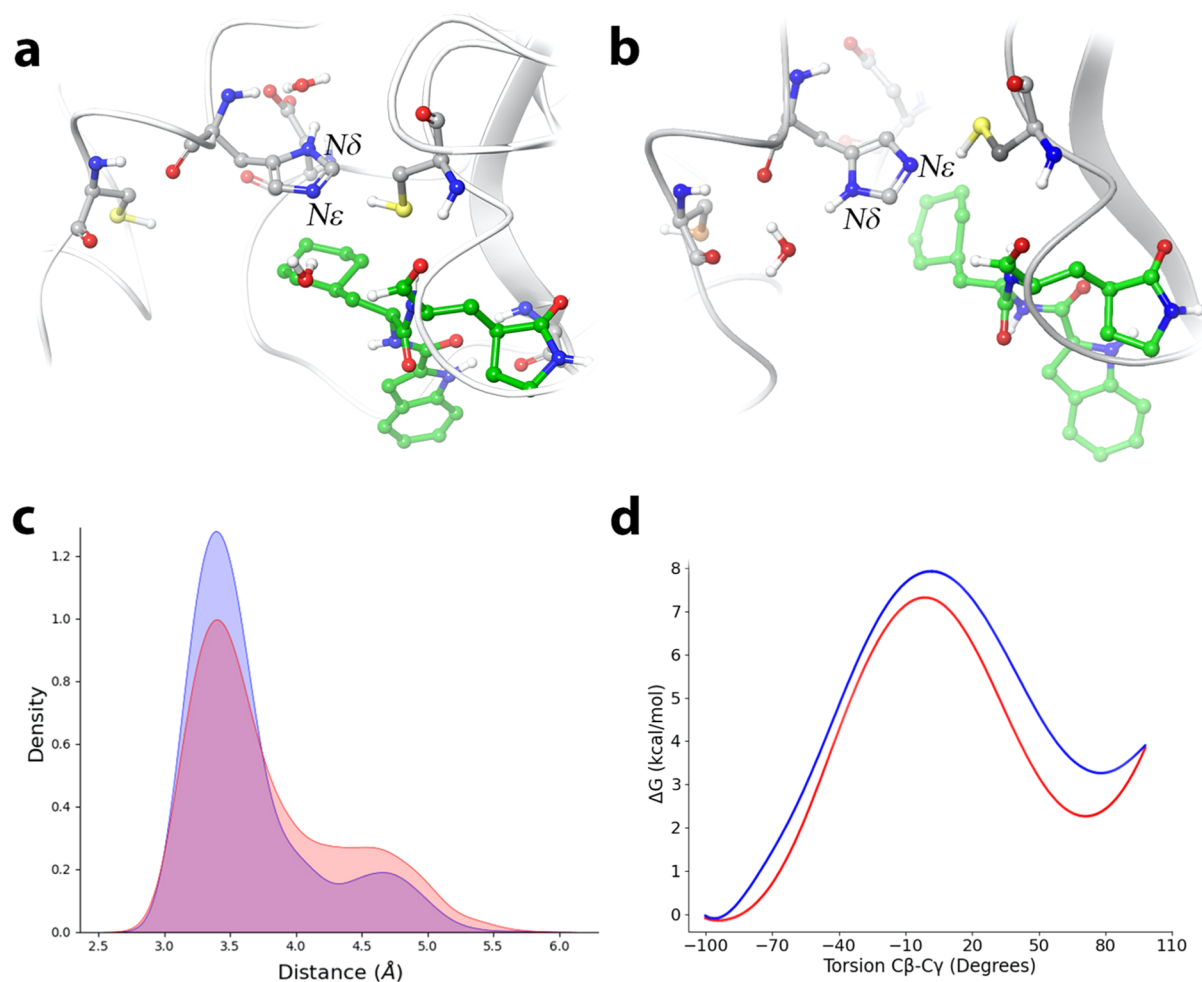


Figure 2. Rotameric state of the catalytic His41. (2a) Representation of the noncovalent complex of **11a** (carbon atoms in green) with SARS-CoV-2 3CL protease with His41 in the ϵ -rotameric state, (2b) representation of the complex presenting the δ -rotameric state of His41, and (2c) probability densities of the distances from the Cys145-S γ atom to the N ϵ atom of His41 for the ϵ - (blue) and δ -rotameric state (red). (2d) Free-energy profile associated to the rotation around the C γ -C β 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 resulting MD simulations of the noncovalent E-I complex (4 replicas of 1 μ s each) were stable in all cases (see RMSD time evolutions in Figure S2), 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 the 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 the 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.^{16,37} As presented above, in the case of the noncovalent E-I complex, these two residues are kept at hydrogen-bond distances during a significant fraction of the simulation (see Figure 2c for one replica; results for other replicas are presented in Figure S3). 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 S γ atom to the aldehyde carbon atom and the protonation of the oxygen atom. Figure 3c shows the distribution of distances between the Cys145 S γ 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 the 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 S γ 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 ± 0.3 Å), Ser144 (2.8

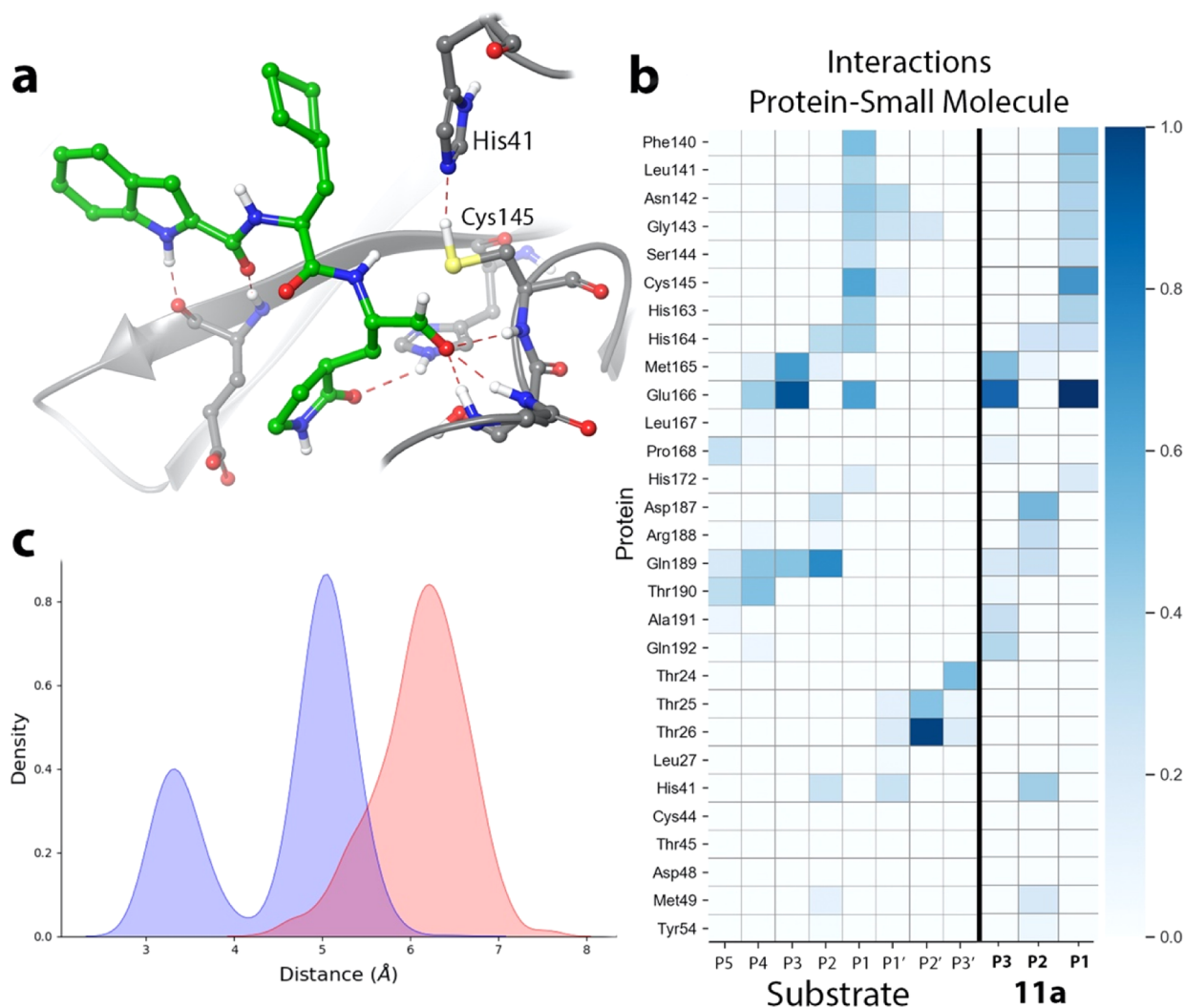


Figure 3. Noncovalent complex formed between the aldehyde derivative **11a** and the 3CL protease of SARS-CoV-2. (3a) Binding pose of the inhibitor in the active site of the protease, showing the location of the catalytic dyad and the oxyanion hole. Note that the aldehyde oxygen points to the oxyanion hole. (3b) Fraction of hydrogen-bond contacts between residues of **11a** and a peptide substrate¹⁶ and those of the protease. A hydrogen-bond contact is counted when the donor–acceptor distance is <3.8 Å and the hydrogen-bond angle is $>120^\circ$. (3c) Probability densities of the distances from the Cys145-S γ atom to the C carbon atom of the substrate, in blue, and from the N ϵ atom of His41 to the aldehyde oxygen atom, in red.

± 0.4 Å), and Gly143 (2.3 ± 0.3 Å), as seen in Figure 3a. These three residues form the oxyanion hole, placed in a U-turn of the loop connecting $\beta 10$ and $\beta 11$ 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 N ϵ atom of His41, as confirmed by the probability distribution of N ϵ –O distances, peaked at 6.3 Å (see Figure 3c). This separation precludes a direct proton transfer between these two atoms after the 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.

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 toward the catalytic His41 instead of toward the oxyanion hole and the distance to the N ϵ 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 obtained for the rotation of the aldehyde group around the C–C(O) bond in the active site. First, in order to test the reliability of our parametrization to describe this rotation of the substrate, we performed gas-phase

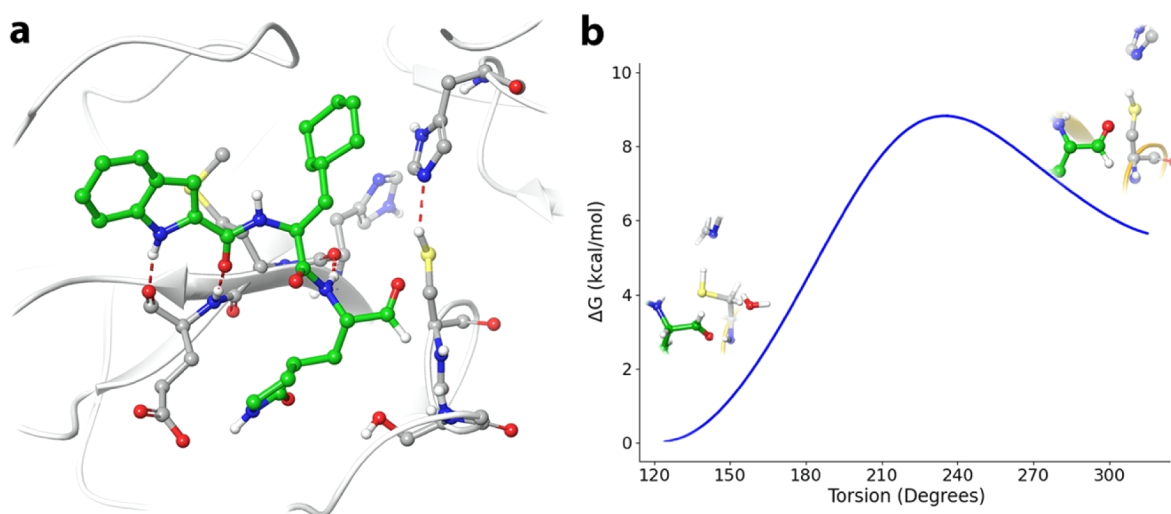


Figure 4. Orientation of the aldehyde group in the noncovalent complex. (4a) Representation of the E–I complex with inhibitor 11a in the pro-(R) conformation. (4b) Classical free-energy profile for the rotation of the aldehyde group from the pro-(S) conformer (left) to the pro-(R) conformer (right).

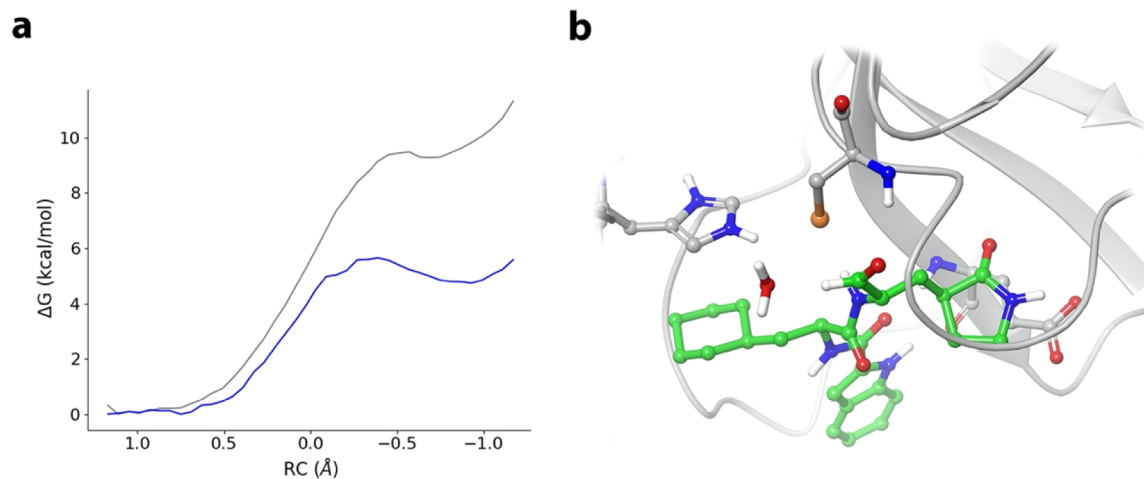


Figure 5. Proton transfer from Cys145 to His41 in SARS-CoV-2 3CL protease. (5a) Free-energy profile for the transformation from the neutral catalytic dyad (left) to the IP (IP, right) in the E–I complex with 11a (gray line) and in the complex with the peptide substrate¹⁶ (blue line). The reaction coordinate (RC) is the antisymmetric combination $d(S\gamma-H)-d(N\epsilon-H)$. (5b) IP structure with 11a showing the presence of a water molecule in between the protonated histidine and the inhibitor aldehyde group.

B3LYPD3/6-31+G* single-point energy calculations on minimum energy structures extracted from a dihedral potential energy scan. In the gas phase, the pro-(R) structure was found to be $1.1 \text{ kcal}\cdot\text{mol}^{-1}$ more stable than the pro-(S) because of an intramolecular interaction between the aldehyde oxygen and the amide NH group, while at the MM level, the energy difference was found to be very similar, $1.6 \text{ kcal}\cdot\text{mol}^{-1}$. This calculation confirms the adequacy of the forcefield to represent the isomerization process. Then, we obtained the isomerization free-energy profile in the enzyme (see Figure 4b), as described in the Methodology section. According to this, the pro-(S) conformation is $5.5 \text{ kcal}\cdot\text{mol}^{-1}$ more stable than the pro-(R) conformer in the enzyme because in the first case, the aldehyde oxygen atom is accommodated in the oxyanion hole, while in the second one, the aldehyde oxygen atom points toward His41. It must be noticed that rotation of the aldehyde group of the inhibitor in the enzymatic active site could be facilitated after the formation of the IP, when the N ϵ atom of His41 is protonated. This rotation would explain that the (R)-

hemithioacetal product can also be formed, as discussed above. Note 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 α -ketoamide inhibitor where the aldehyde hydrogen is substituted by an acetamide group. In this case, the X-ray structure (PDB code 7BRP)⁴¹ 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.

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 proceeds. We traced the QM/MM free-energy profile for the proton transfer from Cys145 to His41 using as a

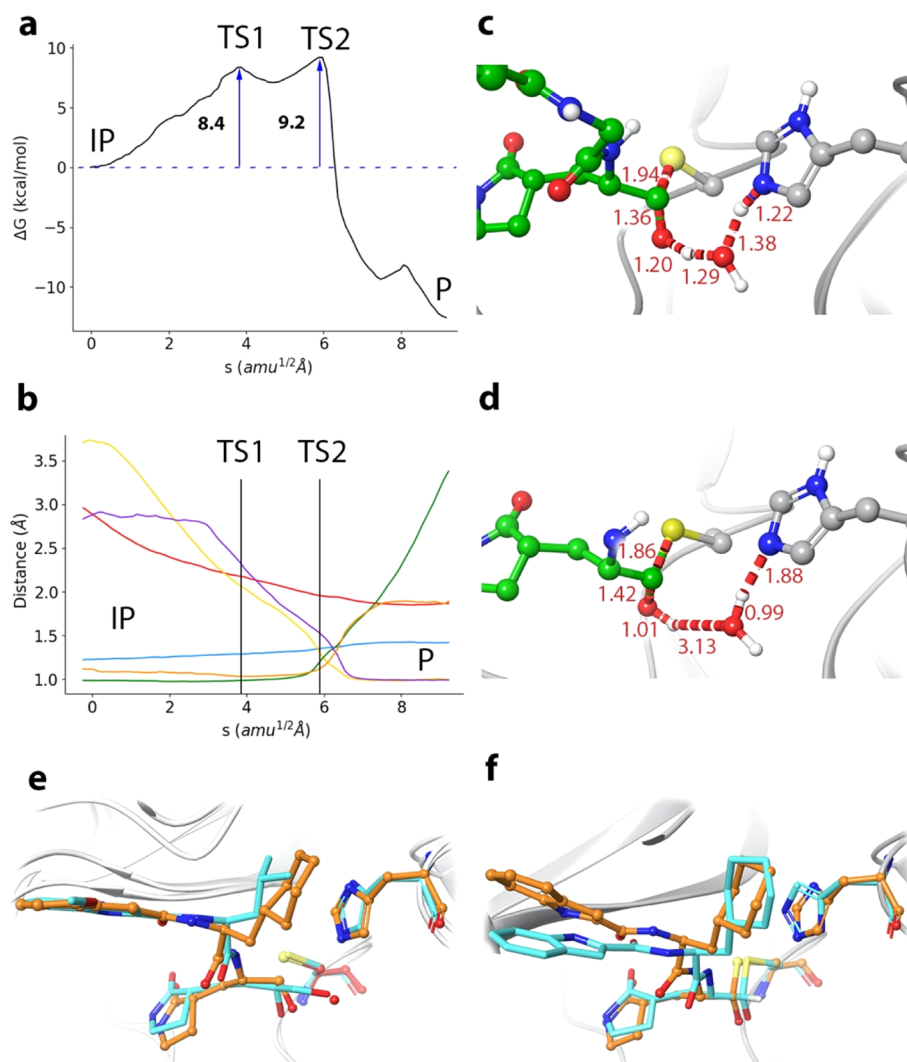


Figure 6. Formation of the (*S*)-hemithioacetal product from the IP. (6a) B3LYPD3/6-31+G*/MM free-energy profile along the path-CV for the formation of the covalent E–I complex from the IP. (6b) Evolution of the selected CVs along the MFEP. The color code corresponds to Scheme 1. (6c) Representation of the rate-determining TS (TS2). The values of the distances correspond (in Å) to the coordinates of the MFEP where the TS is located. (6d) Representation of the (*S*)-hemithioacetal complex. (6e) Overlap of the TS2 structure (balls and 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 and sticks with carbon atoms in orange) with the X-ray structure 6LZE (licorice with carbon atoms in light blue).

distinguished coordinate the antisymmetric combination of proton distances to the donor and acceptor atoms [$d(\text{S}\gamma\text{--H})\text{--}d(\text{N}\epsilon\text{--H})$, see Figure 5a]. Starting from the E–I 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 the IP under the presence of other inhibitors: 7.3 kcal·mol^{−1} with an α -ketoamide inhibitor²⁰ and 10.3 with a Michael acceptor.¹⁸ These values 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 studies,^{16,20} 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 the formation of the IP, the inhibition reaction must proceed with the attack of the activated nucleophile (the *S*

atom of Cys145) on the aldehyde carbon atom and the proton transfer from the N ϵ 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.

Starting from the structure presented in Figure 5b, we obtained the corresponding QM/MM MFEP for the formation of the (*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 *S*

atom to the aldehyde carbon atom, reducing the $S\gamma$ -C distance from 2.95 to 2.18 Å at the first transition state (TS1, shown in Figure S4). 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 6c) where the formation of the S–C bond (presenting a distance of 1.94 Å) is accompanied by 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 His41 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 $S\gamma$ -C bond (1.85 Å) and 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 as 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 the 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).

Combining the free-energy profiles presented in Figures 5 and 6, we can obtain a full picture of the transformation from the noncovalent E–I complex formed between the SARS-CoV-2 protease and the inhibitor 11a to the covalent E–I complex, the (S)-hemithioacetal. A representation of the complete free-energy profile is given in Figure 7. 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 E–I 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),¹⁸ 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

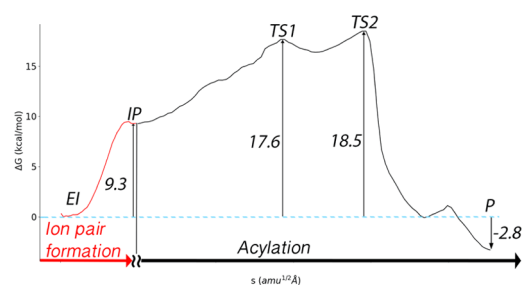


Figure 7. Representation of the complete free-energy profile for the inhibition of 3CL SARS-CoV-2 protease with 11a, including IP formation and the formation of the covalent E–I complex (product P in the figure).

more reversible than that 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.⁴²

Regarding the kinetics of the process, the first-order rate constant for covalent inhibition of the main protease by 11a is determined by the free-energy difference between the rate-determining states,⁴³ the highest energy TS (TS2) and the noncovalent E–I complex. Accordingly, our free-energy simulations provide the activation free energy as the sum of two contributions: the free-energy cost of forming the reactive IP from E–I (9.3 kcal·mol⁻¹, Figure 5a) plus the free energy of the TS2 relative to IP (9.2 kcal·mol⁻¹, Figure 6a). This gives in a total activation free energy of 18.5 kcal·mol⁻¹, as shown in Figure 7. While the inhibition rate constant of SARS-CoV-2 protease by 11a has not been experimentally determined, the rate constant for the inhibition of the protease with GC376 (the prodrug of the aldehyde inhibitor GC373) has been measured to be $2.45 \times 10^{-3} \text{ s}^{-1}$ at 30 °C.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 E–I 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 $C\beta$ – $C\gamma$ bond. Both for the apo form and for the noncovalent complex formed between the 3CL protease and

the **11a** inhibitor, we found that the rotamer where the N ϵ 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 the formation of the catalytic dyad IP, 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 S γ 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 E–I complex. After IP 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 concordance 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 the 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.

■ ASSOCIATED CONTENT

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acscatal.0c05522>.

Rotameric states of His41 in X-ray structures of SARS-CoV-2 3CL protease; converged CVs obtained with different time steps in the string method; RMSD plots of MD simulations; probability distributions of distances obtained for the 4 MD replicas; and representation of TS1 (PDF).

■ AUTHOR INFORMATION

Corresponding Authors

J. Javier Ruiz-Pernía – *Departamento de Química Física, Universidad de Valencia, Burjassot 46100, Spain;*

ORCID: orcid.org/0000-0002-4640-0419;

Email: ignacio.tunon@uv.es

Iñaki Tuñón – *Departamento de Química Física, Universidad de Valencia, Burjassot 46100, Spain;* ORCID: orcid.org/0000-0002-6995-1838; Email: jjavier.ruiz@uv.es

Author

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

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acscatal.0c05522>

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors acknowledge financial support from Feder funds and the Ministerio de Ciencia, Innovación y Universidades (project PGC2018-094852-B-C22), and Conselleria 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. Authors also deeply acknowledge Dr. Kirill Zinovjev for assistance in the adaptation of the code and helpful discussions on the use of the string method.

■ 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*, 1345–1352.
- (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. Rep.* **2020**, *6*, 56–70.
- (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*, 3177–3189.
- (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*, 6595–6628.
- (6) Brocklehurst, K.; Willenbrock, F.; Sauh, E. Cysteine Proteinases. *Hydrolytic Enzymes*; Elsevier: New York, 1987; pp 39–158.
- (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.

(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.; You, T.; Liu, X.; Yang, X.; Bai, F.; Liu, H.; Liu, 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*, 289–293.

(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* **2020**, *368*, 409–412.

(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* **2020**, *368*, 1331–1335.

(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*, 4282.

(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*, 12725–12747.

(13) Mole, J. E.; Horton, H. R. Kinetics of Papain-Catalyzed Hydrolysis of α -N-Benzoyl-L-Arginine-p-Nitroanilide. *Biochemistry* **1973**, *12*, 816–822.

(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*, 678–692.

(15) Berman, H.; Henrick, K.; Nakamura, H. Announcing the Worldwide Protein Data Bank. *Nat. Struct. Mol. Biol.* **2003**, *10*, 980.

(16) Ramos-Guzmán, C. A.; Ruiz-Pernía, J. J.; Tuñón, I. Unraveling the SARS-CoV-2 Main Protease Mechanism Using Multiscale Methods. *ACS Catal.* **2020**, *10*, 12544–12554.

(17) Swiderek, K.; Moliner, V. Revealing the Molecular Mechanisms of Proteolysis of SARS-CoV-2 M pro by QM/MM Computational Methods. *Chem. Sci.* **2020**, *11*, 10626–10630.

(18) 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. *Chem. Sci.*, [Online early access] DOI: 10.1039/D0SC04978F. Published online Jan 29, 2021. <https://doi.org/10.1039/D0SC04978F>.

(19) Arafet, K.; Serrano-Aparicio, N.; Lodola, A.; Mulholland, A. J.; González, F. V.; Swiderek, K.; Moliner, V. Mechanism of Inhibition of SARS-CoV-2 M pro by N3 Peptidyl Michael Acceptor Explained by QM/MM Simulations and Design of New Derivatives with Tunable Chemical Reactivity. *Chem. Sci.* **2021**, *12*, 1433–1444.

(20) Mondal, D.; Warshel, A. Exploring the Mechanism of Covalent Inhibition: Simulating the Binding Free Energy of α -Ketoamide Inhibitors of the Main Protease of SARS-CoV-2. *Biochemistry* **2020**, *59*, 4601–4608.

(21) 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*, 247–260.

(22) 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. University of California: San Francisco, 2018.

(23) 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*, 10269–10280.

(24) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.

(25) Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theory Comput.* **2011**, *7*, 525–537.

(26) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.

(27) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An $N \log(N)$ Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(28) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.

(29) Le Grand, S.; Götz, A. W.; Walker, R. C. SPFP: Speed without Compromise—A Mixed Precision Model for GPU Accelerated Molecular Dynamics Simulations. *Comput. Phys. Commun.* **2013**, *184*, 374–380.

(30) Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9*, 3878–3888.

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

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

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

(34) Becke, A. D. Density-functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

(35) 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.

(36) 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.

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

(38) Zinoviev, K. String-Amber <https://github.com/kzinoviev/string-amber> (accessed Jun 24, 2020).

(39) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.;

Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16*, Revision C.01, 2016.

(40) Kufareva, I.; Ilatovskiy, A. V.; Abagyan, R. Pocketome: An Encyclopedia of Small-Molecule Binding Sites in 4D. *Nucleic Acids Res.* **2011**, *40*, D535–D540.

(41) 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*, 4417.

(42) Drag, M.; Salvesen, G. S. Emerging Principles in Protease-Based Drug Discovery. *Nat. Rev. Drug Discovery* **2010**, *9*, 690–701.

(43) Kozuch, S.; Shaik, S. How to Conceptualize Catalytic Cycles? The Energetic Span Model. *Acc. Chem. Res.* **2011**, *44*, 101–110.

Supporting Information

Multiscale Simulations of SARS-CoV-2 3CL Protease Inhibition with Aldehyde Derivatives. Role of Protein and Inhibitor Conformational Changes 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

Table S1. Rotameric state of the catalytic histidine in different x-ray structures	S2
Figure S1. Evolution of CVs in the string method as a function of the time step.	S5
Figure S2. RMSD values	S6
Figure S3. Probability distributions of distances	S7
Figure S4. Representation of TS1	S8
References	S9

Table S1. Rotameric state of the catalytic histidine (His41) and enantiomeric form of the hemithioacetal product in different x-ray structures.

PDB code	Protomer A	Protomer B	inhibitor	Enantiomer for aldehyde derived hemithioacetals
5R7Y	δ -rotamer	δ -rotamer	Z45617795	
5R7Z	ϵ -rotamer	ϵ -rotamer	Z1220452176	
5R80	ϵ -rotamer	ϵ -rotamer	Z18197050	
5R81	ϵ -rotamer	ϵ -rotamer	Z1367324110	
5R82	ϵ -rotamer	ϵ -rotamer	Z219104216	
5R83	ϵ -rotamer	ϵ -rotamer	Z44592329	
5R84	ϵ -rotamer	ϵ -rotamer	Z31792168	
5RE4	ϵ -rotamer	ϵ -rotamer	Z1129283193	
5REB	ϵ -rotamer	ϵ -rotamer	Z2856434899	
5REJ	ϵ -rotamer	ϵ -rotamer	PCM-0102241	
5REL	ϵ -rotamer	ϵ -rotamer	PCM-0102340	
5REM	ϵ -rotamer	ϵ -rotamer	PCM-0103016	
5REN	ϵ -rotamer	ϵ -rotamer	PCM-0102425	
5REO	ϵ -rotamer	ϵ -rotamer	PCM-0102578	
5RER	ϵ -rotamer	ϵ -rotamer	PCM-0102615	
5RES	ϵ -rotamer	ϵ -rotamer	PCM-0102281	
5RET	ϵ -rotamer	ϵ -rotamer	PCM-0102269	
5REU	ϵ -rotamer	ϵ -rotamer	PCM-0102395	
5REV	ϵ -rotamer	ϵ -rotamer	PCM-0103072	
5REW	δ -rotamer	δ -rotamer	PCM-0102275	
5REX	ϵ -rotamer	ϵ -rotamer	PCM-0102287	
5REY	ϵ -rotamer	ϵ -rotamer	PCM-0102911	
5REZ	ϵ -rotamer	ϵ -rotamer	POB0129	
5RF1	ϵ -rotamer	ϵ -rotamer	NCL-00023830	
5RF3	ϵ -rotamer	ϵ -rotamer	Z1741970824	
5RF6	δ -rotamer	δ -rotamer	Z1348371854	
5RF7	ϵ -rotamer	ϵ -rotamer	Z316425948_minor	
5RFE	ϵ -rotamer	ϵ -rotamer	Z509756472	
5RFF	ϵ -rotamer	ϵ -rotamer	PCM-0102704	
5RFG	ϵ -rotamer	ϵ -rotamer	PCM-0102372	
5RFH	ϵ -rotamer	ϵ -rotamer	PCM-0102277	
5RFI	ϵ -rotamer	ϵ -rotamer	PCM-0102353	
5RFJ	ϵ -rotamer	ϵ -rotamer	PCM-0103067	
5RFK	ϵ -rotamer	ϵ -rotamer	PCM-0102575	

5RFL	ϵ -rotamer	ϵ -rotamer	PCM-0102389	
5RFM	ϵ -rotamer	ϵ -rotamer	PCM-0102539	
5RFN	ϵ -rotamer	ϵ -rotamer	PCM-0102868	
5RFO	ϵ -rotamer	ϵ -rotamer	PCM-0102972	
5RFP	ϵ -rotamer	ϵ -rotamer	PCM-0102190	
5RFQ	ϵ -rotamer	ϵ -rotamer	PCM-0102179	
5RFT	ϵ -rotamer	ϵ -rotamer	PCM-0102432	
5RFU	ϵ -rotamer	ϵ -rotamer	PCM-0102121	
5RFV	ϵ -rotamer	ϵ -rotamer	PCM-0102306	
5RFW	ϵ -rotamer	ϵ -rotamer	PCM-0102243	
5RFX	ϵ -rotamer	ϵ -rotamer	PCM-0102254	
5RFY	ϵ -rotamer	ϵ -rotamer	PCM-0102974	
5RFZ	ϵ -rotamer	ϵ -rotamer	PCM-0102274	
5RG0	ϵ -rotamer	ϵ -rotamer	PCM-0102535	
5RG1	δ -rotamer	δ -rotamer	NCL-00024905	
5RG2	ϵ -rotamer	ϵ -rotamer	NCL-00025058	
5RG3	ϵ -rotamer	ϵ -rotamer	NCL-00025412	
5RGH	ϵ -rotamer	ϵ -rotamer	Z1619978933 (Mpro-x0395)	
5RGI	ϵ -rotamer	ϵ -rotamer	Z369936976 (Mpro-x0397)	
5RGL	ϵ -rotamer	ϵ -rotamer	PCM-0102962 (Mpro-x0705)	
5RGM	ϵ -rotamer	ϵ -rotamer	PCM-0102142 (Mpro-x0708)	
5RGO	ϵ -rotamer	ϵ -rotamer	PCM-0102248 (Mpro-x0736)	
5RGT	ϵ -rotamer	ϵ -rotamer	Z4439011607 (Mpro-x2540)	
5RGY	ϵ -rotamer	ϵ -rotamer	Z1535580916 (Mpro-x2581)	
5RGU	ϵ -rotamer	ϵ -rotamer	Z4444622180 (Mpro-x2562)	
5RGV	ϵ -rotamer	ϵ -rotamer	Z4444622066 (Mpro-x2563)	
5RGW	ϵ -rotamer	ϵ -rotamer	Z4444621910 (Mpro-x2569)	
5RH0	ϵ -rotamer	ϵ -rotamer	Z1286870272	
5RH1	ϵ -rotamer	ϵ -rotamer	Z2010253653	
5RH5	ϵ -rotamer	ϵ -rotamer	Z4439011520	
5RH6	ϵ -rotamer	ϵ -rotamer	Z4439011588	
5RH7	ϵ -rotamer	ϵ -rotamer	Z4439011584	
5RH8	ϵ -rotamer	ϵ -rotamer	Z4444621965	
5RH9	ϵ -rotamer	ϵ -rotamer	Z4438424255	
5RH9	ϵ -rotamer	ϵ -rotamer	Z4438424255 (Mpro-x2776)	
6LU7	ϵ -rotamer	ϵ -rotamer	N3	
6LZE	δ -rotamer	δ -rotamer	11a aldehyde	S(2)
6M0K	δ -rotamer	δ -rotamer	11b aldehyde	S(2)
6M2N	δ -rotamer	δ -rotamer	Baicalein	
6M2Q	ϵ -rotamer	ϵ -rotamer	Apo	
6W63	δ -rotamer	δ -rotamer	X77	
6WNP	ϵ -rotamer	ϵ -rotamer	Boceprevir	

6WTJ	ϵ -rotamer	ϵ -rotamer	GC-373	S(2)
6WTK	ϵ -rotamer	ϵ -rotamer	GC-373	S(2)
6WTT	ϵ -rotamer	ϵ -rotamer	GC-373	S (2) and R (1)
6XHM	δ -rotamer	ϵ -rotamer	PF-00835231	
6XA4	δ -rotamer	δ -rotamer	UAW241	S(2)
6Y2F	ϵ -rotamer	ϵ -rotamer	13b alpha-ketoamide	
6Y2G	ϵ -rotamer	ϵ -rotamer	13b alpha-ketoamide	
6YNQ	δ -rotamer	δ -rotamer	2-Methyl-1-tetralone	
6YZ6	ϵ -rotamer	ϵ -rotamer	Leupeptin	
7BQY	ϵ -rotamer	ϵ -rotamer	N3	
7BRP	ϵ -rotamer	ϵ -rotamer	Boceprevir	
7BUY	δ -rotamer	δ -rotamer	Carmofur	
7JP1	ϵ -rotamer	ϵ -rotamer	Apo form	
7K3T	ϵ -rotamer	ϵ -rotamer	Apo form	
7KHP	ϵ -rotamer	ϵ -rotamer	C-terminal sequence	

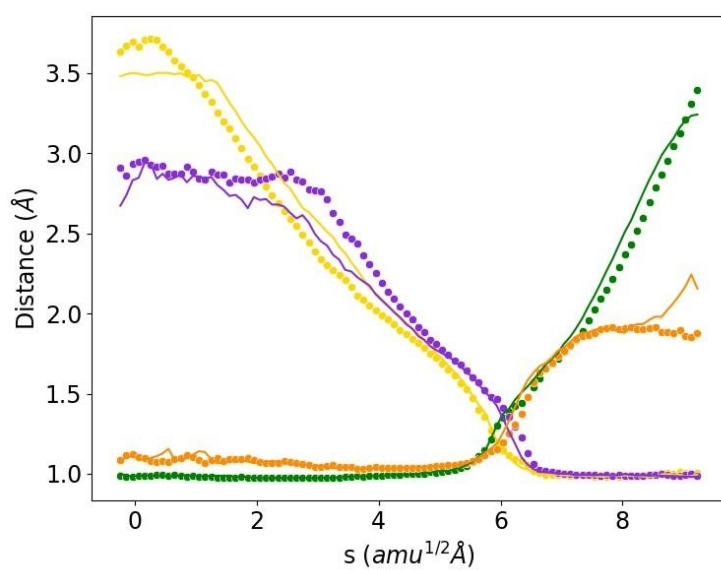


Figure S1. Plot of the evolution of the string-converged CVs involving hydrogen atoms (CV3 to CV6, as shown in Scheme 1) obtained using a simulation time step of 0.5 fs (continuous lines) or 1 fs (symbols). The color code corresponds to the same used in Scheme 1 of the main text.

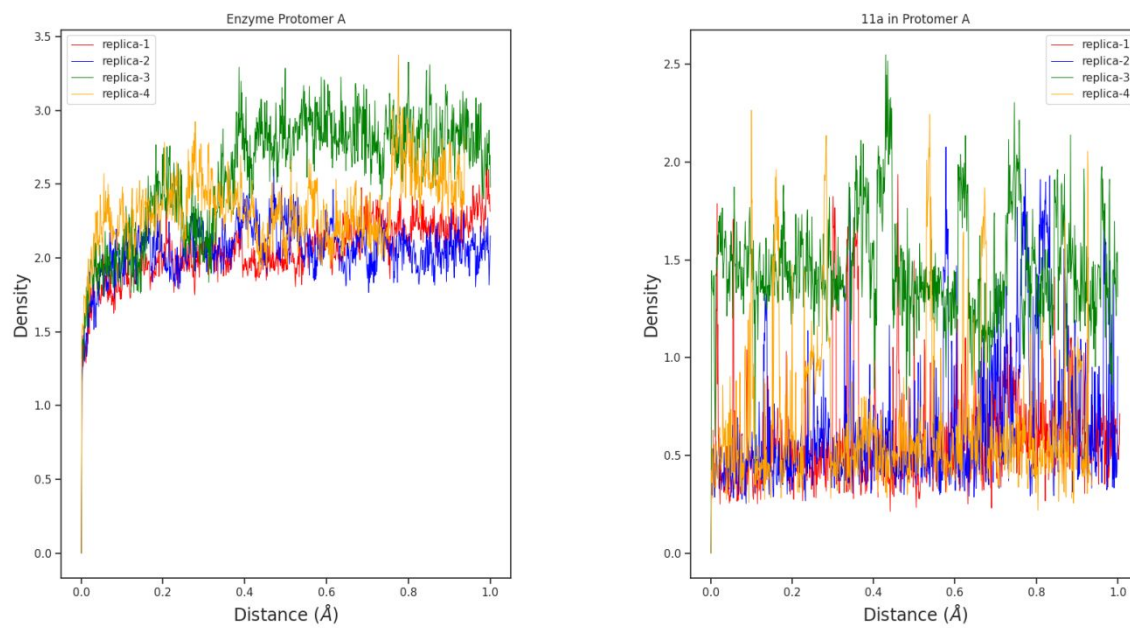


Figure S2. Plot of the RMSD values obtained for the protein (left) and the substrate (right) during the 4 replicas (simulation time of 1.0 μ s) of the noncovalent enzyme-inhibitor complex of the SARS-CoV-2 protease with **11a**. Results correspond to protomer A

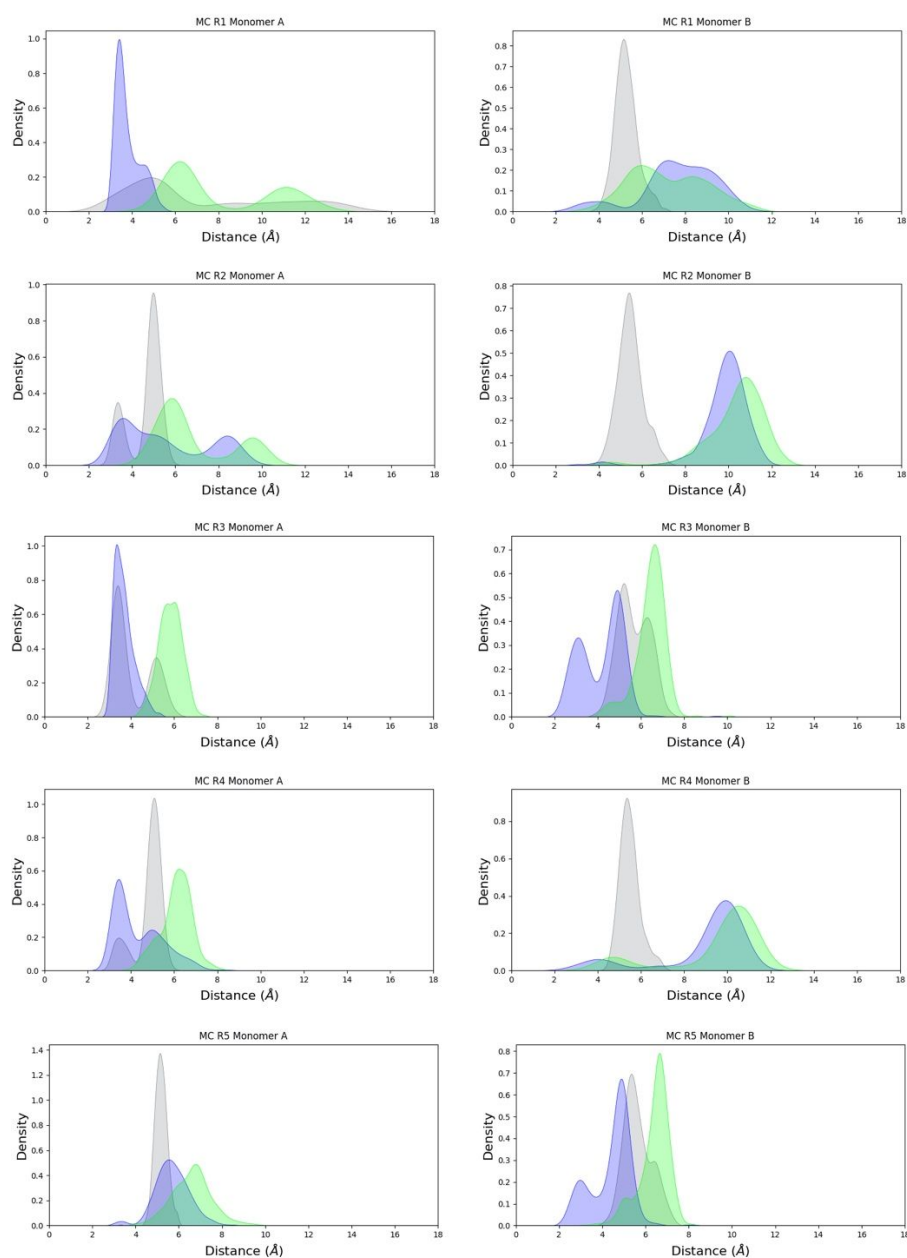


Figure S3. Probability distributions of distances obtained for the 4 replicas in protomer A (left) and B (right). $S\gamma$ - C_{aldh} distance in grey, $S\gamma$ - $N\epsilon$ in blue and $N\epsilon$ - O_{aldh} in green. Results for protomer A are quite similar in all the replicas. However, results for protomers B differ significantly, with the inhibitor moving far from the catalytic dyad in some cases and a substantial increase in the distance between Cys145 and His41 in others. The differences observed between the two protomers are consistent with the expected asymmetry between the two protomers and the fact that only one of them is active at a time.¹

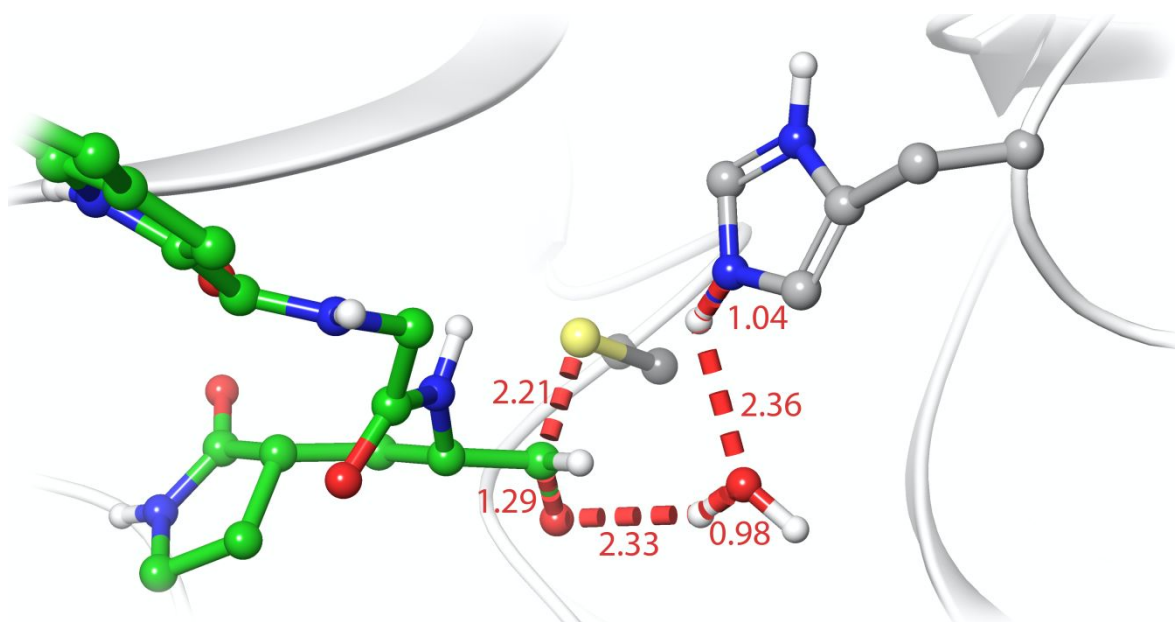


Figure S4. Representation of TS1 for the reaction mechanism of formation of the (S)-hemithioacetal formed by 3CL protease of SARS-CoV-2 and inhibitor **11a**. Distances in Å.

References

1. Chen, H.; Wei, P.; Huang, C.; Tan, L.; Lai L. Only one protomer is active in the dimer of SARS 3C-like proteinase *J. Biol. Chem.* **2006**, *281*, 13894-13898

Appendix 5

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



Cite this: DOI: 10.1039/d1cc03953a

 Received 21st July 2021,
 Accepted 10th August 2021

DOI: 10.1039/d1cc03953a

rsc.li/chemcomm

Computational simulations on the binding and reactivity of a nitrile inhibitor of the SARS-CoV-2 main protease†

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

We present a detailed computational analysis of the binding mode and reactivity of the novel oral inhibitor PF-07321332 developed against the SARS-CoV-2 3CLpro protease. Alchemical free energy calculations suggest that positions P3 and P4 could be susceptible to improvement in order to get a larger binding strength. QM/MM simulations unveil the reaction mechanism for covalent inhibition, showing that the nitrile warhead facilitates the recruitment of a water molecule for the proton transfer step.

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 that is in charge of the proteolysis of the polyproteins produced after translation of its genome in the infected cells.¹ 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 compounds.² Some of the compounds reported to have an inhibitory capacity against SARS-CoV-2 3CLpro are covalent inhibitors that contain an electrophilic site able to react with the catalytic cysteine, Cys145.^{3–7}

Last March, Pfizer announced that one of the antiviral compounds developed by this company against SARS-CoV-2, PF-07321332, went under clinical trial.⁸ PF-07321332 (see Fig. 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.⁹ The main advantage of

PF-07321332 with respect to other inhibitors of 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 γ -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 inhibitors developed against SARS-CoV-2 3CLpro, taking advantage of the selectivity of this enzyme using a glutamine residue in the equivalent position of the cleaved polyprotein. A bicycloproline moiety, also present in boceprevir and other 3CLpro inhibitors¹⁰ is present at the P2 position, which takes into account the preference of the enzyme for a hydrophobic leucine residue at P2 in the natural substrate. Larger variations are observed among different inhibitors regarding the groups present at the P3 and P4 positions (see Fig. S1, ESI†). At the P3 position PF-07321332 presents a *tert*-leucine residue, while at P4 the substituent is a trifluoromethyl group.

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 thioimidate complex (see Fig. 1). We explored the binding mode of the noncovalent complex (EI) formed by the PF-07321332 inhibitor in the active site of SARS-CoV-2 3CLpro by means of Molecular Dynamics (MD) simulations. Alchemical transformations were then used to evaluate the contribution of the different P_i groups of the inhibitor to the binding free energy of the EI complex, with the goal of providing a free

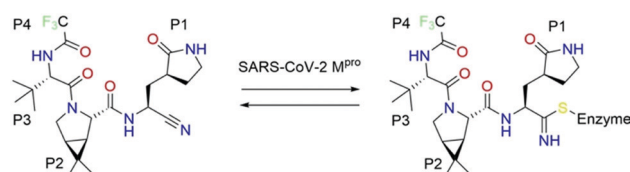


Fig. 1 Chemical structure of the PF-07321332 inhibitor and the complex formed after reaction with the SARS-CoV-2 main protease. P_i sites are shown.

Departamento de Química Física, Universidad de Valencia, 46100 Burjassot, Spain.

E-mail: j.javier.ruiz@uv.es, ignacio.tunon@uv.es

† Electronic supplementary information (ESI) available: Computational details, table of results of alchemical transformations and additional figures. See DOI: 10.1039/d1cc03953a

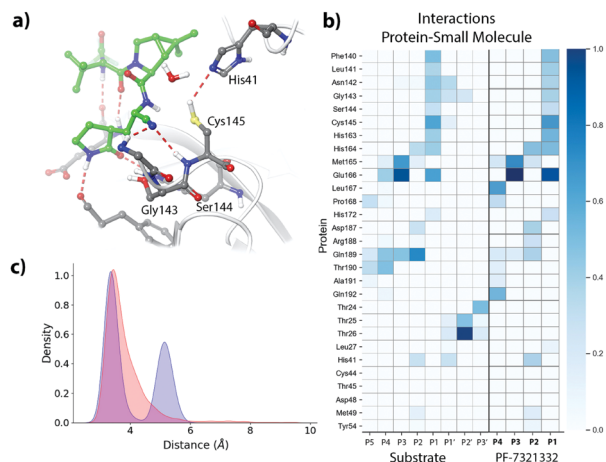


Fig. 2 Noncovalent complex formed between PF-07321332 and the 3CL protease of SARS-CoV-2. **(2a)** Binding pose of the inhibitor in the active site of the protease, showing the location of the catalytic dyad and the oxyanion hole. **(2b)** Fraction of contacts between the residues of PF-07321332 and a peptide substrate¹³ and the protease. A contact is counted when the donor-acceptor distance is < 3.8 Å and the hydrogen bond angle is $> 120^\circ$. **(2c)** Pair distribution function of the distances between the S_γ atom of Cys145 and the electrophilic atom of the inhibitor (blue) and the N_ϵ atom of His41 (red).

energy map as a guide for the design of future generations of the inhibitor. Then, we explored the free energy profile associated with the reaction of the inhibitor with Cys145 to form the covalent complex ($E-I$) characterizing the reaction mechanism and the associated activation free energy.

MD simulations (see the ESI† for methodological details) of the noncovalent complex were started from a binding pose derived from the X-ray structure of the related inhibitor PF-00835231 (PDB code: 6XHM).⁷ 3 MD replicas of 1 μ s each show that the binding pose of the PF-07321332 inhibitor is very similar to that observed for related aldehydes or hydroxymethylketone inhibitors.^{11,12} The nitrile group is placed in the so-called oxyanion hole, formed by interactions with the main chain NH groups of Cys145 (2.79 ± 0.57 Å), Ser144 (3.46 ± 0.46 Å) and Gly143 (2.54 ± 0.34 Å), as seen in Fig. 2a. Fig. 2b compares the fraction of contacts established between the inhibitor and the enzyme with those obtained from MD simulations of a peptide substrate with the sequence Ac-Ser-Ala-Val-Leu-Gln|Ser-Gly-Phe-NMe (the vertical line corresponds to the cleavage site).¹³ The P1 γ -lactam ring of the inhibitor almost perfectly mimics the interactions established by the P1 glutamine residue of the peptide. As said before, the presence of glutamine just before the cleavage site is a requirement in SARS-CoV and SARS-CoV-2 3CLpro¹⁴ and then the presence of the γ -lactam ring at this position is a 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 the 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, the main chain N and O atoms are hydrogen bonded to the 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 to contact with Glu166, Gln189 and Thr190. In general, the binding mode observed here for the PF-07321332 inhibitor is in very good agreement with that obtained previously using supervised MD simulations.¹⁵

Regarding the catalytic dyad, the distance between the Cys145- S_γ atom and the electrophilic carbon atom of the inhibitor is 4.05 ± 0.88 Å. The pair distribution functions (see Fig. 2c, blue) shows two peaks corresponding to the *trans* and *gauche* conformers of the Cys145 side chain. This residue remains hydrogen bonded to His41 most of the time, as reflected by the pair distribution distance between the Cys145- S_γ and His41- N_ϵ atoms (Fig. 2c, red) and the average distance (3.82 ± 0.76 Å). 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.^{11,16}

Previous descriptions of the contacts established between the PF-07321332 inhibitor and enzymatic residues in the active site of SARS-CoV-2 3CLpro stress the role of the 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 Fig. 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 Fig. 3 with their standard deviation, while the values obtained for each replica are shown in Table S1 (ESI†). As can be observed, the free energy changes obtained from our simulations indicate that all the P_i groups of

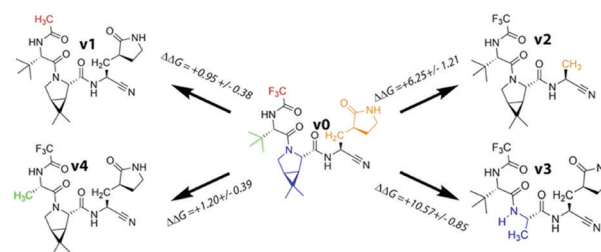


Fig. 3 Binding free energy changes associated with the transformation of different P_i groups of the PF-07321332 inhibitor into methyl groups.

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⁻¹, respectively. These large values clearly indicate the preference of the enzyme for these two groups and the difficulties of 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 by 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⁻¹, respectively. Comparison among structurally similar inhibitors (see Fig. S1, ESI†) suggests that these positions could be changed to improve the binding strength.

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_γ atom of the catalytic cysteine and the electrophilic carbon atom of the nitrile group. This process requires activation of the nucleophile by means of a proton transfer from the thiol group to the neighboring 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^{11,16} or a hydroxyl group.¹² In the case of the present inhibitor, we also found a water molecule 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)¹⁷ 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 (see the ESI† for a discussion of the QM description). The MM region explicitly includes the full protein and solvent. The MFEP was determined through biased MD simulations that explored 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 (see Fig. 4 and the ESI†). 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 the ESI† for details). The resulting free energy profile and the evolution of the CVs along the reaction is shown in Fig. 4a and b. 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_γ-C distance decreases from 3.0 to 2.1 Å) and its separation from the positively charged His41 (the distance between the S_γ atom and the transferred proton increases from 1.8 to 2.9 Å). The free energy profile shows a consistent increase associated with the separation of the

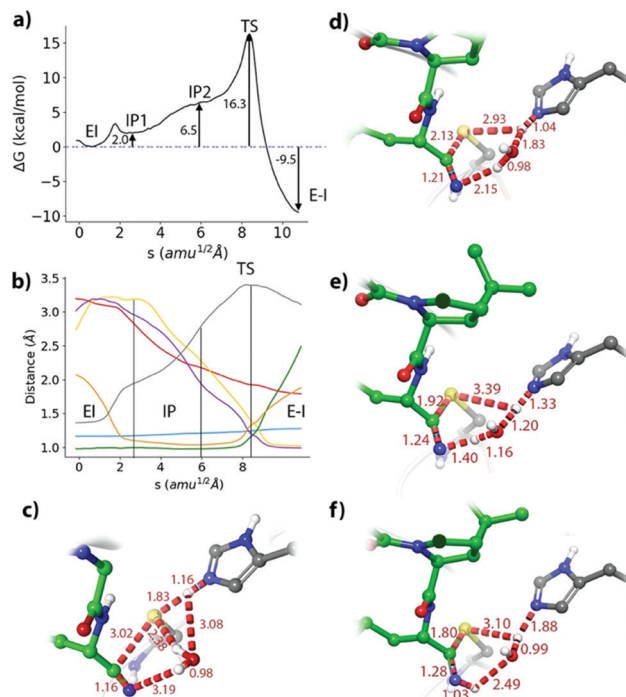


Fig. 4 Formation of the thioimide product (**4a**) B3LYPD3/6-31 + G*/MM free energy profile along the path-CV for the formation of the covalent *E-I* complex from the noncovalent (EI) one. (**4b**) Evolution of the selected CVs along the MFEP. Blue: C–N, orange: H–N_e, green: H_w–O_w, black: S_γ–H, yellow: N–H_w, red: S_γ–C, purple: O_w–H. (**4c**) Representation of IP1. The values of the distances correspond (in Å) to the coordinates of the MFEP. (**4d**) Representation of IP2. (**4e**) Representation of TS. (**4f**) Representation of the thioimide complex.

catalytic dyad until a plateau region is reached (IP2, in Fig. 4a). From this plateau the reaction continues up to the rate-determining Transition State (TS), where the nucleophilic attack is almost completed (the S_γ-C distance is reduced to 1.9 Å, see Fig. 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 (Fig. 4e). Finally, at the product state the water-mediated proton transfer is completed and the thioimide product is formed, presenting a S_γ-C distance of 1.8 Å (Fig. 4f).

The catalytic ion pair IP1 in Fig. 4a is only 2.0 kcal mol⁻¹ above the EI complex. Significantly larger values were found for the IP obtained in the presence of other inhibitors: 9.3 kcal mol⁻¹ for the aldehyde 11a¹¹ and 10.3 kcal mol⁻¹ for the Michael acceptor N3.¹⁶ 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⁻¹, where the charge-separated state can be stabilized by water molecules.¹³ The larger stabilization observed for the IP in the case of the PF-07321332 inhibitor when compared to 11a or N3 is related to the 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 Fig. 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 Fig. S2, ESI†).

The activation free energy obtained for the inactivation of SARS-CoV-2 3CLpro with PF-07321332 is 16.3 kcal mol⁻¹. Although there are no 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⁻¹,¹⁸ while the value we determined theoretically for another aldehyde, 11a, was 18.5 kcal mol⁻¹.¹¹ 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⁻¹.¹⁶ 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 with this proton transfer (see Fig. 4e).

Regarding the reaction free energy, Fig. 4a shows that the covalent complex (*E-I*) is more stable than the noncovalent complex (EI), the free energy difference being -9.5 kcal mol⁻¹. The reaction free energy obtained for the covalent inhibition of 3CLpro by PF-07321332 is larger than that for inhibitors 11a (-2.8 kcal mol⁻¹)¹¹ and PF-00835231 (-6.2 kcal mol⁻¹),¹² but smaller than that for the N3 inhibitor (-15.4 kcal mol⁻¹).¹⁶ 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 that using Michael acceptors.

Our simulations have 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 γ -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, also lowering the activation free energy and then increasing 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 that with aldehyde or ketone derivatives.

The authors acknowledge financial support from Generalitat Valenciana (GVCOV19/Decreto180/2020) and the Barcelona Supercomputing Center for access to MareNostrum. We also acknowledge the use of the Tirant supercomputer at the Universitat de València and the support of Alejandro Soriano.

Conflicts of interest

The authors declare no conflict of interest.

Notes and 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, E. R. Davies, R. A. Ferre, K. S. Gajiwala, M. He, R. J. Hogan, K. F. 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, **371**, 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 M. Pavan, G. Bolcato, D. Bassani, M. Sturlese and S. Moro, *J. Enzyme Inhib. Med. Chem.*, 2021, **36**, 1646–1650.
- 16 C. A. Ramos-Guzmán, J. J. Ruiz-Pernía and I. Tuñón, *Chem. Sci.*, 2021, **12**, 3489–3496.
- 17 K. Zinovjev and I. Tuñón, *J. Phys. Chem. A*, 2017, **121**, 9764–9772.
- 18 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.

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

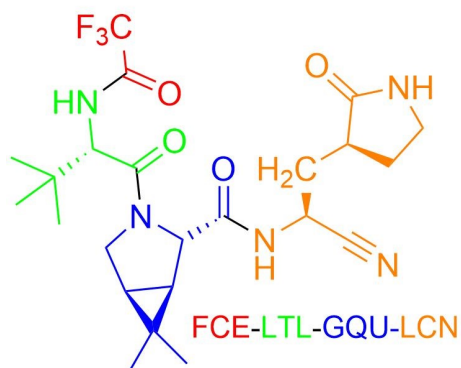
j.javier.ruiz@uv.es

Methodological details	S2
Table S1. Results for alchemical transformations	S5
Figure S1. Comparison of structurally similar inhibitors of SARS-CoV-2 protease	S6
Figure S2. Superposition of Ion Pair structures	S7
References	S8

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 program¹ from the AmberTools18² 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 Rios⁴ 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 6XHM⁵ 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 Maestro⁶ and the protonation states for the amino acid side chains in the enzyme were obtained with PROPKA3.0⁷ 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 of steepest descent algorithm followed by conjugate gradient minimization steps the RMS gradient was below 10^{-3} kcal·mol⁻¹Å⁻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⁻¹Å⁻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. Shake⁸ 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 al¹³ 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 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 (\square 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 .

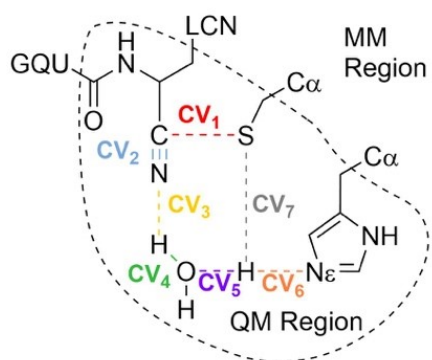
QM/MM calculations

To obtain the minimum free energy pathway (MFEP) along the chemical reaction the Adaptive 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 $0.2 \text{ amu}^{1/2} \cdot \text{\AA}$ 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 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 nucleophilic attack.¹⁹

To run the string a modified version of AMBER18²¹ with Gaussian16²² 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		$\Delta\Delta G_{\text{bind}}$				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		$\Delta\Delta G_{\text{bind}}$				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		$\Delta\Delta G_{\text{bind}}$				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		$\Delta\Delta G_{\text{bind}}$				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		

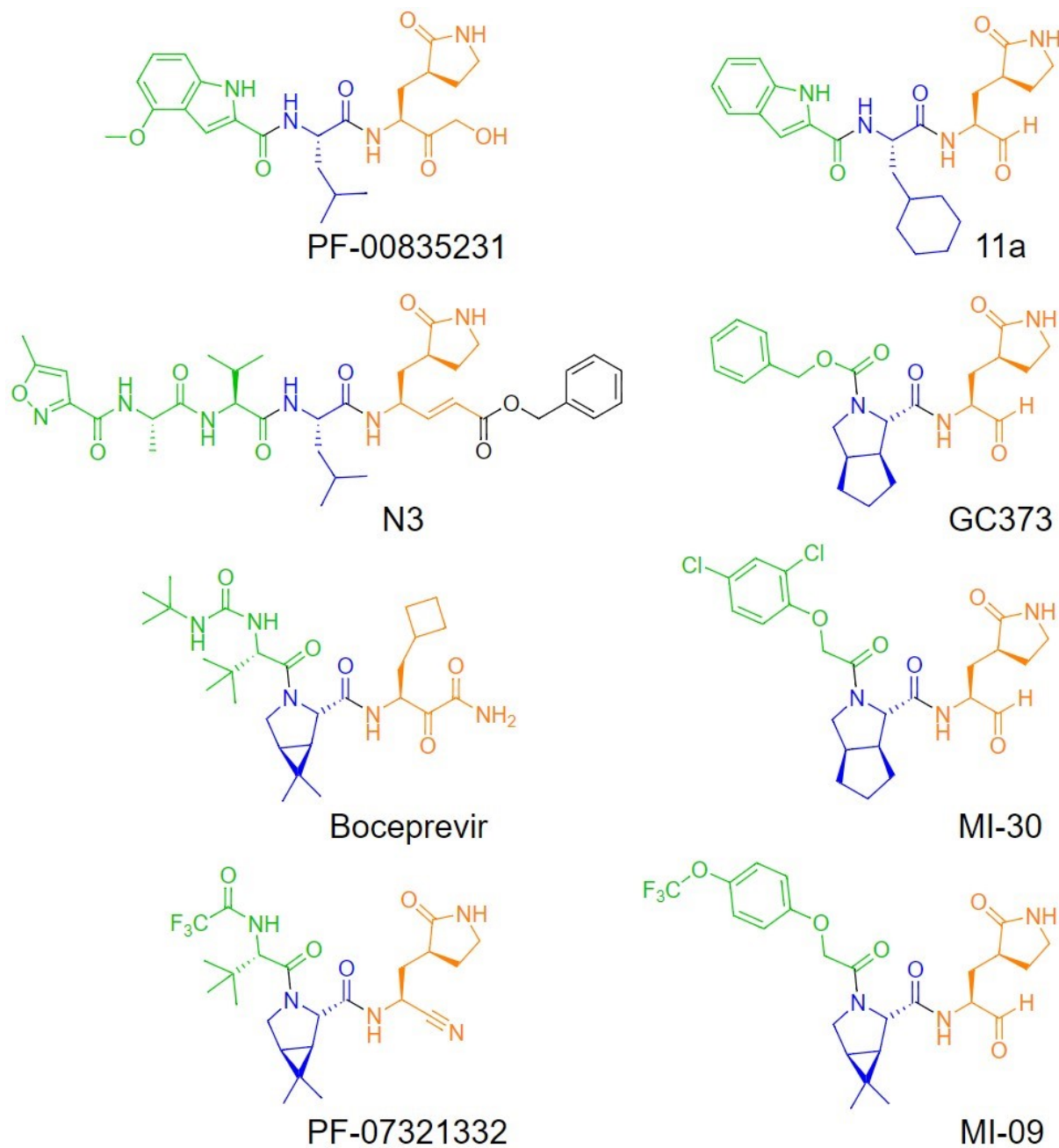


Figure S1. Some of the known inhibitors of the main protease of SARS-CoV-2. P1 group is shown in orange, P2 in blue and P3/P4 in green.

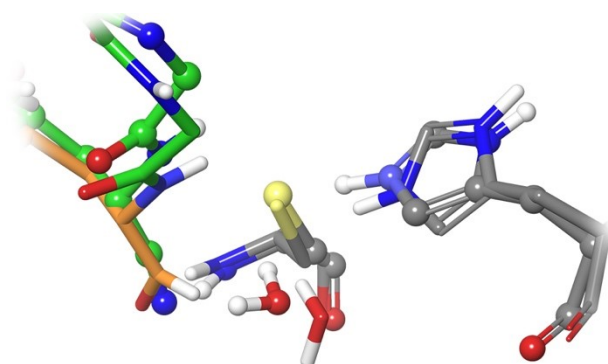


Figure S2. Overlap of the Ion Pair structures of the SARS-CoV-2 3CLpro active site obtained with PF-07321332 inhibitor (green, CPK) and 11a (orange, licorice). Note that the presence of the hydrogen atom bonded to the carbonyl electrophilic carbon in 11a hinders the approach of the water molecule.

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.

Appendix 6

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

**SARS-CoV-2 Inhibitors Hot Paper**

How to cite:

International Edition: doi.org/10.1002/anie.202110027

German Edition: doi.org/10.1002/ange.202110027

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,* and Iñaki Tuñón*

Abstract: We present the results of classical and QM/MM simulations for the inhibition of SARS-CoV-2 3CL protease by a hydroxymethylketone inhibitor, PF-00835231. In the non-covalent complex the carbonyl oxygen atom of the warhead is placed in the oxyanion hole formed by residues 143 to 145, while P1–P3 groups are accommodated in the active site with interactions similar to those observed for the peptide substrate. According to alchemical free energy calculations, the P1' hydroxymethyl group also contributes to the binding free energy. Covalent inhibition of the enzyme is triggered by the proton transfer from Cys145 to His41. This step is followed by the nucleophilic attack of the S_γ atom on the carbonyl carbon atom of the inhibitor and a proton transfer from His41 to the carbonyl oxygen atom mediated by the P1' hydroxyl group. Computational simulations show that the addition of a chloromethyl substituent to the P1' group may lower the activation free energy for covalent inhibition

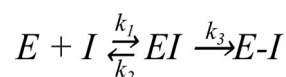
Introduction

Inhibition of the activity of the SARS-CoV-2 3CL protease (or main protease) is one of the therapeutic strategies to treat the COVID-19 pandemic. This enzyme is essential in the vital cycle of this and other related coronaviruses, being in charge of the cleavage of the long polyproteins resulting from the translation of the viral genome, in order to produce the non-structural proteins needed for virus replication.^[1] The 3CL or main protease of SARS-CoV-2 is a cysteine protease that uses a Cys-His catalytic dyad to hydrolyze peptide bonds at specific positions of the polyprotein chain. The proteolysis mechanism of the 3CL

protease involves two basic steps: i) the formation of a thiohemiacetal intermediate resulting from the attack of the S_γ atom of the catalytic cysteine at the carbonyl carbon atom of the peptide bond, which results in the release of the N-fragment and ii) the hydrolysis of the acyl-enzyme intermediate to release the C-terminal fragment and to recover the resting state of the active site (see Figure 1 a).^[2–5] The 3CL protease exclusively cleaves the polyprotein after a glutamine residue, a sequence specificity not shown by any known human protease,^[5] which may facilitate the development of antiviral drugs with reduced side effects.

The reactive cysteine present in 3CL protease is an attractive target for the development of covalent inhibitors of this enzyme, using as warheads functional groups able to form a chemical bond with the S_γ atom.^[6] Examples of warheads already tested in inhibitors of the SARS-CoV-2 3CL protease are Michael acceptors,^[7] α-ketoamides,^[8] aldehydes,^[9–11] ketones^[12] and others.^[13] In these inhibitors warheads are flanked by different groups that try to mimic the interactions established by the fragments of the peptide substrate placed before (P_i) and after (P_{i'}) the cleaved peptide bond with the S_i and S_{i'} sites of the protease (see Figure 1 b).

Covalent inhibitors first bind into the active site of the protease forming a complex (EI) governed by noncovalent interactions. After binding, the noncovalent EI complex reacts with the thiol group of the catalytic cysteine to yield the E-I covalent complex.



Covalent inhibitors can be reversible or irreversible depending on the relative stability of the E-I complex.^[14,15]

One of the most promising families of inhibitors is constituted by aldehyde and ketone derivatives. At least three aldehydes presenting large inhibitory capacities have been already successfully tested in animals.^[9–11] These compounds, see Figure 2 a, present a γ-lactam ring at the P1 position, taking advantage of the selectivity of this enzyme for a glutamine residue before the bond to be cleaved. Another common characteristic is the presence of a hydrophobic group at P2 position, reproducing the preference of the enzyme for a leucine residue in the natural substrate. Combination of kinetic and structural studies demonstrated that aldehydes react with the enzyme forming a hemithioacetal complex, where the electrophilic carbon atom of the aldehyde group is bonded to the S_γ atom of the cysteine.^[9–11] We recently

[*] C. A. Ramos-Guzmán, Dr. J. J. Ruiz-Pernía, Prof. I. Tuñón
Departamento de Química Física
Universidad de Valencia
46100 Burjassot (Spain)
E-mail: j.javier.ruiz@uv.es
ignacio.tunon@uv.es

[**] A previous version of this manuscript has been deposited on a preprint server (<https://doi.org/10.33774/chemrxiv-2021-tndbs>).

Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under:
<https://doi.org/10.1002/anie.202110027>.

© 2021 The Authors. Angewandte Chemie International Edition published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution Non-Commercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.



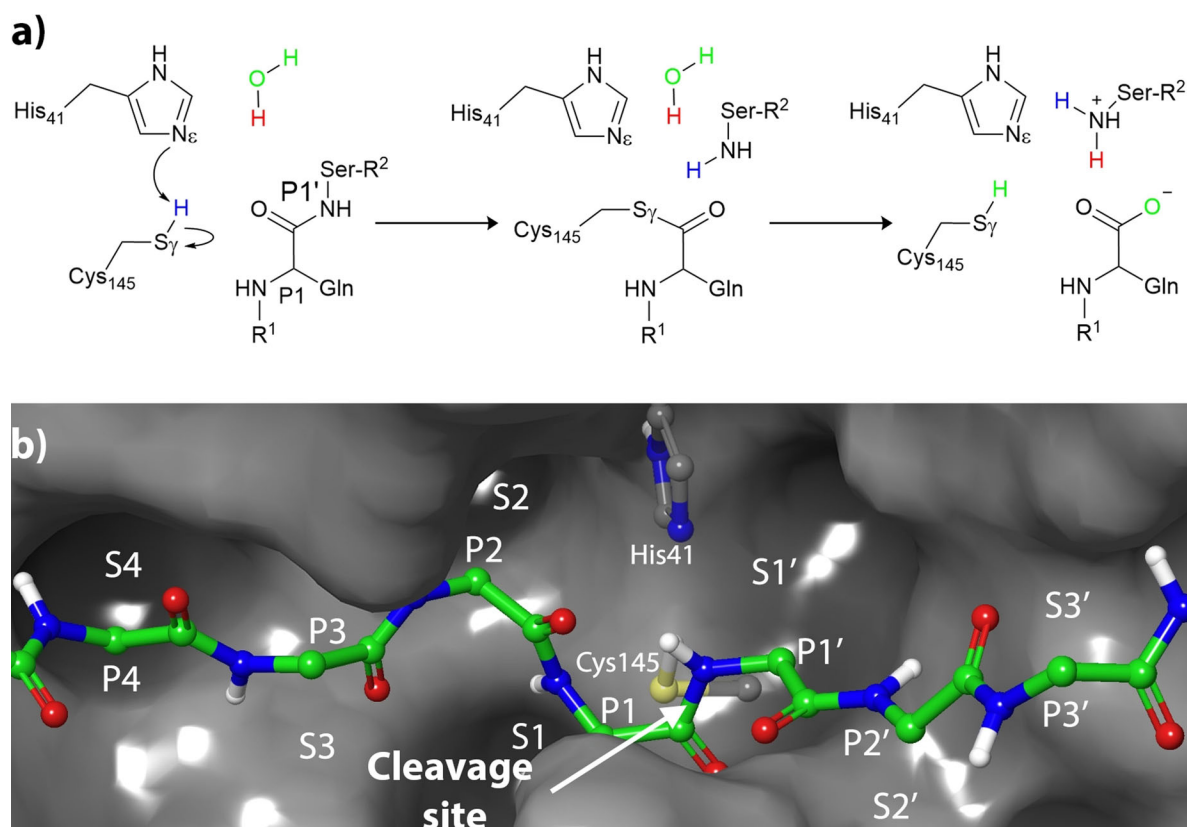


Figure 1. a) Proposed proteolysis mechanism of SARS-CoV-2 3CL protease.^[3] b) Schematic representation of the interactions in the active site of the proteases according to the Schechter and Berger nomenclature.

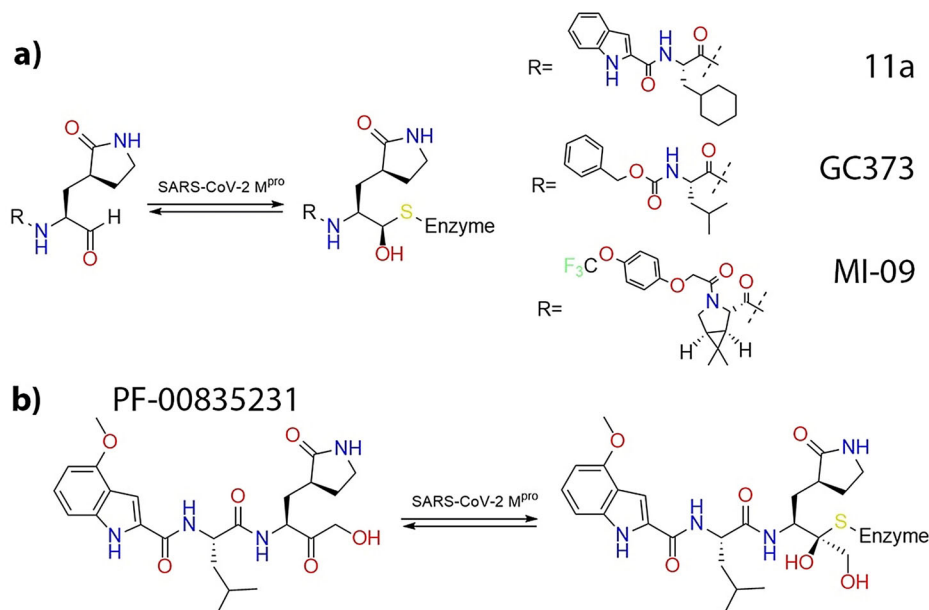


Figure 2. a) Aldehyde inhibitors of the SARS-CoV-2 3CL protease. b) PF-00835231 ketone-based inhibitor.

performed computational simulations that showed that a water molecule is recruited in the reaction mechanism to participate in the proton transfer from the catalytic histidine to the aldehyde oxygen atom.^[16] A very promising inhibitor,

already under clinical trial, is the ketone-based inhibitor developed by Pfizer and known as PF-00835231 (see Figure 2b) that inhibits the 3CL protease forming also a hemithioacetal complex.^[12] PF-00835231 is a hydroxymethyl ke-

tone obtained from the metabolization of the phosphate prodrug PF-07304814 that shows potent SARS-CoV-2 inhibition, good solubility and stability in antiviral assays, converting it to an excellent candidate for therapeutic treatment of COVID-19.^[12] The hydroxymethyl group of this ketone inhibitor could potentially mimic the interactions established by a serine residue at the P1' position of the peptide substrate of the protease, interactions that are obviously absent in aldehyde inhibitors. In addition, the hydroxymethyl group could also actively participate in the reaction mechanism, playing the role of the recruited water molecule in 3CL inhibition by aldehyde derivatives.^[16]

In this work we present the results of classical and hybrid QM/MM molecular dynamics (MD) simulations of the ketone-based inhibitor PF-00835231 in the active site of the SARS-CoV-2 3CL protease. We have first carried out classical MD simulations of the noncovalent complex (EI) identifying the most relevant hydrogen-bond interactions established between the inhibitor and the active site residues. Second, we used hybrid QM/MM methods to explore the reaction mechanism for the covalent inactivation of the enzyme. As reported previously,^[3,4,16] the reaction process is initiated with the activation of the catalytic dyad: a proton transfer from Cys145 to His41 that results in the catalytic dyad ion pair (IP). After this proton transfer the reaction proceeds with the nucleophilic attack of the activated cysteine on the carbonyl carbon atom and the proton transfer from the catalytic histidine to the carbonyl oxygen atom. In this last step the hydroxyl group of the inhibitor acts as a proton relay, accepting the proton from His41 and giving a proton to the carbonyl oxygen atom. The structure obtained for the reaction product agrees with the X-ray structure of the 3CL protease inhibited by PF-00835231 (6XHM).^[12] From the knowledge of the interactions established in the noncovalent complex and the mechanistic details of the covalent inhibition reaction we propose a chemical modification of the PF-00835231 inhibitor that could present improved kinetic properties. Our design is computationally tested by means of alchemical free energy calculations and QM/MM analysis of the reaction process. The simulations presented in this work could be used to improve, by rational design, future generations of antivirals.

Results and Discussion

Binding of the PF-00835231 Inhibitor

The starting point for our simulations is the X-ray structure with PDB code 6XHM. This structure corresponds to the dimeric enzyme with the two active sites inhibited by PF-00835231. In both active sites (corresponding to chains A and B) the bond between the S_γ atom of Cys145 and the carbonyl carbon atom is formed, with distances of 1.86 and 1.80 Å in chains A and B, respectively. In both cases the pose of the inhibitor is very similar (see Figure S1a) and the only significant difference appears in the rotameric state of the catalytic histidine (His41). In both active sites the N_ε atom of His41 is close to the S_γ atom of Cys145, displaying the same distance, 3.71 Å. However, in chain B the N_ε atom is

significantly closer to the inhibitor than in chain A, suggesting that this conformation would be more adequate for a subsequent proton transfer to the inhibitor. The distance from this atom to the hydroxyl oxygen atom of the inhibitor is only 2.65 Å in chain B, while in chain A the distance is increased to 3.80 Å. In the active site of this chain the N_δ atom of His41 is only slightly closer, 3.67 Å. These two rotameric states, hereafter denoted as ε-rotamer (chain B) and δ-rotamer (chain A), are connected by means of a 180° rotation around the C_β-C_γ bond of His41. We thus started our simulations of the noncovalent EI complex studying the preferred rotameric state of the catalytic His41 (see Figure S1b). To this aim (see SI), we traced the free energy profiles from the ε-rotamer to the δ-rotamer and backwards. The PMF displayed in Figure S1c correspond to the average of the two profiles. The ε-rotamer is the preferred conformation, being more stable than the δ-rotamer by 6.1 kcal mol⁻¹. This finding agrees with the simulations performed for the noncovalent complex formed between the 3CL protease and the 11a aldehyde inhibitor shown in Figure 2. Also in this case the ε-rotamer was found to be more stable than the δ-rotamer, by 3.2 kcal mol⁻¹.^[16] It must be also noticed that the larger stability observed for the ε-rotamer agrees with the observation that this conformer appears more frequently in the X-ray structures of the hemithioacetal complexes formed between the SARS-CoV-2 3CL protease and aldehyde or ketone inhibitors.^[16]

Once the preferred conformation for the catalytic histidine has been determined, we analyze the binding pose of the inhibitor and the interactions established with the enzyme in the noncovalent EI complex by means of MD simulations (5 replicas of 1 μs each). These MD simulations were stable in all cases (see RMSD time evolutions in Figure S2), showing a binding pose consistent with the X-ray structure of the hemithioacetal complex (see Figure 3a). P1, P2 and P3 sites of the inhibitor present an interaction pattern similar to that of a peptide substrate with sequence -Val-Leu-Gln|Ser- (where the vertical line indicates the scissile bond)^[3] and also similar to those of other peptidyl inhibitors, such as Michael acceptors^[17] and aldehyde derivatives.^[16] Figure 3b compares the fraction of hydrogen bonds established between the peptide substrate or the inhibitor with enzymatic residues as obtained from MD simulations. The P1 group establishes hydrogen bonds with His163, Glu166 and Phe140. The isobutyl hydrocarbon group at the P2 position stacks with the His41 imidazole ring, interacting also with other nearby residues, such as His164, Met165 and Gln189. Finally, the P3 group of the PF-00835231 inhibitor is exposed to the solvent and stabilized by hydrogen bond interactions with main chain atoms of Met165, Glu166 and Glu189. The position of the carbonyl oxygen atom of the inhibitor is stabilized by means of hydrogen bond interactions with the main chain NH groups of Cys145 (2.3 ± 0.3 Å), Ser144 (2.8 ± 0.4 Å) and Gly143 (2.5 ± 0.3 Å), as seen in Figure 3a. These interactions are also observed in the X-ray structure of the hemithioacetal product.

The formation of a covalent bond between the enzyme and the inhibitor requires the activation of the S_γ atom of Cys145 by means of a proton transfer from Cys145 to His41 and the subsequent nucleophilic attack of this atom on the



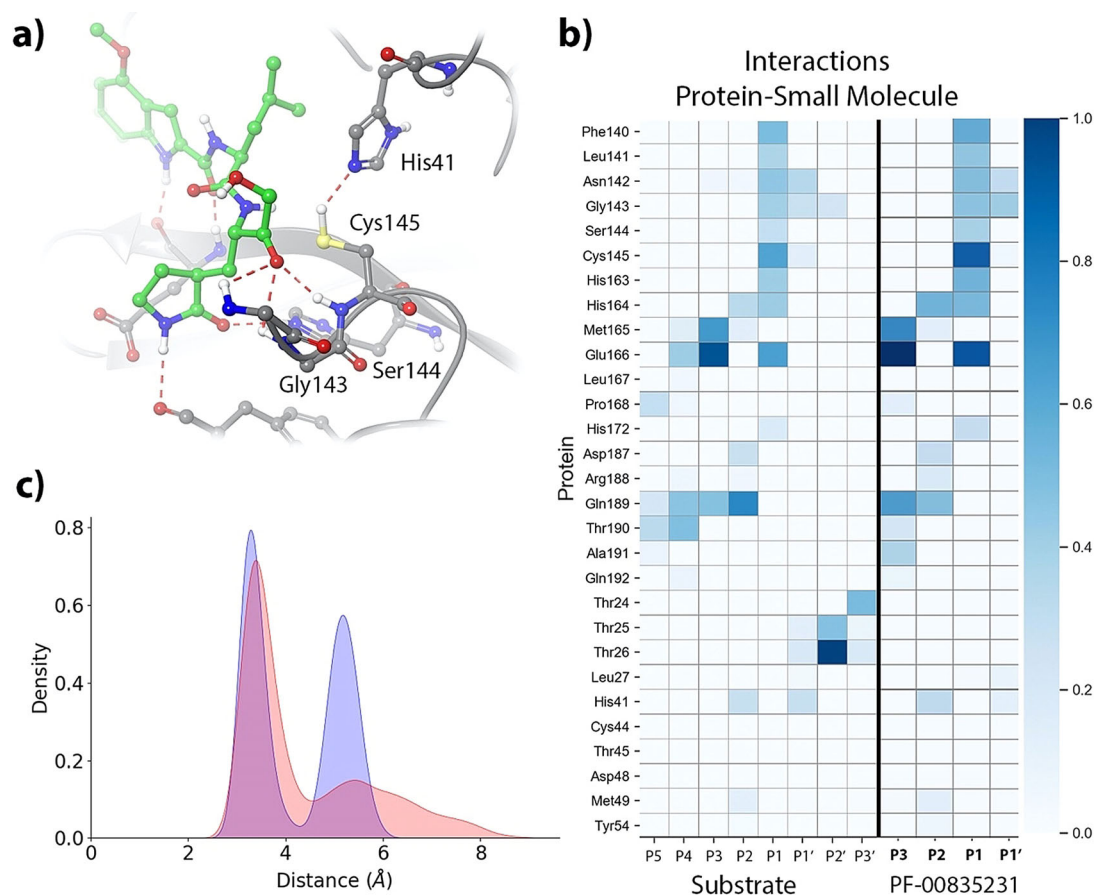


Figure 3. Noncovalent complex formed between PF-00835231 and the 3CL protease of SARS-CoV-2. a) Binding pose of the inhibitor in the active site of the protease, showing the location of the catalytic dyad and the oxyanion hole. Note that the carbonyl oxygen is accommodated in the oxyanion hole. b) Fraction of hydrogen bond contacts between residues of PF-00835231 and a peptide substrate^[3] and those of the protease. A hydrogen bond contact is counted when the donor–acceptor distance is < 3.8 Å and the hydrogen bond angle is $> 120^\circ$. c) Pair distribution functions between the Cys145-S γ atom and the carbonyl carbon atom of the substrate (in blue) and to the N ϵ atom of His41 (in red).

electron deficient carbonyl carbon atom of the inhibitor.^[3,17] We monitored the distances of the Cys145 S γ atom with the His41 N ϵ atom and with the carbonyl carbon atom of the inhibitor. The pair distribution functions are shown in Figure 3c. Both of them display bimodal distributions that can be attributed to the presence of *trans* and *gauche* conformers of the Cys145 side chain.^[18] The S γ -C and S γ -N ϵ distributions are peaked at 3.4/5.4 and 3.3/5.2 Å, respectively; showing that a significant fraction of the noncovalent EI complex conformations observed during the MD simulations are ready to proceed to the formation of the hemithioacetal product.

One of the main novelties of the PF-00835231 inhibitor is the inclusion of a hydroxymethyl group at the P1' position. This group resembles a serine residue, which is one of the preferences of 3CL proteases at this position. In fact, comparison of the hydrogen bond interactions established by Ser-P1' in the peptide substrate and the hydroxymethyl group in the PF-00835231 inhibitor shows that this last is able to recover a fraction of the interactions established by the peptide. The hydroxyl group of the inhibitor partially mimics the interactions made by the serine side chain, in particular with the catalytic dyad, His41 and Cys145. The main difference is caused by the larger rotational freedom of the P1' group in the inhibitor than in the peptide (compare

the P1' interactions for the peptide and the inhibitor in Figure 3b). In the case of the peptide substrate, the presence of P2' and subsequent groups reduces the conformational flexibility of the Ser-P1' side chain, favoring the formation of more stable interactions with the catalytic dyad. Instead, in the case of the inhibitor, the P1' group can rotate and then establish more frequent interactions with other residues of the active site, mainly with Gly143 and Asn142. We have evaluated the contribution of the hydroxymethyl group to the binding free energy of the inhibitor by means of the alchemical transformation from the corresponding aldehyde, where the hydroxymethyl group is substituted by a hydrogen atom (see v0 \rightarrow v1 transformation in Figure 4). The binding free energy difference between the PF-00835231 inhibitor and the aldehyde was computed by means of Thermodynamic Integration to be -1.00 ± 0.39 kcal mol⁻¹ (Table S1 provides the values corresponding to the five independent alchemical transformations in solution and in the protein environment, see SI for details). This means that the P1' group moderately contributes to increase the affinity of the inhibitor by the protein, a contribution that can be rationalized in terms of the protein-inhibitor interactions observed in the noncovalent complex.



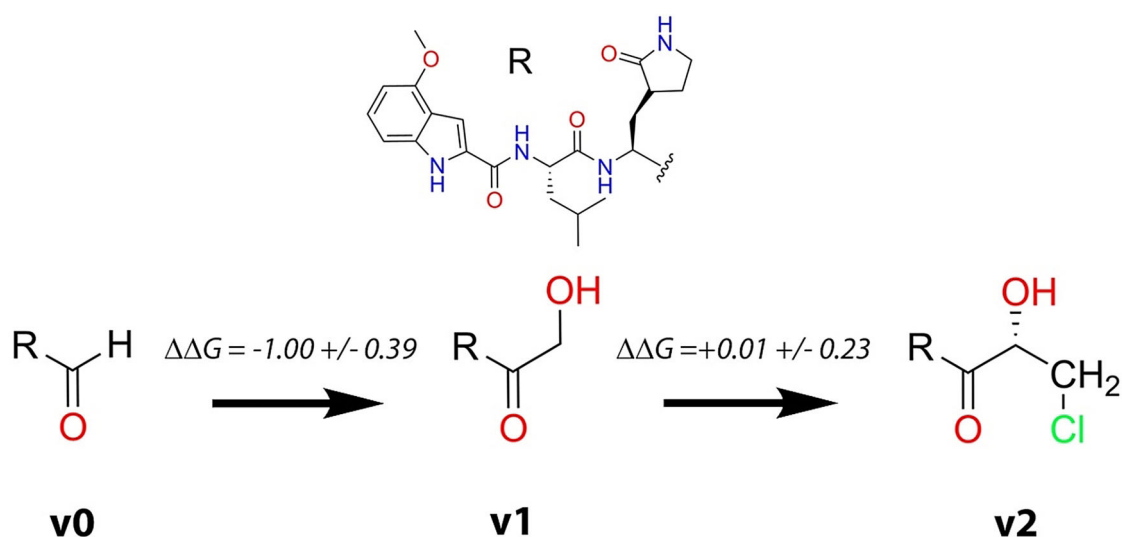


Figure 4. Alchemical transformations considered to evaluate the impact of chemical substitutions on the binding free energy. The original PF-00835231 inhibitor is denoted as v1, the aldehyde derivative as v0 and the inhibitor containing a chloromethyl substituent as v2. The free energy changes provided here correspond to the average of five independent simulations in solution and in the protein environment. Details of free energy simulations are given in Table S1.

Formation of the Covalent Hemithioacetal Product

In order to explore the reaction mechanism for the formation of the covalent E-I complex from the noncovalent one, we traced the corresponding MFEP using the string method at the B3LYPD3/6–31 + G*/MM level (see SI for details). According to previous studies on the 3CL protease activity with peptide substrates^[3,4] and other inhibitors,^[16,17,19,20] formation of a covalent bond with the S γ atom of Cys145 requires its activation by means of a proton transfer to His41. The reaction must be completed with a proton transfer from His41 to the inhibitor. In our previous study with aldehyde derivatives, we found that this proton transfer takes place mediated by a water molecule that occupies the same position as the hydroxyl group in the PF-00835231 inhibitor, suggesting the participation of this group as a proton relay during the proton transfer from His41 to the carbonyl oxygen atom of the inhibitor.^[16]

Our string simulations converged to a two-steps reaction mechanism, as shown in Figures 5 a and b (and video S1). In the first step a metastable ion pair (IP) catalytic dyad is formed after the proton transfer from Cys145 to His41. The IP is found to be 10.7 kcal mol⁻¹ above the noncovalent EI complex (see Figure 5 a). We confirmed this value by means of umbrella sampling simulations along the antisymmetric proton transfer coordinate ($d(\text{S}\gamma\text{H})-d(\text{N}\epsilon\text{H})$), which gave a free energy difference of 10.5 kcal mol⁻¹ (see Figure S3). This free energy difference between the neutral and IP forms of the dyad is very close to the values reported for other inhibitors. In particular, formation of the IP in presence of the 11a aldehyde^[16] requires 9.3 kcal mol⁻¹ and 10.3 in presence of a Michael acceptor.^[17] Instead, formation of the IP has a significantly smaller free energy cost (4.8 kcal mol⁻¹) when a peptide with a serine residues at the P1' position is present in the active site^[3] because in that case the hydroxyl group is better positioned to make favorable contacts with the

catalytic dyad. The IP, which is represented in Figure 5 c, is found in a free energy plateau, with a very low free energy barrier for its reversion back to the neutral dyad. At the IP, the proton transfer from Cys145 to His45 has been completed (the N ϵ -H and S γ -H distances are 1.04 and 2.61 Å, respectively) and His41 is now oriented towards the hydroxyl group of the inhibitor, forming a hydrogen bond with a N ϵ H-O distance of 1.93 Å. This first proton transfer from Cys145 to His41 is accompanied by the approach of the activated S γ atom to the carbonyl carbon: from 3.09 to 2.25 Å, when going from the noncovalent EI complex to the IP. From this state, the reaction proceeds completing the nucleophilic attack of the S γ atom of Cys145 to the carbonyl carbon atom of the inhibitor and the proton transfer from His41 to the hydroxyl group of the inhibitor and from this to the carbonyl oxygen atom of the ketone group (see evolution of the CVs in Figure 5 b). The rate-limiting transition state (TS) for the formation of the hemithioacetal is presented in Figure 5 d. At the TS the enzyme-inhibitor covalent bond (S γ -C) is almost completely formed, displaying a distance of 1.99 Å and the carbonyl double bond of the inhibitor has been elongated from 1.22 (EI complex) to 1.35 Å. The two proton transfers events are not very advanced at the TS. The N ϵ -H distance has been slightly lengthened up to 1.12 Å, while the distance of this proton to the hydroxyl oxygen atom of the inhibitor is 1.47 Å. The proton transfer from the hydroxyl group to the carbonyl oxygen atom is found at a slightly more advanced stage, being the distances of the proton to the donor and acceptor atoms of 1.16 and 1.39 Å, respectively. From this TS the reaction is completed to yield the hemithioacetal product (see Figure 5 e) where the two proton transfers are now completed. In our simulations, the covalent S γ -C bond presents a distance of 1.88 Å in the product state, in agreement with the X-ray observations (1.80/1.86 Å). In general, the configuration obtained for the covalent E-I complex agrees with the X-ray 6XHM structure (see Figure 5 e), with

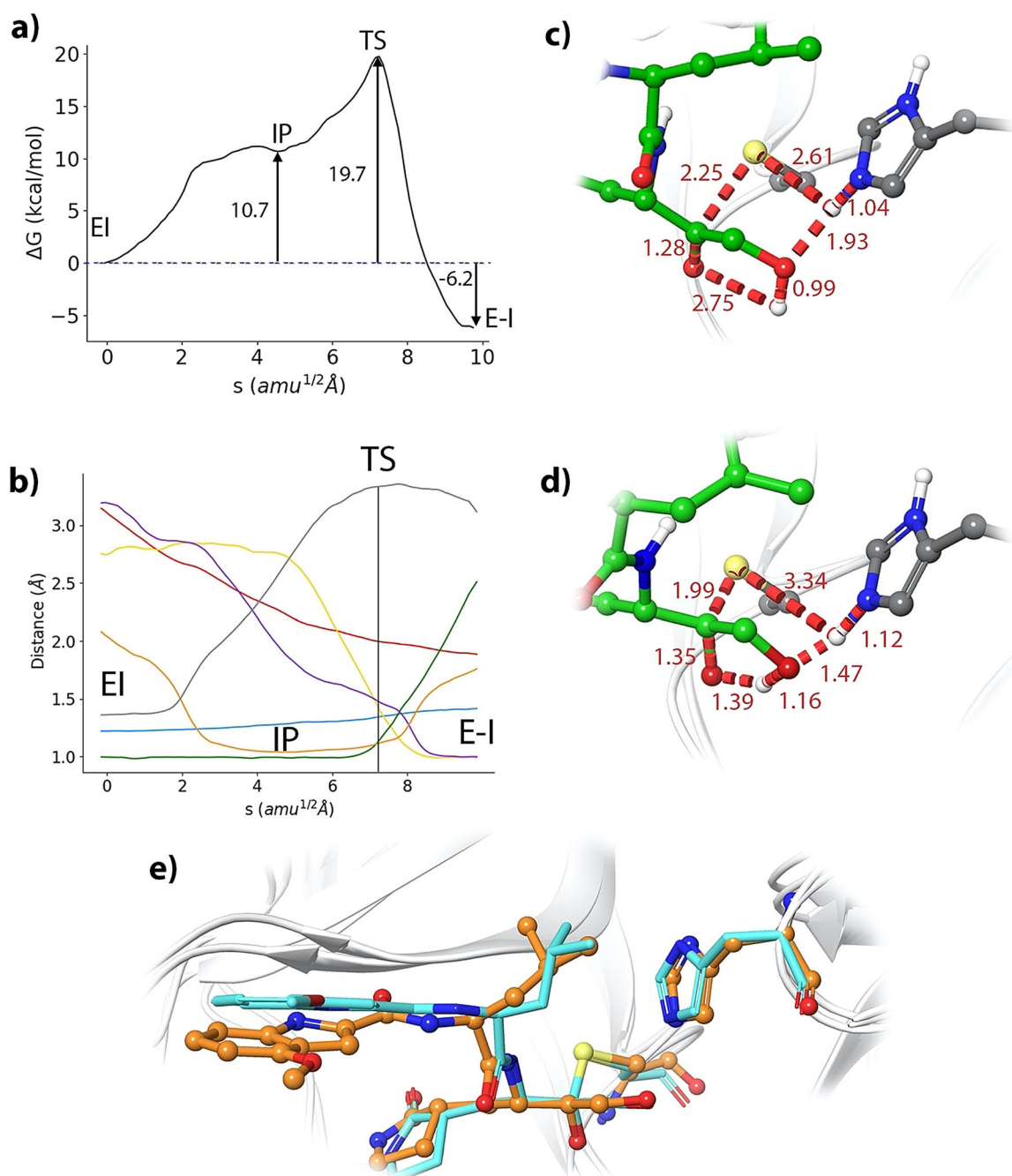


Figure 5. Formation of the (S)-hemithioacetal product. a) B3LYPD3/6–31 + G*/MM free energy profile along the path-CV for the formation of the covalent E-I complex from the EI complex. b) Evolution of the selected CVs along the MFEP: $S\gamma$ -C (red), $S\gamma$ -H (grey), $N\epsilon$ -H (orange), $H-O_{hid}$ (purple), $O_{hid}-H_{hid}$ (green), $H_{hid}-O$ (yellow), C-O (pale blue); see Scheme S1. c) Representation of the IP state. The values of the distances (in Å) correspond to the coordinates of the MFEP where the IP is located. d) Representation of the rate-limiting TS. e) Overlap of the product structure (balls and sticks with carbon atoms in orange) with the X-ray structure 6XHM containing a hydroxymethylketone inhibitor PF-00835231 (rods with carbon atoms in light blue).

an average RMSD calculated for the main chain non-hydrogen atoms, including those of the inhibitor, of 1.54 Å.

The free energy profile for the covalent inactivation of the 3CL protease with PF-00835231 (see Figure 5a) can be compared with experimental and theoretical results obtained for the inhibition with aldehyde derivatives 11a and GC-373. Recent analyses indicate that the PF-00835231 inhibitor has similar or higher potency against SARS-CoV-2 in human

A549 cells than GC-376 (the prodrug of GC-373).^[21] Our calculations predict that the formation of the hemithioacetal product from the noncovalent complex is exergonic, with a reaction free energy of $-6.2 \text{ kcal mol}^{-1}$. This value is larger, in absolute value, than the reaction free energy obtained for SARS-CoV-2 3CL inhibition with aldehyde 11a using the same computational Scheme, $-2.8 \text{ kcal mol}^{-1}$,^[16] but smaller than the value predicted for a Michael acceptor, -15.0 kcal

mol^{-1} .^[17] These two inhibitors, the 11a aldehyde and the N3 Michael acceptor, are examples of covalent reversible^[10] and irreversible^[7] inhibitors, respectively. Thus, our calculations predict that the ketone-based PF-00835231 inhibitor would display an intermediate behavior, with more irreversible character than other aldehyde inhibitors. The main difference found between the hemithioacetal products obtained with the aldehyde and the ketone-based inhibitors that can explain the larger stabilization of the latter is the formation of a hydrogen bond between the hydroxyl group of the inhibitor and the N ϵ atom of His41. The average O-N ϵ distance observed in our simulations of the hemithioacetal product of the PF-00835231 inhibitor was $2.77 \pm 0.11 \text{ \AA}$, close to the observed X-ray value (2.65 \AA , see Figure S1).

The predicted activation free energy for the process is $19.7 \text{ kcal mol}^{-1}$. Unfortunately, there are not experimental determinations of the first order inactivation rate constant for this inhibitor, but we can again compare our results with the values obtained for the inhibition of the SARS-CoV-2 3CL protease with aldehyde derivatives. The rate constant for the inhibition of the protease with the inhibitor GC373 has been measured to be $2.45 \times 10^{-3} \text{ s}^{-1}$ at 30°C ,^[22] a value that, according to Transition State Theory, can be translated into an activation free energy of $21.1 \text{ kcal mol}^{-1}$. For 11a our simulations predicted an activation free energy of $18.5 \text{ kcal mol}^{-1}$,^[16] the PF-00835231 inhibitor would present similar kinetic properties for the covalent inhibition of the SARS-CoV-2 3CL enzyme as the aforementioned aldehyde-based inhibitors.

Improving the Kinetic Properties of the Inhibitor

After the analysis of the characteristics of the binding and covalent inactivation processes of the PF-00835231 inhibitor, our next goal was to try to find a strategy to design an inhibitor with improved properties. This *in silico* design can also be considered as a further test of the conclusions reached with our simulations. As discussed in the previous section, our simulations indicate that a significant contribution to the inactivation free energy barrier comes from the formation of the IP and that this free energy cost significantly depends on the nature of the substrate/inhibitor and the interactions established in the active site. Our initial hypothesis was that the substitution of one of the hydrogen atoms of the P1' site of the inhibitor by a bulkier group could reduce its rotational freedom, favoring the interaction of the hydroxyl moiety with the catalytic dyad. This substitution could then lower the energetic cost due to IP formation, at the cost of reducing the binding entropy. A comparison between the binding pose of the PF-00835231 inhibitor at the non-covalent EI complex with that of a peptide substrate (see Figure S3) shows that such a substitution may be carried out on the pro-*R* hydrogen atom of the hydroxymethyl group. We selected a chloromethyl group (see variant v2 in Figure 4) as an appropriate candidate to be placed at the pro-*R* position for two reasons: i) it is a relatively small substituent and ii) this group polarizes the ketone group of the inhibitor, increasing the positive charge on the electrophilic carbon atom.

The binding pose of this new inhibitor (2 MD replicas of $1.5 \mu\text{s}$) is shown in Figure 6a. The interaction mode of the chloromethyl containing variant in the active site of the 3CL protease is very similar to that observed for the original inhibitor (compare Figures 3a and 6a). The new group is accommodated in the active site by means of hydrogen bond contacts with the side chain of Asn142 and the main chain NH group of Gly143 (see Figure 6b). As expected, the attachment of the methylchloride group reduces the rotation freedom of the P1' fragment of the inhibitor in the active site. This restriction is reflected in the distribution function of the distance between the oxygen atom of the hydroxyl group and the N ϵ atom of His41 (see Figure 6c) as compared to that of PF-00835231. However, these changes found in the interaction pattern are not reflected in an improved binding free energy. As explained in the methodological section, we used Thermodynamic Integration to evaluate the impact of the alchemical transformation of the original inhibitor into the chloromethyl variant (v1 \rightarrow v2 in Figure 4). According to our simulations, addition of the chloromethyl group has an almost neutral impact on the binding free energy, $0.01 \pm 0.23 \text{ kcal mol}^{-1}$ (see Table S1 for details). The entropic effect associated to the rotational restriction of the P1' fragment seems to compensate the enthalpic gain obtained from the new interactions established between the inhibitor and the protein.

We also evaluated the effect of the inclusion of the chloromethyl group on the formation of the covalent complex, tracing the corresponding free energy profile with the string method. These calculations were carried out using the same QM/MM description used for the PF-00835231 inhibitor, but including now the chloromethyl group into the QM region. The new profile, showed in Figure 6d, displays an activation free energy of $15.1 \text{ kcal mol}^{-1}$, significantly smaller than the value predicted for the original inhibitor ($19.7 \text{ kcal mol}^{-1}$, see Figure 5a). This reduction is essentially due to the lower free energy cost associated to the formation of the IP, $7.0 \text{ kcal mol}^{-1}$ in the presence of the new inhibitor versus $10.7 \text{ kcal mol}^{-1}$ obtained with the PF-00835231 inhibitor (we compare similar configurations of the IPs). The larger stabilization of the IP observed for the chloromethyl variant with respect to the original inhibitor is due to the enhanced interactions observed between the hydroxyl group of the inhibitor and the catalytic dyad as a consequence of the restricted mobility of the P1' group. The hydroxyl group has now the ability to stabilize the two charged fragments of the ionic pair catalytic dyad, acting simultaneously as proton donor and proton acceptor with Cys145 and His41, respectively. As expected, the reaction mechanism, described by the evolution of the collective variables and the geometry of the TS, remains essentially unchanged with respect to the original PF-00835231 inhibitor (see Figure S5 for details). The observed reduction in the activation free energy barrier in the new inhibitor can be translated into an increase of the covalent inactivation rate constant [k_3 in Eq. (1)] of 3 orders of magnitude at 300 K. This result shows that the kinetic properties of hydroxymethylketone inhibitors could probably be improved attaching new substituents to the hydroxymethyl group. These substituents play the role of P' fragments in

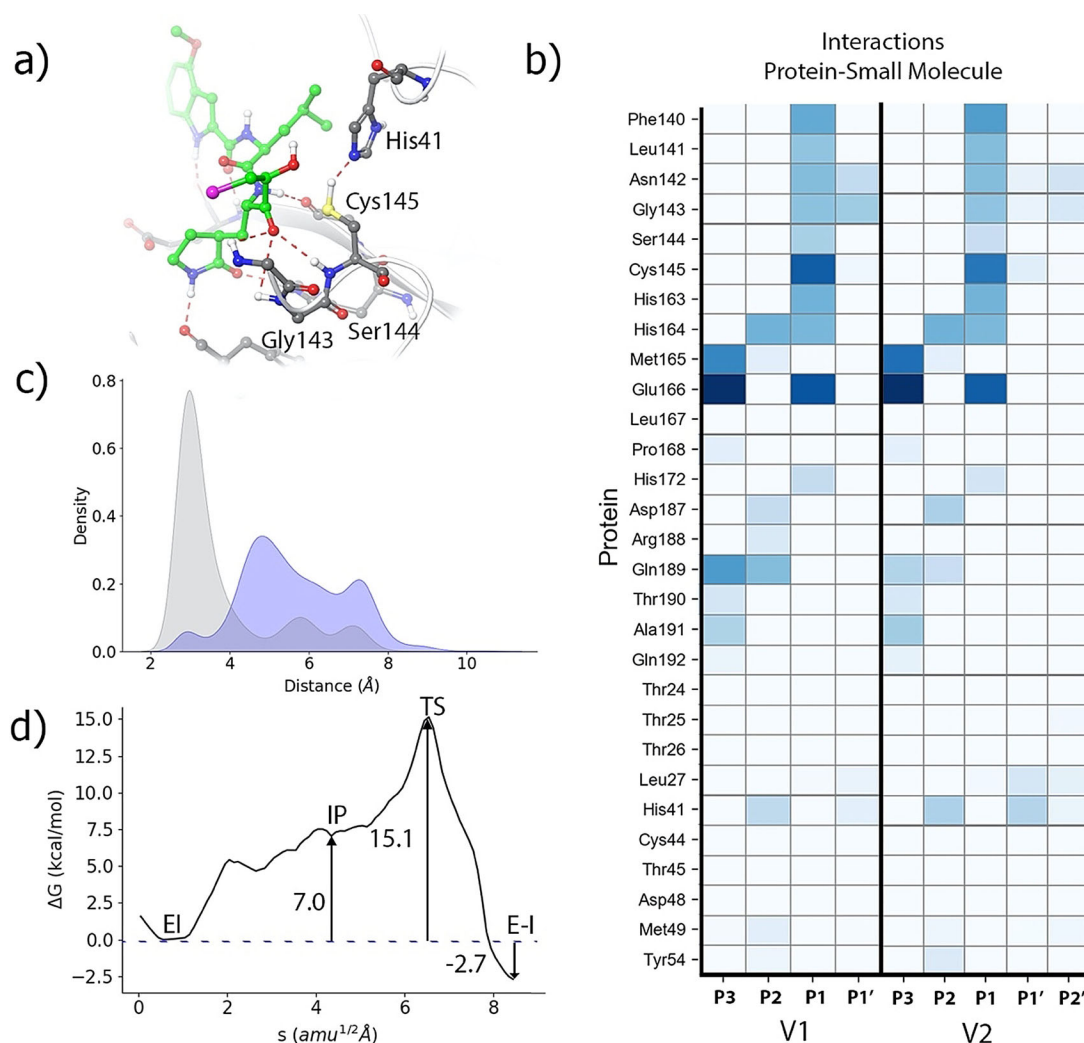


Figure 6. Improved design of a hydroxymethylketone inhibitor. a) Binding pose of the chloromethyl-containing inhibitor (v2 in Figure 4) in the active site of the SARS-CoV-2 3CL protease (chlorine atom shown in purple). b) Fraction of hydrogen bond contacts between residues of PF-00835231 inhibitor (v1) and the chloromethyl variant (v2) with the protease. c) Pair distribution functions from the oxygen atom of the inhibitor's hydroxyl group to the Ne atom of His41; the PF-00835231 inhibitor in blue and the chloromethyl variant in grey. d) B3LYPD3/6-31 + G*/MM free energy profile along the path-CV for the formation of the covalent E-I complex from the EI complex for the chloromethyl inhibitor.

peptidic substrates. Regarding the reaction free energy, that determines the irreversibility of the process, the process is now exergonic by only $-2.7 \text{ kcal mol}^{-1}$, smaller in absolute value than for the original inhibitor ($-6.2 \text{ kcal mol}^{-1}$).

Conclusion

We have used a combination of classical and hybrid QM/MM molecular dynamics simulations to explore the covalent inhibition mechanism of the SARS-CoV-2 3CL protease by a ketone-based inhibitor that is currently under clinical trial, PF-00835231. We first explored the binding mode and interactions established by the inhibitor starting from the X-ray structure of the hemithioacetal complex. We determined the preferred rotameric state of the catalytic His41, since the X-ray structure shows a different conformation for this residue in each of the active sites of the dimer. The

conformation determined as the most stable is that presenting the Ne atom of His41 close to the inhibitor warhead.

Our simulations emphasize the role played by the hydroxymethyl group at the P1' position of this inhibitor. Alchemical transformations from the corresponding aldehyde derivative show that the hydroxymethyl group contributes to increase the binding free energy of the inhibitor. Furthermore, this group plays also an active role during the formation of the hemithioacetal complex, mediating the proton transfer from the catalytic histidine to the carbonyl oxygen atom. The formation of the covalent complex is initiated after a proton transfer from Cys145 to His41 to form an ion pair. Once the catalytic dyad is activated, the process continues with the nucleophilic attack of the cysteine sulfur atom on the carbonyl carbon atom of the inhibitor and the proton transfer from His41 to the carbonyl oxygen atom. The rate limiting TS shows a short carbon-sulfur bond distance, while the proton transfers from the catalytic histidine to the hydroxyl group of



the inhibitor and from this group to the carbonyl oxygen atom are found at an earlier stage. The activation free energy associated to this TS is similar to that found for aldehyde inhibitors, while formation of the hemithioacetal complex is more exergonic for the ketone-based inhibitor.

Finally, we have explored the possibility to improve the kinetic properties of the inhibitor adding a substituent on the pro-*R* position of the P1' hydroxymethyl group. We tested the effect of the addition of a chloromethyl group which was introduced to favor the positioning of the P1' hydroxyl group in the vicinity of the catalytic dyad and to increase the positive charge on the carbonyl carbon atom. Our simulations show that the modification had a neutral effect on the binding free energy but increased the inactivation rate constant, due to the stabilization of the ion pair formed by the catalytic dyad. This new inhibitor should not be considered as end point in the design process, but rather as an indication of a possible route to further improve the design of hydroxymethylketone inhibitors. Our simulations indicate that substitutions at the pro-*R* hydrogen atom of the hydroxymethyl group of the PF-PF-00835231 inhibitor offer the possibility to improve the kinetic properties of this inhibitor of the SARS-CoV-2 3CL protease.

Acknowledgements

The authors acknowledge financial support from Consellería de Innovación, Universidades, Ciencia y Sociedad Digital, Generalitat Valenciana (GVCOV19/Decreto180/2020) and from Feder funds and the Ministerio de Ciencia, Innovación y Universidades (project PGC2018-094852-B-C22). We want to acknowledge Barcelona Supercomputing Center (BSC) for providing us access to MareNostrum and the staff from BSC for the technical support (Project QSB-2021-1-0006). We also acknowledge the use of the Tirant supercomputer at the Universitat de València (financed by the FEDER funds for Scientific Infrastructures; IDIFEDER-2018-063) and the support of Alejandro Soriano from Servei d'Informàtica from the Universitat de València.

Conflict of Interest

The authors declare no conflict of interest.

Keywords: 3CL protease · inhibitors · molecular modeling · PF-00835231 · SARS-CoV-2

- [1] C. R. M. Bangham, *J. Gen. Virol.* **2003**, *84*, 3177–3189.
- [2] J. Solowiej, J. A. Thomson, K. Ryan, C. Luo, M. He, J. Lou, B. W. Murray, *Biochemistry* **2008**, *47*, 2617–2630.
- [3] C. A. Ramos-Guzmán, J. J. Ruiz-Pernía, I. Tuñón, *ACS Catal.* **2020**, *10*, 12544–12554.
- [4] K. Świderek, V. Moliner, *Chem. Sci.* **2020**, *11*, 10626–10630.
- [5] R. Hilgenfeld, *FEBS J.* **2014**, *281*, 4085–4096.

- [6] K. Steuten, H. Kim, J. C. Widen, B. M. Babin, O. Onguka, S. Lovell, O. Bolgi, B. Cerikan, C. J. Neufeldt, M. Cortese, R. K. Muir, J. M. Bennett, R. Geiss-Friedlander, C. Peters, R. Bartschlagler, M. Bogyo, *ACS Infect. Dis.* **2021**, *7*, 1457–1468.
- [7] 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, H. Yang, *Nature* **2020**, *582*, 289–293.
- [8] L. Zhang, D. Lin, X. Sun, U. Curth, C. Drosten, L. Sauerhering, S. Becker, K. Rox, R. Hilgenfeld, *Science* **2020**, *368*, 409–412.
- [9] 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, H. Liu, *Science* **2020**, *368*, 1331–1335.
- [10] 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, M. J. Lemieux, *Nat. Commun.* **2020**, *11*, 4282.
- [11] 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, S. Yang, *Science* **2021**, *371*, 1374–1378.
- [12] 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, B. Taggart, *J. Med. Chem.* **2020**, *63*, 12725–12747.
- [13] C. A. Menéndez, F. Byléhn, G. R. Perez-Lemus, W. Alvarado, J. J. de Pablo, *Sci. Adv.* **2020**, *6*, eabd0345.
- [14] J. C. Powers, J. L. Asgian, Ö. D. Ekici, K. E. James, *Chem. Rev.* **2002**, *102*, 4639–4750.
- [15] M. Drag, G. S. Salvesen, *Nat. Rev. Drug Discovery* **2010**, *9*, 690–701.
- [16] C. A. Ramos-Guzmán, J. J. Ruiz-Pernía, I. Tuñón, *ACS Catal.* **2021**, *11*, 4157–4168.
- [17] C. A. Ramos-Guzmán, J. J. Ruiz-Pernía, I. Tuñón, *Chem. Sci.* **2021**, *12*, 3489–3496.
- [18] H. Yang, M. Yang, Y. Ding, Y. Liu, Z. Lou, Z. Zhou, L. Sun, L. Mo, S. Ye, H. Pang, G. F. Gao, K. Anand, M. Bartlam, R. Hilgenfeld, Z. Rao, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 13190–13195.
- [19] D. Mondal, A. Warshel, *Biochemistry* **2020**, *59*, 4601–4608.
- [20] K. Arafet, N. Serrano-Aparicio, A. Lodola, A. J. Mulholland, F. V. González, K. Świderek, V. Moliner, *Chem. Sci.* **2021**, *12*, 1433–1444.
- [21] M. de Vries, A. S. Mohamed, R. A. Prescott, A. M. Valero-Jimenez, L. Desvignes, R. O'Connor, C. Steppan, J. C. Devlin, E. Ivanova, A. Herrera, A. Schinlever, P. Loose, K. Ruggles, S. B. Korolov, A. S. Anderson, J. Binder, M. Dittmann, *J. Virol.* **2021**, *95*, e01819-20.
- [22] C. Ma, M. D. Sacco, B. Hurst, J. A. Townsend, Y. Hu, T. Szeto, X. Zhang, B. Tarbet, M. T. Marty, Y. Chen, J. Wang, *Cell Res.* **2020**, *30*, 678–692.







Manuscript received: July 27, 2021

Accepted manuscript online: September 28, 2021

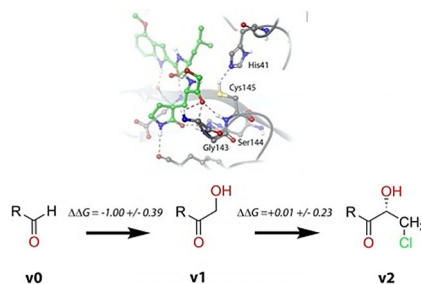
Version of record online: ■ ■ ■ ■ ■ ■ ■ ■ ■ ■

Research Articles

SARS-CoV-2 Inhibitors

C. A. Ramos-Guzmán, J. J. Ruiz-Pernía,*
I. Tuñón*      

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



Multiscale simulations unveil the binding and reaction mechanism of the SARS-CoV-2 main protease inhibitor, PF-00835231 inhibitor. This compound contains a hydroxymethyl group that plays a relevant role in the formation of the noncovalent and covalent complexes. In silico modifications show a possible strategy for the design of new inhibitors.

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, and Iñaki Tuñón**

anie_202110027_sm_miscellaneous_information.pdf

anie_202110027_sm_SI.pdb

anie_202110027_sm_Video_S1.mp4

Supporting Information

Methodology	S2
Figure S1. Rotameric State of His41	S6
Figure S2. RMSD of MD simulations	S7
Figure S3. PMF for Ion Pair formation	S7
Figure S4. Overlap of reactant's structures	S8
Figure S5. Covalent inhibition pathway for the inhibitor variant	S9
Table S1. Free Energy calculations for alchemical transformations	S10
Video S1. Reaction mechanism for covalent inhibition	S11
File S1. Coordinates of the noncovalent complex EI	S11
References	S12

Methodology

Classical Molecular Dynamics simulations

The PDB structure used to build the noncovalent EI complex was 6XHM (resolution 1.4 Å) that contains the dimeric form of the COVID-19 main protease covalently bonded to the PF-00835231 inhibitor in the two active sites.^[1] The H-bond assignment was optimized using the protein Preparation Wizard tool from Maestro^[2] and the most probable protonation states at pH 7.4 were determined using its PROPKA3.0.^[3] For neutral histidine residues, the δ/ϵ protonation state was confirmed after visual inspection of the x-ray structure. The parameters for the inhibitor were obtained using the non-standard residue parameterization procedure implemented in Amber with the Antechamber program^[4] from the AmberTools18^[5] package. The atomic charges for inhibitor atoms were obtained using the Restrained Electrostatic Potential (RESP) method^[6] at the HF/6-31G* level. Standard amino acids were describe using the ff14SB forcefield.^[7] This complex was immersed in a box of TIP3P water molecules, using the tleap tool from Ambertools18,^[5] in such a way that protein-inhibitor atoms were at least 12 Å from the limits of the simulation box. The charge of the enzyme-inhibitor system was neutralized by the addition of Na⁺ ions.

The noncovalent EI complex was prepared in the two active sites of the x-ray structure. To remove bad contacts the complex was first minimized using 500 steps of steepest descent algorithm. Then the conjugate gradient method was used to find a minimum energy structure until the root mean square of the gradient was below 10^{-3} kcal·mol⁻¹·Å⁻¹. Then, the system was heated slowly from 0 to 300 K. A linear heating ramp was used along 120 ps followed by a 20 ps simulation at 300 K. To avoid large variations in the atomic coordinates while heating, the position of the heavy atoms of the protein backbone were restrained using a harmonic restraint with a force constant of 20 kcal·mol⁻¹·Å⁻². Next, the system was equilibrated in the NPT ensemble (300K and 1 bar). The force constant of the positional restraints was reduced from 15 to 0 kcal·mol⁻¹·Å⁻², decreasing 3 units every 1.25 ns. Finally, the system ran free of restraints for 1.25 ns. In order to get enough sampling 5 replicas of 1 μs of the noncovalent enzyme inhibitor complex were simulated. Bond distances involving a hydrogen atom were constrained using SHAKE.^[8] Electrostatic interactions were treated using the Particle Mesh Ewald^[9,10] and a 10 Å cut-off radius was used for short-range interactions. In NPT simulations the Berendsen barostat and Langevin thermostat were used to control pressure and temperature respectively, while the time step was set equal to 2 fs. For all classical molecular dynamic simulations the AMBER19 GPU version of PMEMD^[11,12] was employed.

To study the free energy profile change to the rotation of the catalytic histidine side chain a classical potential of mean force (PMF) was performed using umbrella sampling.^[13] The dihedral angle formed by the $C\alpha-C\beta-C\gamma-N\delta$ atoms of the residue was used as the distinguished coordinate. In order to obtain the starting structures for the PMF a potential energy scan was done along the reaction coordinate. Starting from -120° , initial structures were sequentially minimized under a harmonic restraint using a force constant of $100 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{rad}^{-2}$. For each minimization 2000 steps of steepest descent method were followed by the conjugate gradient method until the root mean square of the gradient was below $10^{-3} \text{ kcal}\cdot\text{mol}^{-1}\text{\AA}^{-1}$. With this procedure we obtained a set of 49 structures separated by 5° from -120° to 120° and then, a total of 52 ns of classical MD was performed for each of them, the first 2 ns were run for relaxation followed by 50 ns of production. Then the process was repeated to obtain a PMF in the backward direction. The free energy profiles were integrated using the Weighted Histogram Analysis Method (WHAM).^[14]

Alchemical Transformations

Alchemical transformations of the inhibitor were performed to evaluate the changes in the binding free energy related to ligand modifications ($\Delta\Delta G_{\text{bind}}$). For this, the Amber Thermodynamic Integration protocol as proposed by He et al,^[15] was used. Briefly, for each transformation in aqueous solution and the protein environment 5 independent replicas were run with 9 values of the coupling parameter (λ) corresponding to the gaussian quadrature schedule. For each λ value, we first run a 200ps trajectory using pmemd.MPI (CPU) and subsequently 5 ns of simulation were made using pmemd.cuda (GPU). To compute the average value of $dU/d\lambda$ only the last 4 ns were employed, while the first nanosecond was considered as an equilibration run. In our case the $\Delta\Delta G_{\text{bind}}$ was calculated as the difference in the averages of the values for the transformation in aqueous solution and in the protein.

The structure of the PF-00835231 inhibitor was used as a starting point for the two alchemical transformations considered in this work (see Results). All the inhibitor atoms were included in the thermodynamic integration region, while only the atoms appearing/disappearing during the transformation were included in the soft-core region.

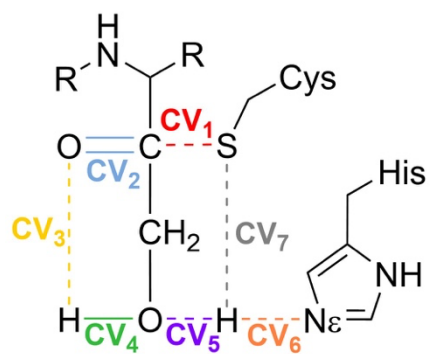
QM/MM calculations

The reaction under analysis is a multidimensional process that involves changes in several valence coordinates, preventing the use of free energy surfaces of low dimensionality. A better approach to study chemical reactions with complex reaction coordinates is Adaptive String Method (ASM).^[16] With this method the minimum free energy pathway (MFEP) can be traced on a multidimensional free energy surface of arbitrary dimensionality

and then an adequate reaction coordinate for the process can be defined as the advance along this path. It is worth to mention that in this method the number of collective variables describing the reaction coordinate can be as large as the system needs without implying an additional computational cost. Scheme S1 describe the set of 7 collective variables (CVs) used to describe the chemical transformation of the noncovalent complex to the hemithioacetal. In this case we used the distances of all the bonds whose formal order could be changed during the chemical transformation.

In our string calculations the structures of reactants and products were connected by 96 replicas of the system (or string nodes). Using QM/MM MD simulations, at every simulation step the nodes are moved according to their free energy gradient but keeping them redistributed equidistantly along the string. This procedure is continued until the string converges to the MFEP displaying a RMSD below $0.1 \text{ amu}^{1/2} \cdot \text{\AA}$ for at least 2 ps. To increase the convergence speed, replica exchange between nodes was used, with attempts done every 50 steps. After convergence, a path-CV (denoted as s) that measure the advance of the system along the MFEP is defined and used as reaction coordinate to trace the corresponding reaction free energy profile. Then 10 ps of QM/MM simulations were run for every node and the sampled values were integrated with WHAM^[14] to obtain the free energy profile along the path-CV. The values of the force constants employed to bias the ASM simulations were determined on-the-fly^[16] in order to ensure a homogeneous probability density distribution of the reaction coordinate. In the string simulations the mass of the protons in flight was set equal to 2 amu and the time step was of 1 fs.

The B3LYP functional^[17,18] with a 6-31+G* basis set and D3 dispersion corrections^[19] was selected to describe the QM region. This computational level has been shown to be a good methodological combination to obtain activation energies in agreement with the experimental data for the acylation of a peptide substrate^[20] and an inhibitor of the SARS-CoV-2 protease.^[21] QM/MM calculations were performed using a modified version of Amber18^[5,22] and Gaussian16^[23] for Density Functional Theory calculations. For all the QM-MM interactions the cutoff-radius used was 15 Å. The QM region included the side chains of the catalytic dyad (His41 and Cys145), the backbone atoms of residues P1 and the hydroxymethyl P1' fragment in the PF-00835231 inhibitor. Any other atom was described at the MM level and the link atom approach was used to describe the boundary between the two subsystems.



Scheme S1. Representation of the Collective Variables (CVs) selected for the string calculations on 3CL inhibition by PF-00835231.

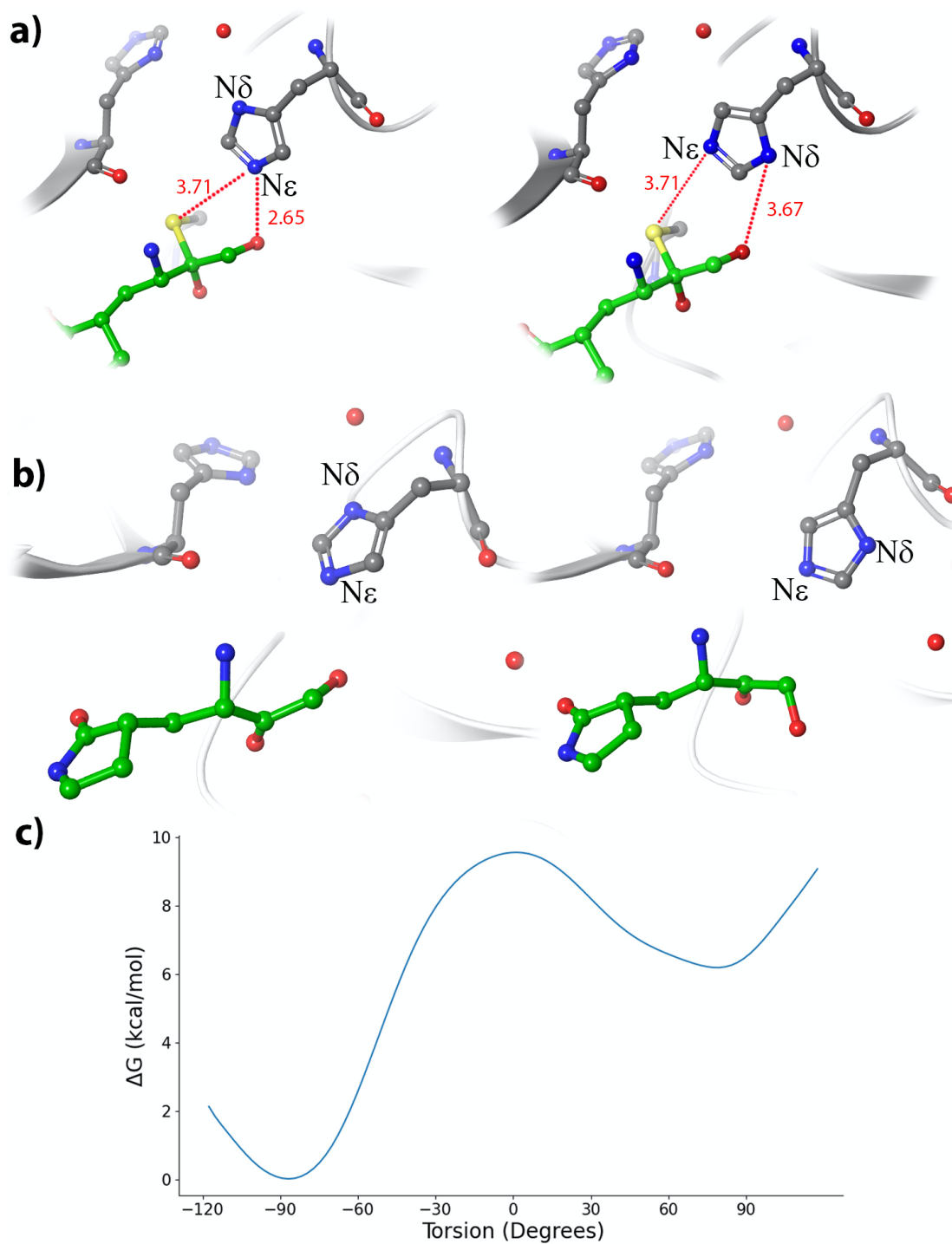


Figure S1. His41 rotameric state in the active site of 3CL protease. **(a)** Covalent product of PF-00835231 (carbon atoms in green) with the SARS-CoV-2 3CL protease in the ϵ -rotameric state found in chain B of 6XHM (left) and in the δ -rotameric found in chain A of 6XHM (right). **(b)** Representation of the noncovalent EI complex in the ϵ -rotameric state (left) and in the δ -rotameric state (right). **(c)** Free energy profile for the rotation of His41 along the C γ -C β bond, the ϵ - and δ -rotameric states appear at dihedral angle values of approximately -90 and 90 degrees.

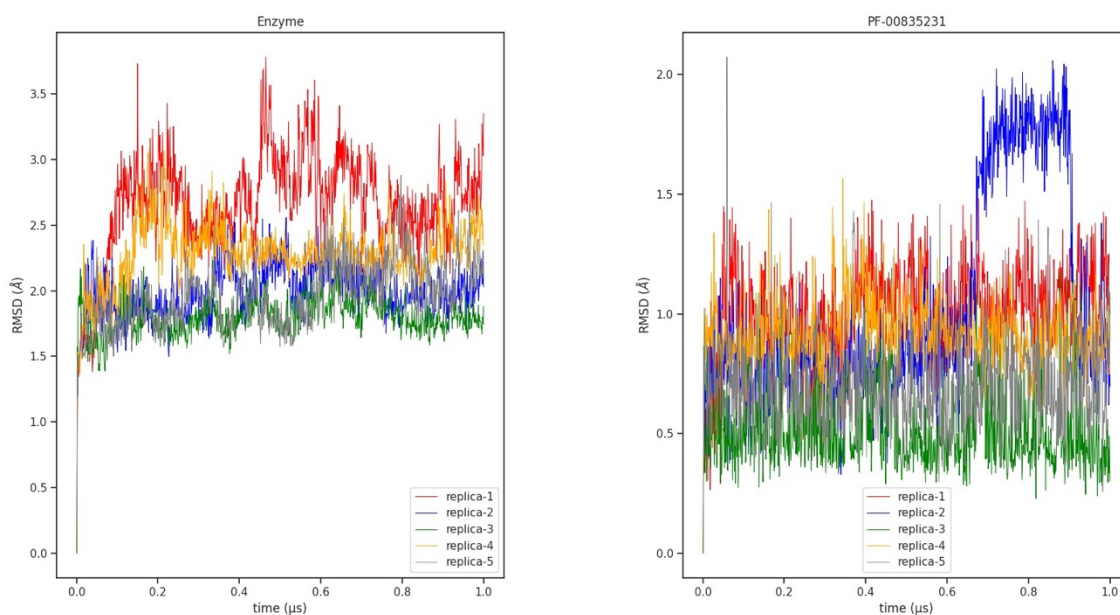


Figure S2. Plot of the RMSD values obtained for the protein (left) and the substrate (right) during the 5 replicas (simulation times of 1.0 μs) of the noncovalent enzyme-inhibitor complex of the SARS-CoV-2 protease with PF-00835231.

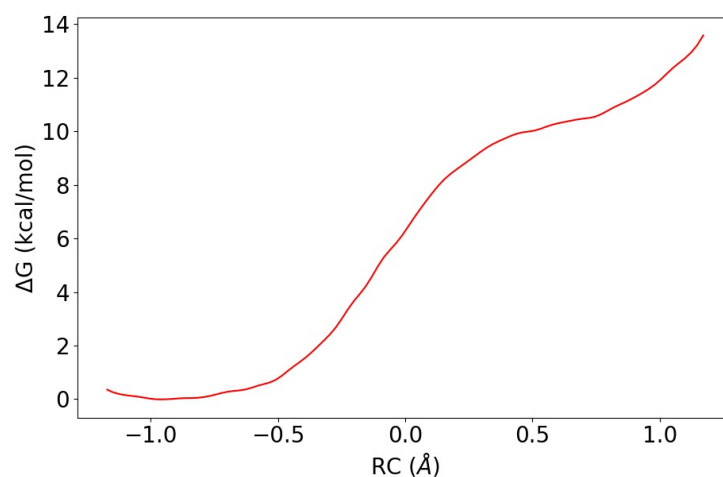


Figure S3. Analysis of the ion pair formation ($\text{Cys145}^- \cdots \text{His41H}^+$, right side) from the neutral form ($\text{Cys145H} \cdots \text{His41}$, left side) in the noncovalent complex. The free energy profile is obtained at the B3LYPD3/6-31+G*/MM level using Umbrella Sampling/WHAM simulations along the $d(\text{S}\gamma\text{-H})\text{-}d(\text{N}\epsilon\text{-H})$ coordinate.

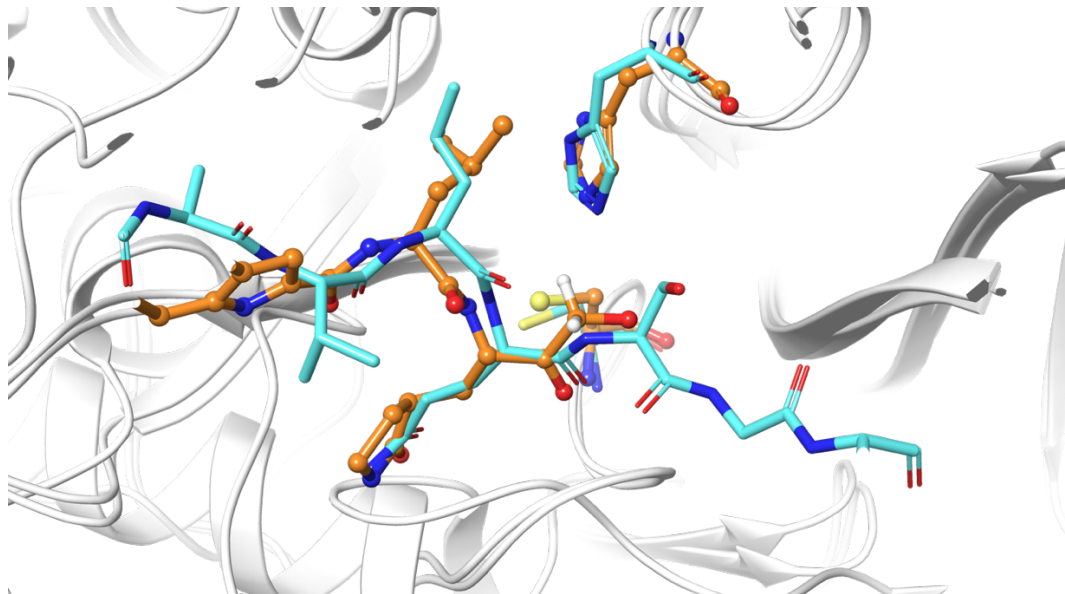


Figure S4. Comparison between the binding pose of a peptide substrate (licorice, carbon atoms in light blue) and the PF-00835231 inhibitor (carbon atoms in orange) in the active site of SARS-CoV-2 3CL protease. Only the hydrogen atoms of the hydroxymethyl group of the inhibitor are shown.

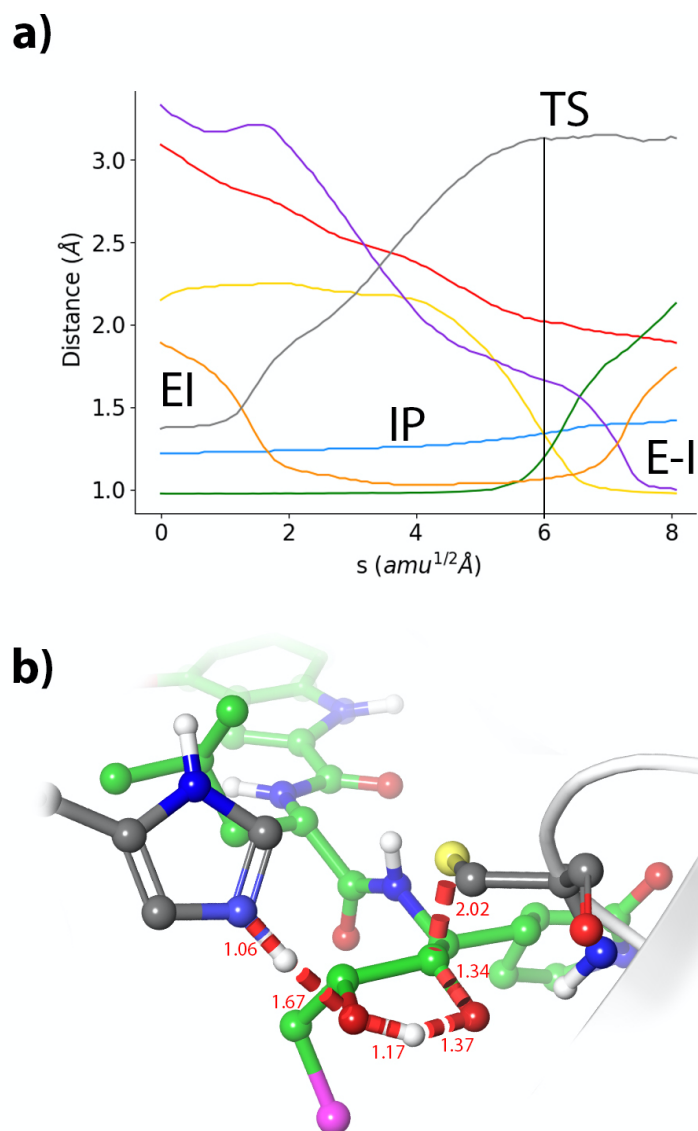
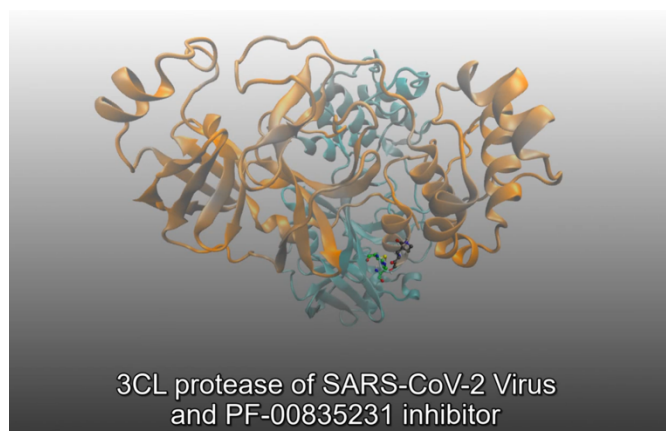


Figure S5. Formation of the hemithioacetal product corresponding to the methylchloride containing inhibitor. **(a)** Evolution of the selected CVs along the MFEP. The color code corresponds to Scheme 1. **(b)** Representation of the rate limiting TS (distances are given in Å)

Table S1. Free energy changes associated to alchemical transformations performed in aqueous and protein environments. 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		$\Delta\Delta G_{\text{bind}}$			
		-1.00 ± 0.39			
environment	replica	ΔG	environment	replica	ΔG
aqueous	1	40.57	protein	1	39.51
aqueous	2	40.67	protein	2	39.33
aqueous	3	40.57	protein	3	39.29
aqueous	4	40.57	protein	4	40.24
aqueous	5	40.52	protein	5	39.53
mean		40.58	mean		39.58
std		0.05	std		0.38

v1 to v2		$\Delta\Delta G_{\text{bind}}$			
		0.01 ± 0.23			
environment	replica	ΔG	environment	replica	ΔG
aqueous	1	-8.16	protein	1	-8.22
aqueous	2	-8.15	protein	2	-7.77
aqueous	3	-8.13	protein	3	-8.12
aqueous	4	-8.12	protein	4	-8.10
aqueous	5	-8.06	protein	5	-8.39
mean		-8.13	mean		-8.12
std		0.04	std		0.23



Video S1. Reaction mechanism for the covalent inhibition of the SARS-CoV-2 3CL protease by the PF-00835231 inhibitor. Trajectory for the TS can be downloaded from: http://disco.uv.es/pub/efme/disco/SARS-CoV-2/3CLPRO_PF835231_TS_traj and also from the MolSSI server: <https://covid.molssi.org//simulations/#10-ps-simulation-of-transition-state-of-the-sars-cov2-3clpro-in-complex-with-the-pf-00832531-inhibitor-all-atom-simulation>

File S1. EI_PF00835231.pdb file containing the coordinates of the simulated noncovalent complex formed between the SARS-CoV-2 3CL protease and the PF-00835231 inhibitor in PDB format can be downloaded also from the MolSSI server: <https://covid.molssi.org//simulations/#10-ps-simulation-of-transition-state-of-the-sars-cov2-3clpro-in-complex-with-the-pf-00832531-inhibitor-all-atom-simulation>

References

- [1] 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, B. Taggart, *J. Med. Chem.* **2020**, *63*, 12725–12747.
- [2] Schrödinger Release 2021-2: Maestro, Schrödinger, LLC, New York, NY, 2021.
- [3] M. H. M. Olsson, C. R. Søndergaard, M. Rostkowski, J. H. Jensen, *J. Chem. Theory Comput.* **2011**, *7*, 525–537.
- [4] J. Wang, W. Wang, P. A. Kollman, D. A. Case, *J. Mol. Graph. Model.* **2006**, *25*, 247–260.
- [5] 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, P. A. Kollman, *Univ. California, San Fr.* **n.d.**
- [6] C. I. Bayly, P. Cieplak, W. Cornell, P. A. Kollman, *J. Phys. Chem.* **1993**, *97*, 10269–10280.
- [7] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, C. Simmerling, *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.
- [8] J.-P. Ryckaert, G. Ciccotti, H. J. . Berendsen, *J. Comput. Phys.* **1977**, *23*, 327–341.
- [9] T. Darden, D. York, L. Pedersen, *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- [10] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, L. G. Pedersen, *J. Chem. Phys.* **1995**, *103*, 8577–8593.
- [11] S. Le Grand, A. W. Götz, R. C. Walker, *Comput. Phys. Commun.* **2013**, *184*, 374–380.
- [12] R. Salomon-Ferrer, A. W. Götz, D. Poole, S. Le Grand, R. C. Walker, *J. Chem. Theory Comput.* **2013**, *9*, 3878–3888.
- [13] G. M. Torrie, J. P. Valleau, *J. Comput. Phys.* **1977**, *23*, 187–199.
- [14] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, P. A. Kollman, *J. Comput. Chem.* **1992**, *13*, 1011–1021.
- [15] X. He, S. Liu, T. Lee, B. Ji, V. H. Man, D. M. York, J. Wang, *ACS Omega* **2020**, *5*, 4611–4619.
- [16] K. Zinovjev, I. Tuñón, *J. Phys. Chem. A* **2017**, *121*, 9764–9772.
- [17] A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648–5652.
- [18] C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* **1988**, *37*, 785–789.
- [19] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* **2010**, *132*, 154104.
- [20] C. A. Ramos-Guzmán, J. J. Ruiz-Pernía, I. Tuñón, *ACS Catal.* **2020**, *10*, 12544–12554.

- [21] C. A. Ramos-Guzmán, J. J. Ruiz-Pernía, I. Tuñón, *ACS Catal.* **2021**, *11*, 4157-4168.
- [22] K. Zinovjev, "String-Amber," can be found under <https://github.com/kzinovjev/string-amber>, **2019**.
- [23] **Gaussian 16, Revision C.01**, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, Gaussian, Inc., Wallingford CT, 2016.