



# New insights into the catalytic mechanism of the SARS-CoV-2 main protease: an ONIOM QM/MM approach

Henrique S. Fernandes<sup>1</sup> · Sérgio F. Sousa<sup>1</sup> · Nuno M. F. S. A. Cerqueira<sup>1</sup>

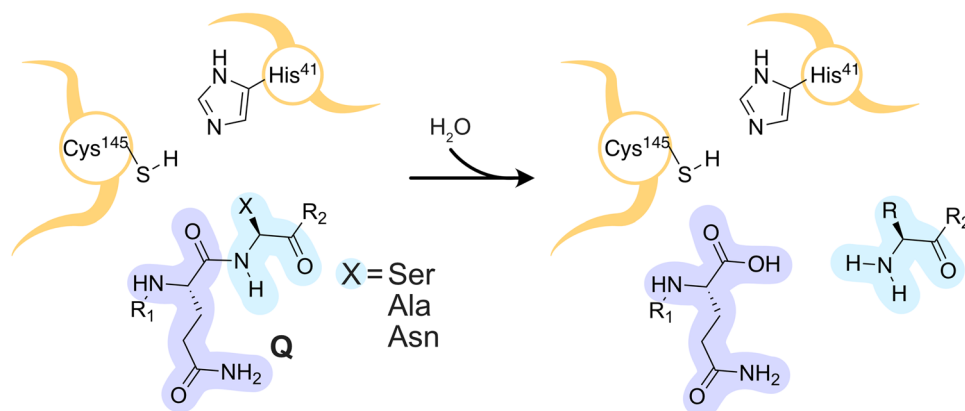
Received: 20 April 2021 / Accepted: 15 June 2021

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2021

## Abstract

SARS-CoV-2 M<sup>pro</sup>, also known as the main protease or 3C-like protease, is a key enzyme involved in the replication process of the virus that is causing the COVID-19 pandemic. It is also the most promising antiviral drug target targeting SARS-CoV-2 virus. In this work, the catalytic mechanism of M<sup>pro</sup> was studied using the full model of the enzyme and a computational QM/MM methodology with a 69/72-atoms QM region treated at DLPNO-CCSD(T)/CBS//B3LYP/6-31G(d,p):AMBER level and including the catalytic important oxyanion-hole residues. The transition state of each step was fully characterized and described together with the related reactants and products. The rate-limiting step of the catalytic process is the hydrolysis of the thioester-enzyme adduct, and the calculated barrier closely agrees with the available kinetic data. The calculated Gibbs free energy profile, together with the full atomistic detail of the structures involved in catalysis, can now serve as valuable models for the rational drug design of transition state analogs as new inhibitors targeting the SARS-CoV-2 virus.

## Graphic abstract



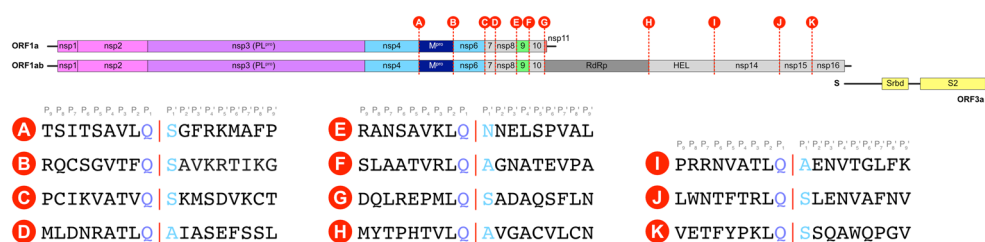
**Keywords** SARS-CoV-2 · Main protease · COVID-19 · Catalytic mechanism · QM/MM

## Introduction

The main protease of SARS-CoV-2, also known as SARS-CoV-2 M<sup>pro</sup>, is an essential enzyme for the replication of the virus causing the COVID-19 pandemic. M<sup>pro</sup> is the enzyme responsible for cleaving the translated polyprotein chain in 11 different regions (A-K) [1] (Fig. 1). This means that effective inhibiting M<sup>pro</sup> activity can stall the viral replication cycle [2]. Since no homology or sequence similarity is

✉ Nuno M. F. S. A. Cerqueira  
nunocerqueira@med.up.pt

<sup>1</sup> UCIBIO@REQUIMTE, BioSIM, Departamento de Biomedicina, Faculdade de Medicina da Universidade do Porto, Alameda Professor Hernâni Monteiro, 4200-319 Porto, Portugal



**Fig. 1** The complete genome sequence of SARS-CoV-2 virus [1]. Each red label (A-K) represents the respective protein cleavage regions catalyzed by  $M^{pro}$ . Each cleavage region is also shown with associated flanking amino acid sequences

found between the virus and human proteases [3],  $M^{pro}$  is seen as an appealing drug target for the development of new antiviral drugs to fight COVID-19, one among other strategies currently in development [4–12].

Similar therapeutic approaches were used with success, in the past, in the treatment of other viral infections, as, for example, the now commercially available inhibitors such as tipranavir, darunavir, amprenavir, lopinavir, among others [13], which target the HIV-1 protease. For this reason and along with the development of vaccines, the search for antiviral compounds continues to be on the top list of the pharmaceutical companies and research laboratories that seek to discover a compound capable of treating and impairing the dissemination of COVID-19 worldwide [14–16].

In record time, a valuable amount of knowledge was obtained regarding SARS-CoV-2 virus and, in particular, about  $M^{pro}$ .  $M^{pro}$  was described to be located between nsp4 and nsp5 segments of the polypeptide, which means that  $M^{pro}$  is able to cleave itself (Fig. 1a and b). Several X-ray structures of  $M^{pro}$  have also been deposited on the Protein Databank (PDB), revealing its three-dimensional structure [4, 17–23].

$M^{pro}$  is a homodimer whose subunits have independent active sites that are exposed to the solvent. Although the active sites are apparently independent, the monomer was shown to be inactive [3].

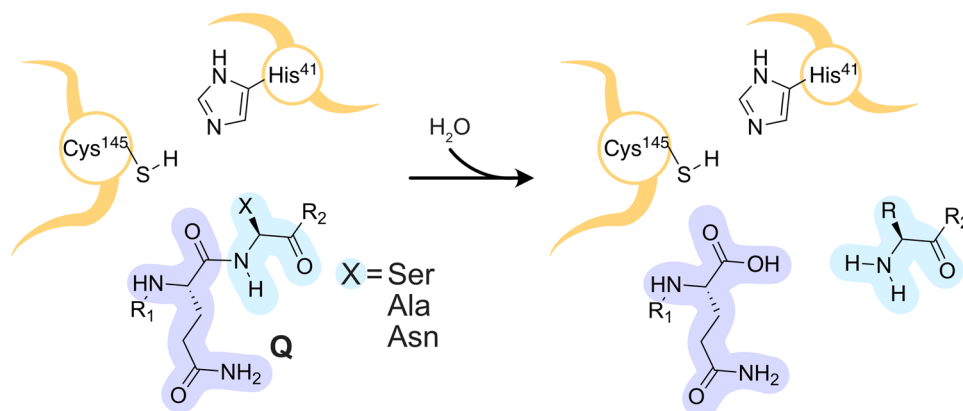
The solvent exposure of the active site favors the binding of the peptide substrates to the active site. Also, it ensures the availability of water molecules that are required for the completion of the catalytic mechanism (Fig. 2).

From the sequence analysis of all the peptides that are cleaved by  $M^{pro}$ , a well-established pattern is observed. The peptide bond that is cleaved by  $M^{pro}$  always occurs between a Gln residue (P1) and a Ser, Ala, or Asn residues (P1'). Moreover, the key Gln residue of the substrate peptide is always preceded by a bulky nonpolar residue: Val, Leu, or Phe (P2) (Fig. 1).

In addition, several residues from the active site (Thr24, Thr26, Asn142, Asn143, His163, His164, Glu166, Gln189, and Thr190) make hydrogen bonds with the peptide. These residues behave as anchors that allow the recognition of the peptide as a substrate and favor the correct alignment and orientation of the peptide inside the active site. Thus, a close contact between the target Gln residue and the catalytic Cys145 residue is ensured. According to the literature, the P4-P1' range comprises the most important residues for the recognition and anchoring of the peptide [24].

The reaction is catalyzed by a Cys-His dyad. The His41 residue activates the nucleophilic character of Cys145 that covalently attaches the peptide substrate to the enzyme. This reaction intermediate is then hydrolyzed, and the product of the reaction released to the active site (Fig. 2).

**Fig. 2** The general reaction catalyzed by  $M^{pro}$ . “R1” and “R2” represent the remaining part of the substrate (peptide). “X” represents the sidechain of the possible amino acid residues (Ser, Ala, or Asn)



So far, there are a panoply of studies devoted to the study of the M<sup>pro</sup> [19, 25–27], including the catalytic mechanism [28, 29]. In this work, the full catalytic mechanism of the SARS-CoV-2 M<sup>pro</sup> was studied using an ONIOM QM/MM methodology. The obtained results provide valuable insights about the structures of the transition state and linked minima for each step of the mechanism. Moreover, a complete atomistic description of the mechanism was obtained.

## Methodology

### Structure preparation

The QM/MM model used in this work was prepared based on the X-ray structure of M<sup>pro</sup> published by Z. Jin *et al.* and deposited on PDB under the 6LU7 code [17]. The structure of M<sup>pro</sup> was completely solved, and all the 306 amino acids of each subunit of the homodimer are available. A synthetic molecule was with the model as an analogue of the substrate. Since the main goal of this work was studying the mechanism of one possible natural substrate, the structure of the synthetic construct was used as template to model the following peptide sequence: ILMVFQSGACN. The co-crystallized synthetic construct shares some similarity with the substrate polypeptide. Accordingly, the “backbone” atoms were used as template to model the peptide sequence using Leap program from Amber package [30–33]. The scaffold of the synthetic construct was used to add the side chains of the residues belonging to one possible substrate of the enzyme.

The protonation state of the amino acid residues of the entire system was assessed using the PROPKA 3 software [34, 35] at pH 7.0 (Table S1). Finally, the resulting structure was solvated using a TIP3P type of water molecules [36, 37] (orthogonal box with 12.0 Å of minimum distance), and 8 sodium ions were added to neutralize the system. Then, the solvated system was minimized through a set of four sequential minimization protocols (non-bonded interactions cutoff: 10 Å), using AMBER software [30–33] and ff14SB force field [38]: 2500 (1250 steepest descent (ST) + 2500 conjugated gradients (CG)) steps for minimization of the water molecules; 2500 (1250 ST + 1250 CG) steps for minimization of all hydrogen atoms; 2500 (1250 ST + 1250 CG) steps for minimization of all system except backbone atoms; 10000 (5000 ST + 5000 CG) steps for minimization of all system. The resulting minimized structure was used to conduct a 60-ns molecular dynamics (MD) simulation: 50 + 50 ps NVT equilibration; 60-ns NPT production. The MD simulation was performed using a 2-fs integration step and the SHAKE algorithm [39] to constraint hydrogen atoms.

The minimized structure was used as a starting point to build and conduct the QM/MM calculations. This strategy

was employed to approximate the starting QM/MM model to the structure obtained experimentally.

### QM/MM model

The molUP plugin [40] for Visual Molecular Dynamics (VMD) [41] software, available through the VMD Store [42], was used to load the minimized structure obtained in the previous step and prepare the Gaussian09 [43] input files.

All the homodimer structure, the substrate peptide, and a 5 Å coat of water molecules were selected for the QM/MM model. The Gln and Ser residues of the substrate, the Cys145 residue, the sidechain of His41 residue, the backbone of Gly143 and Ser144 residues, and a water molecule (steps 3 and 4) were included in the high-level (HL) layer. In total, 69/72 atoms were included in the layer treated using quantum mechanics. The remaining atoms of the systems were considered using MM. A shell of 10 Å from the HL layer was kept free during the geometry optimizations. The relaxed coordinate scans and the geometry optimization of minima and transition states (TS) were performed using the following scheme: B3LYP [44–46]/6-31G(d,p) [47–52]:ff14SB [38]. Hydrogen atoms were used as link atoms to cap all the bond that spans across the QM-MM boundary. The calculations were performed considering the electrostatic embedding approach as implemented in Gaussian09 software. All the minima and TS were confirmed and validated by vibrational frequency analysis. Internal Reaction Coordinate (IRC) calculations were conducted to obtain the reactant and product for each step. The zero-point energy (ZPE), thermal, and entropic energy corrections were estimated at 298.15 K and 1.0 atm during the frequency calculation of TS and related minima structures.

The final energies were refined through single-point (SP) energy calculations with the following scheme: DLPNO-CCSD(T) [53]/CBS[54]:ff14SB. The HL layer was isolated from the QM/MM system using molUP extension for VMD. Hydrogen atoms were used as link atoms for the bonds spanning over the two layers. ORCA software (v4.2.1) [55] was used to perform the SP energy calculations using DLPNO-CCSD(T) and the basis sets cc-pVDZ and cc-pVTZ, from which the energies were extrapolated to complete basis set (CBS). The resulting electronic energies were summed to the LL layer contributions and to the thermal corrections previously computed with B3LYP/6-31G(d,p):ff14SB for the QM/MM system.

The activation Gibbs free energies ( $\Delta G^\ddagger$ ) of each step were calculated through the difference between the energy of the TS and respective reactant. On the other hand, the reaction Gibbs free energies ( $\Delta G_R$ ) were calculated through the difference between the energy of product and reactant for each step.

This computational approach has already been successfully employed to study the catalytic mechanism of several enzymes [56–62], including proteases [63–65].

## Results and discussion

According to the calculations performed in this study, the catalytic mechanism of SARS-CoV-2 M<sup>Pro</sup> occurs through four sequential steps. After an initial recognition process, where specific amino acid sequences of the substrate are recognized by the protease binding pocket, a key Gln residue from the substrate becomes aligned with the Cys145-His41 dyad, and the proteolytic reaction begins.

In the first step, the proton from Cys145 is transferred to His41, forming an ion pair. This reaction is characterized by a TS with an imaginary frequency at  $655.5i\text{ cm}^{-1}$  and an activation Gibbs free energy ( $\Delta G^\ddagger$ ) of 2.4 kcal/mol, and it is exergonic in 0.7 kcal/mol. In the product of the reaction, His41 becomes positively charged, Cys145 becomes anionic and gets close to the C1 carbon of the peptide substrate (SX---C1: 3.06 Å at reactant versus 2.30 Å at product) (Fig. 3 Step 1; Fig. 4). This step is very important to improve the nucleophilic nature of Cys145 and therefore to enhance its reactivity toward the peptide substrate. The nucleophilic nature of the SX of Cys145 is supported by its atomic charge variation from  $-0.10$  to  $-0.25$  a.u. during this first step. In addition, the approximation of SX to the C1 causes a decrease in the atomic charge of the electrophilic carbon of the substrate that passes from  $0.65$  to  $0.47$  a.u.. This shift of the electron density triggers the next step of the mechanism. This first step originates an ion pair where both His41 and Cys145 become charged, as reported in the interesting publication of Ramos-Guzmán et al. [29]. However, the activation and reaction energies reported in that publication are slightly higher than the ones obtained in this work. The difference could be assigned to the different peptide substrates used in both studies. Particularly, the P2 position is occupied by a Phe residue in our study versus a Leu residue, which can lead to a greater stabilization of His41 through a  $\pi$ -stacking interaction. Consequently, the positively charged His41 residue is more stabilized, and the reactions become faster and marginally exergonic. Contrary to our results and the Ramos-Guzmán et al. [29] work, the pioneer mechanism proposal by Swiderek et al. [28] suggests that the deprotonation of Cys145 is concomitant with the nucleophilic attack to the substrate with much higher activation and reaction energies. However, the study was conducted with a substrate analogue with a bulky non-amino acid molecular species at the P1' position, which in our study is occupied by Ser residue. The residue that occupies P1' position has been shown to play an important role in the stabilization of the ion pair either in SARS-CoV orthologs [66, 67] and SARS-CoV-2

protease [29]. Moreover, it seems that the hydroxyl group of Ser residue can stabilize the ion pair in about 1.4 kcal/mol [29] when compared with Gly or Ala residues.

The second step of the catalytic process comprises the nucleophilic attack of the anionic Cys145 to C1 carbon of the peptide substrate (Fig. 3 Step 2). At the same time, one proton migrates from the cationic His41 to the N1 atom of the substrate. This reaction requires an activation Gibbs free energy of 5.1 kcal/mol. However, this second step, where the peptide bond is broken, is considerably more exergonic than the previous one, with a  $\Delta G_R$  of  $-4.6$  kcal/mol. The TS of this step was characterized and confirmed by an imaginary frequency at  $1064.3i\text{ cm}^{-1}$ . In the product of this reaction, a thioester-enzyme adduct is obtained, which covalently attaches the substrate to the enzyme through the Cys145 residue. In this process occurs a change in the hybridization of C1 carbon of the substrate from  $sp^2$  to  $sp^3$ , which is corroborated by the improper torsion angle, C $\alpha$ -C1-N1-O1, that changes from  $170.5^\circ$  to  $131.0^\circ$ .

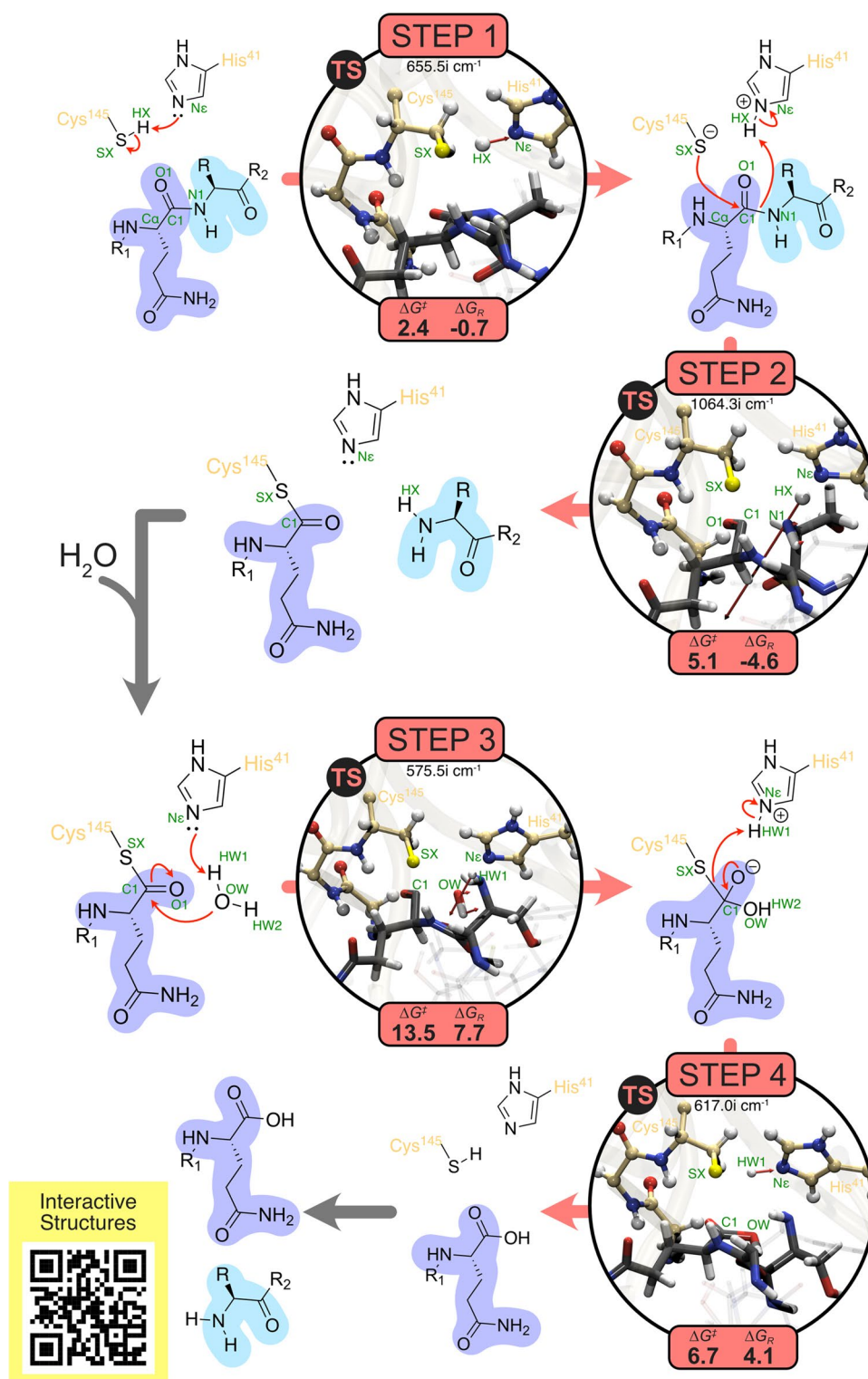
The activation Gibbs free energy obtained in this work is considerably lower than the one predicted by Ramos-Guzmán et al [29] (5.1 kcal/mol vs  $\sim 10$  kcal/mol). Nonetheless, the reaction product, after these two initial steps, shares similar energy in both studies ( $-5.3$  kcal/mol vs  $\sim -2$  kcal/mol). The considerably faster reactions reported by our calculations could be a consequence of the Phe residue at the position P1' as reported in step 1 and also the fact that, in our calculations, the backbone of critical Gly143 and Ser144 residues was included in the QM region since they are known to be important for stabilizing the oxo group of the substrate.

In the next steps of the catalytic process (Fig. 3 Steps 3 and 4), the hydrolysis of thioester-enzyme adduct takes place, requiring the presence of one water molecule, which is widely available in the active site due to its exposure to the solvent. In the optimized structure of the reactants, the water molecule becomes very close to the Cys145 residue, and it is stabilized by three hydrogen bonds provided by His41 (HW1: 1.79 Å), a carbonyl group of the substrate (HW2: 1.92 Å), as well as the amine group of the released peptide (OW: 2.14 Å).

After the appropriate accommodation of the water molecule within the active site, it makes a nucleophilic attack to the C1 carbon of the thiol-enzyme adduct (OW---C1: 3.15 Å at reactant versus 1.59 Å at product). Simultaneously, the His41 residue abstracts a proton from the water molecule (N $\epsilon$ ---HW1: 1.79 Å at reactant versus 1.05 Å) (Fig. 3 Step 3). At this point, the His41 residue becomes again positively charged, showing that it is also important for this second stage of the mechanism. The reaction is characterized by a TS with an imaginary frequency at  $575.5i\text{ cm}^{-1}$  and an activation Gibbs free energy of 13.5 kcal/mol. This reaction is also endergonic in 7.7 kcal/mol, which can be justified by

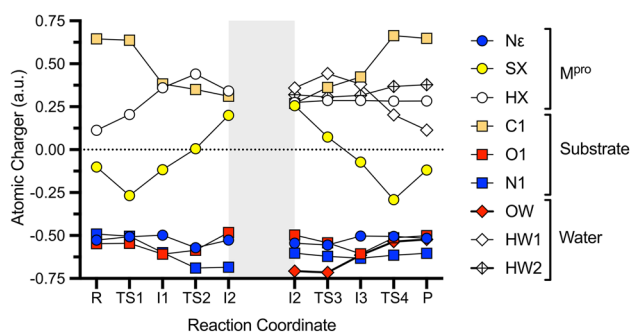


**Fig. 3** Representation of all the 4 steps of the catalytic mechanism of SARS-CoV-2 M<sup>pro</sup>. The activation and reaction Gibbs free energies are expressed in kcal/mol units. The eigenvectors of each TS structures are shown as red arrows



the formation of two punctual charges in the carbonyl group of the substrate (O1) and in the His41 residue. Although the product is less stable, this anionic tetrahedral intermediate (Fig. 4) is deeply stabilized by three hydrogen bonds provided by the NH groups of Gly143, Ser144, and Cys145 to

the negatively charged O1 atom. These three residues of the active site form what is commonly known as the oxyanion hole and have an important role in the catalytic process of M<sup>pro</sup>. This important interaction was not included in the QM part of the previous study of the mechanisms with the

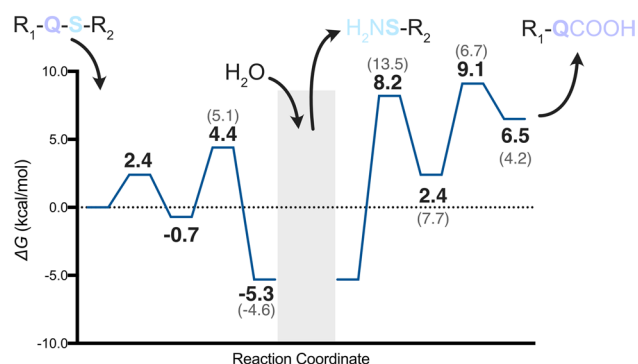


**Fig. 4** Charge fluctuation of the key atoms over the catalytic mechanism of the  $M^{pro}$ . Circles, squares, and lozenges identify atoms belonging to the catalytic dyad, the substrate, and the water molecule, respectively. Markers are colored by element: nitrogen (blue), sulfur (yellow), hydrogen (white), carbon (orange), and oxygen (red). A detailed analysis of the atomic charges is depicted in Figure S1

substrate, possibly leading to higher activation and reaction energies when compared with the results obtained in this work (13.5 kcal/mol vs 15.6 kcal/mol [29] and 7.7 kcal/mol vs ~12 kcal/mol [29]). When we repeated the calculations with a smaller QM region, excluding the backbone of Gly143, Ser144, and Cys145 residues, the estimated activation and reaction energies increase to 21.0 and 20.2 kcal/mol, respectively. This evidence confirms the importance of these residues in the QM region to properly stabilize the oxyanion formed in this step of the mechanism.

In the fourth and last step of the catalytic process, the proton HW1 is transferred from His41 to Cys145 (HW1---SX: 2.61 Å at reactant versus 1.38 Å at product), restoring the initial protonation state of His41 and Cys145. At the same time, the bond between the Cys145 residue and the C1 carbon is cleaved, which releases the second reaction product to the active site (SX---C1: 1.59 Å at reactant versus 3.23 Å at product) (Fig. 3 Step 4). The TS was also verified and presents an imaginary frequency at 617.0i  $\text{cm}^{-1}$ . This last step occurs when the activation Gibbs free energy barrier of 6.7 kcal/mol is overcome. This last step is endergonic in 4.1 kcal/mol. At the end of step 4, the initial protonation state of all the active site residues is restored, and the enzyme is ready to catalyze a new proteolytic reaction.

The Gibbs free energy profile of the proteolytic mechanism catalyzed by  $M^{pro}$  is shown in Fig. 5. In this profile, two well-defined stages can be identified. In the first part of the mechanism, the peptide substrate becomes covalently attached to the enzyme, resulting in the thioester-enzyme adduct and the release of the N-terminal part of the substrate. In the second stage occurs the hydrolysis of the thioester-enzyme adduct through two sequential steps and from which it results in the release of the product of the reaction, i.e., the C-terminal part of the peptide. This reaction is the rate-limiting step of the full catalytic process and requires



**Fig. 5** Complete energy profile for the catalytic mechanism of SARS-CoV-2  $M^{pro}$ . The values placed between parentheses correspond to the activation/reaction Gibbs free energies of each step

total activation Gibbs free energy of 14.4 kcal/mol (Fig. 5). The available kinetic data for Gln-Ser SARS-CoV protease show turnover numbers between 0.847 and 12.2  $\text{s}^{-1}$  [66, 67], corresponding to approximately 16.0 to 17.6 kcal/mol, considering the transition state theory. The maximum activation barrier calculated in this work agrees with the recent evidence that the catalytic efficiency of SARS-CoV-2  $M^{pro}$  is slightly higher than the SARS-CoV ortholog [17], which suggests a faster reaction and, consequently, lower activation energy for the rate-limiting step.

## Conclusion

In the past year, several studies have been published assessing different features of the SARS-CoV-2 virus and associated human disease. The  $M^{pro}$  was one of the most studied drug targets due to the success in the treatment of other viral infections using inhibitors of similar proteases. Valuable computational studies reported important insights about the catalytic mechanism of the  $M^{pro}$  [28, 29, 68, 69]. In this work, we provide new relevant features of the catalytic mechanism of this protease. The calculations reported in this publication corroborate the hypothesis of an ion pair formation prior to the nucleophilic attack of Cys145 to the substrate as proposed in another study [29] and contrary to what has been suggested by Świderek *et al.* [28]. The non-natural and bulky molecular group at the P1' position of the substrate seems to be responsible for the concomitant nucleophilic attack described in that work, as well as the higher activation and reaction energies. Moreover, our results elucidate the key role played by Gly143, Ser144, and Cys145 residues in the stabilization of the oxyanion hole in lowering the activation barrier of the rate-limiting step. Their inclusion in the QM region of any QM/MM model seems essential for properly modeling their important role.

The insights provided by the calculations reported in this work, together with the important conclusions already described in previous publications, can be useful to rationally develop TS analogue inhibitors targeting  $M^{\text{pro}}$ . Usually, inhibitors should have the following characteristics: (i) be recognized by the active site and allow the correct alignment of the Gln or a similar molecular scaffold in relation to the catalytic dyad and (ii) react with the active site residues and irreversibly inhibit the enzyme. To fulfill the first point, the inhibitor should follow a pattern of intermolecular interactions similar to what is observed with the original substrate peptides. Since the drug candidate will compete with the natural substrates of  $M^{\text{pro}}$  to bind into the active site, it is important to achieve a comparable or higher degree of affinity. In order to irreversibly inhibit the enzyme, the compound must react with the catalytic dyad and generate a similar thioester-enzyme adduct, as observed with the natural substrate. This reaction intermediate should be stable and prevent the hydrolysis or the reverse reaction. A good starting point is the three-dimensional structure of the TS from step 2 (available in supporting information), which can be modified to fulfill the requirements described above. We believe that this knowledge can be a step forward in the development of more specific and potent  $M^{\text{pro}}$  inhibitors.

## Associated content

- Atomic charge distribution of the different reactional species over the mechanism (PDF)
- Molecular structures for reactant, transition state, and product of all the steps of the mechanism (PDB files in a ZIP package)
- A web resource for visualization and inspection of the structures of reactant, TS, intermediates and product of the reaction is available at: covid19.biosim.pt

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s11030-021-10259-7>.

**Author contributions** The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

**Funding** This work was supported by the Applied Molecular Biosciences Unit—UCIBIO, which is financed by national funds from FCT (UIDP/04378/2020 and UIDB/04378/2020). HSF acknowledges FCT for his PhD grant SFRH/BD/115396/2016 and SS for grant 2020.01423.CEECIND. Some of the calculations were produced with the support of INCD funded by FCT and FEDER under the projects 01/SAICT/2016 number 022153, and projects CPCA/A00/7140/2020, and CPCA/A00/7145/2020.

## References

1. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG et al (2020) A new coronavirus associated with human respiratory disease in China. *Nature*. 579(7798):265–9. <https://doi.org/10.1038/s41586-020-2008-3>
2. Anand K, Ziebuhr J, Wadhwani P, Mesters JR, Hilgenfeld R (2003) Coronavirus main proteinase (3CL<sup>pro</sup>) structure: basis for design of anti-SARS drugs. *Science*. 300(5626):1763–7. <https://doi.org/10.1126/science.1085658>
3. Ullrich S, Nitsche C (2020) The SARS-CoV-2 main protease as drug target. *Bioorg Med Chem Lett*. 30(17):127377. <https://doi.org/10.1016/j.bmcl.2020.127377>
4. Zhang L, Lin D, Sun X, Curth U, Drosten C, Sauerhering L et al (2020) Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved alpha-ketoamide inhibitors. *Science*. 368(6489):409–12. <https://doi.org/10.1126/science.abb3405>
5. Das S, Sarmah S, Lyndem S, Singha Roy A (2020) An investigation into the identification of potential inhibitors of SARS-CoV-2 main protease using molecular docking study. *J Biomol Struct Dyn*. <https://doi.org/10.1080/07391102.2020.1763201>
6. Ngo ST, Quynh Anh Pham N, Le Thi L, Pham DH, Vu VV (2020) Computational determination of potential inhibitors of SARS-CoV-2 main protease. *J Chem Inf Model* 60(12):5771–80. <https://doi.org/10.1021/acs.jcim.0c00491>
7. Ton AT, Gentile F, Hsing M, Ban F, Cherkasov A (2020) Rapid identification of potential inhibitors of SARS-CoV-2 main protease by deep docking of 1.3 billion compounds. *Mol Inform* 39(8):2000028. <https://doi.org/10.1002/minf.202000028>
8. Eleftheriou P, Amanatidou D, Petrou A, Geronikaki A (2020) In silico evaluation of the effectivity of approved protease inhibitors against the main protease of the novel SARS-CoV-2 virus. *Molecules*. <https://doi.org/10.3390/molecules25112529>
9. Sacco MD, Ma C, Lagarias P, Gao A, Townsend JA, Meng X et al (2020) Structure and inhibition of the SARS-CoV-2 main protease reveal strategy for developing dual inhibitors against M(pro) and cathepsin L. *Sci Adv*. <https://doi.org/10.1126/sciadv.abe0751>
10. Lapaillerie D, Charlier C, Fernandes HS, Sousa SF, Lesbats P, Weigel P et al (2021) In silico, in vitro and in cellulo models for monitoring SARS-CoV-2 spike/human ACE2 complex viral entry and cell fusion. *Viruses*. <https://doi.org/10.3390/v13030365>
11. Kumar Y, Singh H, Patel CN (2020) In silico prediction of potential inhibitors for the main protease of SARS-CoV-2 using molecular docking and dynamics simulation based drug-repurposing. *J Infect Public Health* 13(9):1210–23. <https://doi.org/10.1016/j.jiph.2020.06.016>
12. Li Z, Li X, Huang YY, Wu Y, Liu R, Zhou L et al (2020) Identify potent SARS-CoV-2 main protease inhibitors via accelerated free energy perturbation-based virtual screening of existing drugs. *Proc Natl Acad Sci U S A* 117(44):27381–7. <https://doi.org/10.1073/pnas.2010470117>
13. Lv Z, Chu Y, Wang Y (2015) HIV protease inhibitors: a review of molecular selectivity and toxicity. *HIV AIDS (Auckl)* 7:95–104. <https://doi.org/10.2147/HIV.S79956>
14. Lurie N, Saville M, Hatchett R, Halton J (2020) Developing covid-19 vaccines at pandemic speed. *N Engl J Med* 382(21):1969–73. <https://doi.org/10.1056/NEJMp2005630>
15. Zhang C, Huang S, Zheng F, Dai Y (2020) Controversial treatments: An updated understanding of the coronavirus disease 2019. *J Med Virol* 92(9):1441–8. <https://doi.org/10.1002/jmv.25788>
16. Zhang J, Zeng H, Gu J, Li H, Zheng L, Zou Q (2020) Progress and prospects on vaccine development against SARS-CoV-2. *Vaccines (Basel)*. <https://doi.org/10.3390/vaccines8020153>
17. Jin Z, Du X, Xu Y, Deng Y, Liu M, Zhao Y et al (2020) Structure of M(pro) from SARS-CoV-2 and discovery of its



- inhibitors. *Nature* 582(7811):289–93. <https://doi.org/10.1038/s41586-020-2223-y>
18. Douangamath A, Fearon D, Gehrtz P, Krojer T, Lukacik P, Owen CD et al (2020) Crystallographic and electrophilic fragment screening of the SARS-CoV-2 main protease. *Nat Commun* 11(1):5047. <https://doi.org/10.1038/s41467-020-18709-w>
  19. Wang H, He S, Deng W, Zhang Y, Li G, Sun J et al (2020) Comprehensive insights into the catalytic mechanism of middle east respiratory syndrome 3C-like protease and severe acute respiratory syndrome 3C-like protease. *ACS Catal* 10(10):5871–90. <https://doi.org/10.1021/acscatal.0c00110>
  20. Su HX, Yao S, Zhao WF, Li MJ, Liu J, Shang WJ et al (2020) Anti-SARS-CoV-2 activities in vitro of Shuanghuanglian preparations and bioactive ingredients. *Acta Pharmacol Sin* 41(9):1167–77. <https://doi.org/10.1038/s41401-020-0483-6>
  21. Dai W, Zhang B, Jiang XM, Su H, Li J, Zhao Y et al (2020) Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease. *Science* 368(6497):1331–5. <https://doi.org/10.1126/science.abb4489>
  22. Kneller DW, Phillips G, O'Neill HM, Jedrzejczak R, Stols L, Langan P et al (2020) Structural plasticity of SARS-CoV-2 3CL M(pro) active site cavity revealed by room temperature X-ray crystallography. *Nat Commun* 11(1):3202. <https://doi.org/10.1038/s41467-020-16954-7>
  23. Ma C, Sacco MD, Hurst B, Townsend JA, Hu Y, Szeto T, et al. Boceprevir, GC-376, and calpain inhibitors II, XII inhibit SARS-CoV-2 viral replication by targeting the viral main protease. *bioRxiv*. 2020:2020.04.20.051581. doi: <https://doi.org/10.1101/2020.04.20.051581>.
  24. Muramatsu T, Takemoto C, Kim YT, Wang H, Nishii W, Terada T et al (2016) SARS-CoV 3CL protease cleaves its C-terminal auto-processing site by novel subsite cooperativity. *Proc Natl Acad Sci U S A* 113(46):12997–3002. <https://doi.org/10.1073/pnas.1601327113>
  25. Huynh T, Wang H, Luan B (2020) In silico exploration of the molecular mechanism of clinically oriented drugs for possibly inhibiting SARS-CoV-2's main protease. *J Phys Chem Lett* 11(11):4413–20. <https://doi.org/10.1021/acs.jpclett.0c00994>
  26. Suarez D, Diaz N (2020) SARS-CoV-2 main protease: a molecular dynamics study. *J Chem Inf Model* 60(12):5815–31. <https://doi.org/10.1021/acs.jcim.0c00575>
  27. Frances-Monerris A, Hognon C, Miclot T, Garcia-Iriepa C, Iriepa I, Terenzi A et al (2020) Molecular basis of SARS-CoV-2 infection and rational design of potential antiviral agents: modeling and simulation approaches. *J Proteome Res* 19(11):4291–315. <https://doi.org/10.1021/acs.jproteome.0c00779>
  28. Świderek K, Moliner V (2020) Revealing the molecular mechanisms of proteolysis of SARS-CoV-2 Mpro by QM/MM computational methods. *Chem Sci* 11(39):10626–30. <https://doi.org/10.1039/d0sc02823a>
  29. Ramos-Guzmán CA, Ruiz-Pernía JJ, Tuñón I (2020) Unraveling the SARS-CoV-2 main protease mechanism using multiscale methods. *ACS Catalysis* 10(21):12544–54. <https://doi.org/10.1021/acscatal.0c03420>
  30. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) Development and testing of a general amber force field. *J Comput Chem* 25(9):1157–74. <https://doi.org/10.1002/jcc.20035>
  31. Wang J, Wang W, Kollman PA, Case DA (2006) Automatic atom type and bond type perception in molecular mechanical calculations. *J Mol Graph Model* 25(2):247–60. <https://doi.org/10.1016/j.jmgm.2005.12.005>
  32. University of California, San Francisco. 2020.
  33. Case DA, Darden TA, Cheatham TE, Simmerling CL, Wang J, Duke RE et al (2012) AMBER 12. University of California, San Francisco
  34. Sondergaard CR, Olsson MH, Rostkowski M, Jensen JH (2011) Improved treatment of ligands and coupling effects in empirical calculation and rationalization of pKa values. *J Chem Theory Comput* 7(7):2284–95. <https://doi.org/10.1021/ct200133y>
  35. Olsson MH, Sondergaard CR, Rostkowski M, Jensen JH (2011) PROPKA3: consistent treatment of internal and surface residues in empirical pKa predictions. *J Chem Theory Comput* 7(2):525–37. <https://doi.org/10.1021/ct100578z>
  36. Price DJ, Brooks CL 3rd (2004) A modified TIP3P water potential for simulation with Ewald summation. *J Chem Phys* 121(20):10096–103. <https://doi.org/10.1063/1.1808117>
  37. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79(2):926–35. <https://doi.org/10.1063/1.445869>
  38. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C (2015) ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J Chem Theory Comput* 11(8):3696–713. <https://doi.org/10.1021/acs.jctc.5b00255>
  39. Ryckaert J-P, Ciccotti G, Berendsen HJC (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J Comput Phys* 23(3):327–41. [https://doi.org/10.1016/0021-9991\(77\)90098-5](https://doi.org/10.1016/0021-9991(77)90098-5)
  40. Fernandes HS, Ramos MJ (2018) Cerqueira NMFS. molUP: A VMD plugin to handle QM and ONIOM calculations using the gaussian software. *J Comput Chem* 39(19):1344–1353. <https://doi.org/10.1002/jcc.25189>
  41. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. *J Mol Graph*. [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5)
  42. Fernandes HS, Sousa SF, Cerqueira N (2019) VMD store-A VMD plugin to browse, discover, and install VMD extensions. *J Chem Inf Model*. 59(11):4519–23. <https://doi.org/10.1021/acs.jcim.9b00739>
  43. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR et al (2009) Gaussian 09. Gaussian, Inc., Wallingford, CT, USA
  44. Lee C, Yang W, Parr RG (1988) Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys Rev B Condens Matter* 37(2):785–9. <https://doi.org/10.1103/physrevb.37.785>
  45. Becke AD (1996) Density-functional thermochemistry. IV. A new dynamical correlation functional and implications for exact-exchange mixing. *J Chem Phys*. <https://doi.org/10.1063/1.470829>
  46. Becke AD (1988) Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys Rev A Gen Phys* 38(6):3098–100. <https://doi.org/10.1103/physreva.38.3098>
  47. Spitznagel GW, Clark T, von Ragué Schleyer P, Hehre WJ (1987) An evaluation of the performance of diffuse function-augmented basis sets for second row elements. Na-Cl. *J Comput Chem* 8(8):1109–16. <https://doi.org/10.1002/jcc.540080807>
  48. Hehre WJ, Ditchfield R, Pople JA (1972) Self-consistent molecular orbital methods. XII. Further extensions of Gaussian—type basis sets for use in molecular orbital studies of organic molecules. *J Chem Phys*. <https://doi.org/10.1063/1.1677527>
  49. Hariharan PC, Pople JA (1973) The influence of polarization functions on molecular orbital hydrogenation energies. *Theoret Chim Acta*. 28(3):213–22. <https://doi.org/10.1007/bf00533485>
  50. Gordon MS, Binkley JS, Pople JA, Pietro WJ, Hehre WJ (1982) Self-consistent molecular-orbital methods. 22. Small split-valence basis sets for second-row elements. *J Am Chem Soc* 104(10):2797–2803. <https://doi.org/10.1021/ja00374a017>
  51. Franci MM, Pietro WJ, Hehre WJ, Binkley JS, Gordon MS, DeFrees DJ et al (1982) Self-consistent molecular orbital methods.



- XXIII. A polarization-type basis set for second-row elements. *J Chem Phys* 77(7):3654–3665. <https://doi.org/10.1063/1.444267>
52. Ditchfield R, Hehre WJ, Pople JA (1971) Self-consistent molecular-orbital methods. IX. An extended Gaussian-type basis for molecular-orbital studies of organic molecules. *J Chem Phys* 54(2):724–728. <https://doi.org/10.1063/1.1674902>
  53. Guo Y, Riplinger C, Becker U, Liakos DG, Minenkov Y, Cavallo L et al (2018) Communication: An improved linear scaling perturbative triples correction for the domain based local pair-natural orbital based singles and doubles coupled cluster method [DLPNO-CCSD(T)]. *J Chem Phys* 148(1):011101. <https://doi.org/10.1063/1.5011798>
  54. Truhlar DG (1998) Basis-set extrapolation. *Chem Phys Lett* 294(1–3):45–8. [https://doi.org/10.1016/s0009-2614\(98\)00866-5](https://doi.org/10.1016/s0009-2614(98)00866-5)
  55. Neese F (2011) The ORCA program system. *WIREs Comput Mol Sci* 2(1):73–8. <https://doi.org/10.1002/wcms.81>
  56. Magalhães RP, Fernandes HS, Sousa SF (2020) Modelling enzymatic mechanisms with QM/MM approaches: current status and future challenges. *Israel J Chem* 60(7):655–66. <https://doi.org/10.1002/ijch.202000014>
  57. Paiva P, Sousa SF, Fernandes PA, João Ramos M (2019) Human fatty acid synthase: a computational study of the transfer of the acyl moieties from MAT to the ACP domain. *Chem Cat Chem* 11(16):3853–64. <https://doi.org/10.1002/cctc.201900548>
  58. Teixeira CSS, Ramos MJ, Sousa SF (2019) Cerqueira NMFSA solving the catalytic mechanism of tryptophan synthase: an emergent drug target in the treatment of tuberculosis. *Chem Cat Chem* 12(1):227–237. <https://doi.org/10.1002/cctc.201901505>
  59. Fernandes HS, Ramos MJ (2018) Cerqueira NMFSA catalytic mechanism of the serine hydroxymethyltransferase: a computational ONIOM QM/MM study. *ACS Catalysis* 8(11):10096–10110. <https://doi.org/10.1021/acscatal.8b02321>
  60. Fernandes HS, Ramos MJ, Cerqueira N (2017) The catalytic mechanism of the pyridoxal-5'-phosphate-dependent enzyme histidine decarboxylase: a computational study. *Chemistry* 23(38):9162–73. <https://doi.org/10.1002/chem.201701375>
  61. Paiva P, Ramos MJ, Fernandes PA (2020) Assessing the validity of DLPNO-CCSD(T) in the calculation of activation and reaction energies of ubiquitous enzymatic reactions. *J Comput Chem* 41(29):2459–68. <https://doi.org/10.1002/jcc.26401>
  62. Bistoni G, Polyak I, Sparta M, Thiel W, Neese F (2018) Toward accurate QM/MM reaction barriers with large QM regions using domain based pair natural orbital coupled cluster theory. *J Chem Theory Comput* 14(7):3524–31. <https://doi.org/10.1021/acs.jctc.8b00348>
  63. Lawal MM, Sanusi ZK, Govender T, Tolufashe GF, Maguire GEM, Honarparvar B et al (2018) Unraveling the concerted catalytic mechanism of the human immunodeficiency virus type 1 (HIV-1) protease: a hybrid QM/MM study. *Struct Chem* 30(1):409–17. <https://doi.org/10.1007/s11224-018-1251-9>
  64. Barman A, Prabhakar R (2013) Elucidating the catalytic mechanism of beta-secretase (BACE1): a quantum mechanics/molecular mechanics (QM/MM) approach. *J Mol Graph Model* 40:1–9. <https://doi.org/10.1016/j.jmgm.2012.12.010>
  65. Ma S, Devi-Kesavan LS, Gao J (2007) Molecular dynamics simulations of the catalytic pathway of a cysteine protease: a combined QM/MM study of human cathepsin K. *J Am Chem Soc* 129(44):13633–45. <https://doi.org/10.1021/ja074222+>
  66. Fan K, Wei P, Feng Q, Chen S, Huang C, Ma L et al (2004) Biosynthesis, purification, and substrate specificity of severe acute respiratory syndrome coronavirus 3C-like proteinase. *J Biol Chem* 279(3):1637–42. <https://doi.org/10.1074/jbc.M310875200>
  67. Solowiej J, Thomson JA, Ryan K, Luo C, He M, Lou J et al (2008) 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* 47(8):2617–30. <https://doi.org/10.1021/bi702107v>
  68. Ramos-Guzmán CA, Ruiz-Pernía JJ, Tuñón I (2021) A microscopic description of SARS-CoV-2 main protease inhibition with Michael acceptors. Strategies for improving inhibitor design. *Chem Sci* 12(10):3489–3496. <https://doi.org/10.1039/d0sc04978f>
  69. Arafet K, Serrano-Aparicio N, Lodola A, Mulholland AJ, González FV, Świderek K et al (2021) Mechanism of inhibition of SARS-CoV-2 Mpro by N3 peptidyl Michael acceptor explained by QM/MM simulations and design of new derivatives with tunable chemical reactivity. *Chemical Science*. 12(4):1433–44. <https://doi.org/10.1039/d0sc06195f>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.