Impact of warhead modulations on the covalent inhibition of SARS-CoV-2 M^pro explored by QM/MM simulations

Sergio Martí,[1] Kemel Arafet,[1] Alessio Lodola,[2] Adrian Mulholland,[3] Katarzyna Świderek,[1,*] Vicent Moliner.[1,*]

1. BioComp Group, Institute of Advanced Materials (INAM), Universitat Jaume I, 12071 Castelló, Spain.
2. Dipartimento di Scienze degli Alimenti e del Farmaco, Università degli Studi di Parma, Italy.
3. Centre for Computational Chemistry, School of Chemistry, University of Bristol, United Kingdom.

ABSTRACT

The COVID-19 pandemic, caused by the novel severe acute respiratory syndrome coronavirus-2, SARS-CoV-2, shows the need for effective antiviral treatments. Here, we present a simulation study of the inhibition of the SARS-CoV-2 main protease (M^{pro}), a cysteine hydrolase essential for the life cycle of the virus. The free energy landscape for the mechanism of the inhibition process is explored by QM/MM umbrella sampling and free energy perturbation simulations at the M06-2X/MM level of theory for two proposed peptidyl covalent inhibitors sharing the same recognition motif while featuring distinct cysteine-targeting warheads. Regardless of intrinsic reactivity of the modelled inhibitors, namely a Michael acceptor and a hydroxymethylketone activated carbonyl, our results confirm that the inhibitory process takes place by means of a two-step mechanism, in which the formation of an ion pair C145/H41 dyad precedes the protein-inhibitor covalent bond formation, in both cases. The nature of this second step appears to be strongly dependent on the functional groups introduced in the warhead: in the present study, while the nucleophilic attack of the C145 sulfur atom on the Cα of the double bond of the Michael acceptor takes place concertedly to the proton transfer from H41 to Cβ, in the compound with an activated carbonyl the sulfur attacks the carbonyl carbon concomitant to the proton transfer from H41 to the carbonyl oxygen through the hydroxyl group. Analysis of the free energy profiles, structures along the reaction path, and interactions between the inhibitors and the different pockets of the active site on the protein shows a measurable impact of the warhead on the kinetics and thermodynamics of the process. The present results can be used as a guide to select warheads to design efficient irreversible and reversible inhibitors of SARS-CoV-2 M^{pro}.

INTRODUCTION

Despite the development of efficient vaccines, the impact of the COVID-19 pandemic around the world, caused by the novel severe acute respiratory syndrome coronavirus-2 – SARS-CoV-2 – has emphasized the need for effective antiviral treatments. Moreover, considering the capabilities of the virus to mutate, as any other viruses that contain RNA genetic material such as this or the influenza viruses, the corresponding risk of a decrease of the effectiveness of the vaccines urged the need of complementary strategies to fight against the pandemia. Many efforts have focused on understanding the life cycle of SARS-CoV-2, which can provide information about possible targets for drug

development.[1-3] Among the proteins involved in the replication of the virus, the main coronavirus protease (SARS-CoV-2 M^{pro}) is a most attractive target due to its intrinsic features, including its distinguishing ability with respect to human proteases to cleave proteins after glutamine residue,[4] a catalytic features which makes M^{pro} unique with respect to human proteases. The most effective M^{pro} inhibitors so far identified, including the clinical candidates PF-00835231, incorporate a glutamine residue or a bioisostere at P1 position (see below) to obtain potency and selectivity and a peptidomimetic scaffold of moderate size endowed with branched and yet hydrophobic substituents at both P2 and P3 positions.[5-7] These compounds act by a covalent-based mechanism, so a reactive 'warhead' is required, i.e. an electrophilic group responsible for the covalent bond formation between the active site cysteine residue (Cys145), previously activated by a histidine residue (His41), and the inhibitor. Warheads so far employed ranged from classical Michael acceptors (MAs) to activated carbonyl derivatives, including alpha-ketoamides, aldehydes and hydroxymethylketone (HMK).[4,6,8] Nevertheless, previous studies of the inhibition of M^{pro} do not show which is the most reactive warhead to design an efficient inhibitor, nor which is the most appropriate recognition part to increase the affinity of the inhibitor by the enzyme active site. Thus, there is a need to understand the effects of warheads on reactivity and e.g. reversibility of inhibitors, which may significantly affect their pharmacokinetics.[9] A plethora of different computational methods have used since the emergence of COVID-19 for the discovery of small-molecule therapeutics.[7,10] Regarding the inhibition of M^{pro}, modelling can contribute effectively to noncovalent inhibitor development,[11-13] while the use of methods based on multiscale quantum mechanics/molecular mechanics (QM/MM) potentials can help with covalent inhibitor design. We recently studied the mechanism of the covalent inhibition of the peptidyl-MA compound **N3** designed by Jin and colleagues,[6] and two designed MA compounds (**B1** and **B2**) by QM/MM molecular dynamics (MD) methods.[14] Our results indicated that both designed compounds may be promising candidates as drug leads against COVID-19, one as an irreversible inhibitor and one as a potential reversible inhibitor.

As previously proposed from X-ray diffraction studies,[3] and later supported by computational studies using different approaches, the chemical reaction leading to M^{pro} inactivation requires the imidazole group of H41 to activate the SH group of C145 to generate a highly nucleophilic thiolate (CysS⁻) that would readily react with the inhibitor.[14-19] According to the recent literature, the equilibrium between the neutral

dyad and the CysS⁻/HisH⁺ ion pair appears to be tipped in favour of the neutral pair by ligand binding, but it may depend on the stereoelectronic properties of the ligand itself. Thus, while some studies report a neutral dyad significantly more stable than the ion pair (by ca. 11 kcal·mol⁻¹) and even not being an stable state,[19] others suggest the ion pair is not so destabilized with respect the initial neutral dyad,[14,16] or even slightly more stable than the initial neutral dyad (i.e. our previous study with **B1**).[14]

Our previous study on the proteolysis reaction of SARS-CoV-2 M^{pro} using the polypeptide Ac-Val-Lys-Leu-Gln-ACC as a substrate (with ACC being a fluorescent 7-amino-4-carbamoylmethylcoumarin tag) suggests that the mechanism of action of this enzyme slightly differs from other cysteine proteases.[15] First of all, the enzyme:substrate initial complex is the neutral C145/H41 dyad (**E:I**) instead of the ion pair dyad C145⁻/H41⁺ (**E$^{(+/-)}$:I**). This result is in agreement with studies carried out by us, and others, using different inhibitors and substrates,[14,17,19] but in contrast with the protonation state of the catalytic dyad suggested from the ligand-free SARS–CoV-M^{pro} recently solved by neutron crystallography at pH 6.6.[20] Nevertheless, as already pointed out, questions remain about how pH or the presence of an inhibitor or substrate influence the protonation state of the dyad in SARS-CoV M^{pro}.[17]

The acylation reaction consists of a proton transfer from C145 to the H41 concomitant, or not, with the nucleophilic attack on the carbonyl carbon atom of the peptide bond by the sulfur atom of C145, leading to a pseudo-stable intermediate. Then, the cleavage of the peptide bond by M^{pro} is assisted by proton transfer from the protonated H41⁺ to the substrate, forming a stable acyl-enzyme covalent intermediate. Thus, while modelling of the proteolysis reaction showed that the formation of the ion pair and the attack of the Cys145 to the carbonyl carbon of the substrate took place in a single step,[15] in the inhibition reaction by **N3**, or our designed **B1** and **B2** MA compounds,[14] or the simulation with M^{pro}-substrate peptide models,[17] the reaction appears to proceed in a stepwise manner: in the first step, Cys145 is activated by His41, forming the ion pair **E$^{(+/-)}$:I**, followed in the second step by attack of the sulfur atom of Cys145 on the C$_\beta$ atom of the inhibitor and proton transfer from His41 to the C$_\alpha$ atom of the inhibitor, leading to a stable covalent **E-I** intermediate. The rate-limiting step of the process, in all three cases, was enzyme-inhibitor covalent bond formation, with activation free energies ranging from 11.8 to 9.8 kcal·mol⁻¹.[14] From the thermodynamic point of view, the exergonic process obtained with **N3** (reaction energy –17.9 kcal·mol⁻¹) was consistent with its experimentally observed stability (e.g. revealed by X-ray crystallographic structures)·[6]

while the inactivation reactions of M^{pro} with **B1** and **B2** were also exergonic but very different from each other (–27.9 and –11.4 kcal·mol^{-1}, respectively), predicting that compound **B1** would be an irreversible inhibitor, but that compound **B2** would have a more reversible character. Analysis of the QM-MM interaction energies between the different residues located in the substrate-binding pockets of M^{pro} and the peptide (in the study of the proteolysis reaction) or the inhibitor (in the case of the inhibitors) confirms the predictions assumed during the design of **B1** and **B2**; i.e. the interactions between the protein and the inhibitors are dominated by those in the P1:::S1 site. Thus, our previous results indicate that a low barrier Cys145 covalent modification can be obtained by modulating either the recognition portion or the warhead. Interactions between the recognition moiety and M^{pro} active site affect the chemical step because they dictate the pose of the inhibitor in the active site of the enzyme. Consequently, to design an efficient inhibitor, the presence of a reactive warhead and the interactions between the recognition moiety of the inhibitor and the different sub-sites of the binding pocket of the protein must be taken into account.[21] The results from previous studies on this and related cysteine proteases can be used to guide the design and QM/MM simulations can provide a useful tool to investigate the reactivity of designed covalent inhibitors within their protein targets.

B3

B4
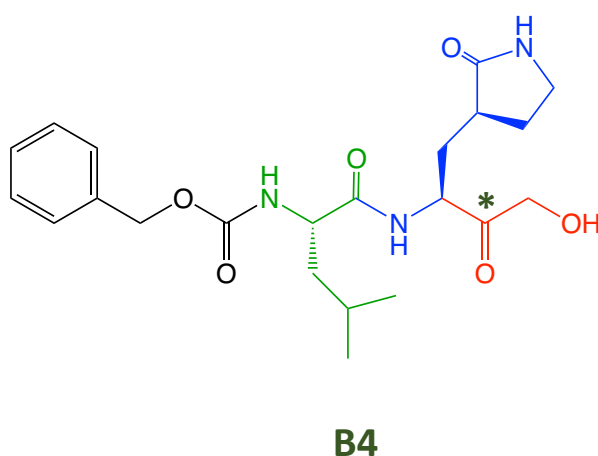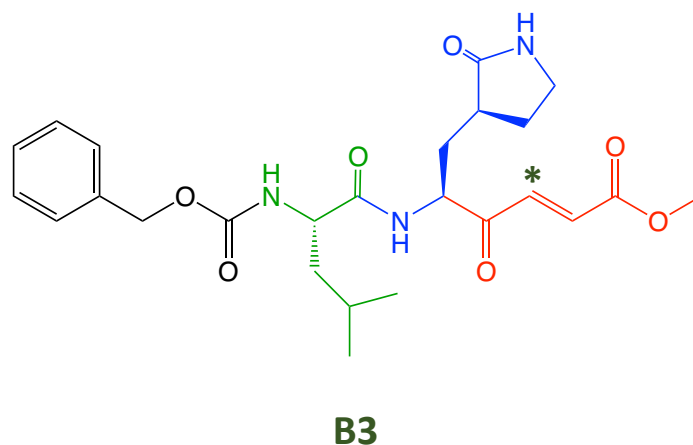
Scheme 1. Chemical structures of the proposed (**B3** and **B4**) Michael acceptor inhibitors of SARS-CoV-2 M[pro]. The warheads are highlighted in red, while P1, P2 and P3 fragments are in blue, green and black, respectively. The subpockets of the active site are labeled with S numbering complementary to fragments of the inhibitor. Asterisks indicate the main reactive centre of the inhibitors.

Here, we propose and investigate the inhibition of the SARS-CoV-2 M[pro] by two potential covalent (peptidyl) inhibitors endowed with two chemically diverse warheads. Building upon the findings on information derived from our previous studies on the proteolysis of M[pro],[15] and on the reaction of the inhibition with several peptidyl irreversible inhibitors,[14] two compounds **B3** and **B4** are proposed (Scheme 1). A methyl oxo-enoate was used in **B3,** inspired by dimethylfumarate structure,[22,23] while a hydroxymethylketone (HMK) was used as warhead in **B4** compound. This reactive group also present in the structure of PF-00835231 M[pro] inhibitor, now in clinical trial[5]. The recognition part possessed by both **B3** and **B4** compounds, was selected on the based on QM/MM results obtained with previously proposed inhibitors **B1** and **B2** as well as by suggestions emerged, from QM-MM protein-substrate interactions. S2 appears to be a

6

small hydrophobic pocket without strong hydrogen bond interactions with P2. Therefore, an isobutyl group was kept at the P2 site. The S3 subsite is completely exposed to the solvent and only three interactions between the peptide backbone atoms of Lys3 of the substrate and the protein were observed. In addition, the lack of strong hydrogen bond interactions in the S4 sub-site supports the strategy of reducing the size of the inhibitor. Small moieties at position further than P3 were introduced with the aim of improving synthetic accessibility. Also, previous studies[6] suggest that P3 can tolerate substituents of different shape and size, which motivated keeping the same full recognition part and changing only the warhead. This strategy allows deciphering of the effect of the warhead in the inhibition process.

From a mechanistic point of view, the two proposed compounds could potentially react in different manners in the active site of the enzyme, also because their key electrophilic centers not only possess a different chemical environment but they are not topologically equivalent. Thus, as shown in Scheme 2, after the formation of the ion pair $E^{(+/-)}$:I reactant complex with **B3**, the attack of the sulfur of Cys145 to the β-carbon of the substrate can take place, followed by the proton transfer from the protonated His41 to the α-carbon, leading to a stable covalent product **E-I**. Nevertheless, considering the nature of the warhead in **B3**, the final proton transfer could also take place to the carbonyl oxygen atom (**E-I'**). In the case of the inhibition with **B4**, this dual possibility of the final proton transfer does not appear after the acylation of the enzyme because the proton from His41 can only be transferred to the carbonyl oxygen atom of the inhibitor (Scheme 3).
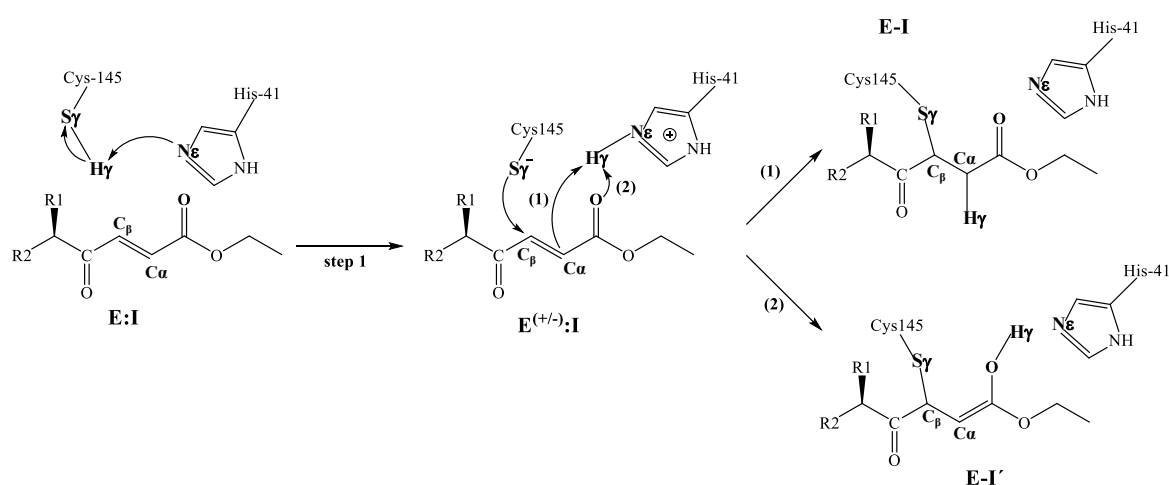
Scheme 2. Proposed mechanism of SARS-CoV-2 M^{pro} cysteine protease inhibition by **B3**. R1 and R2 represent different substituents, as shown in Scheme 1.
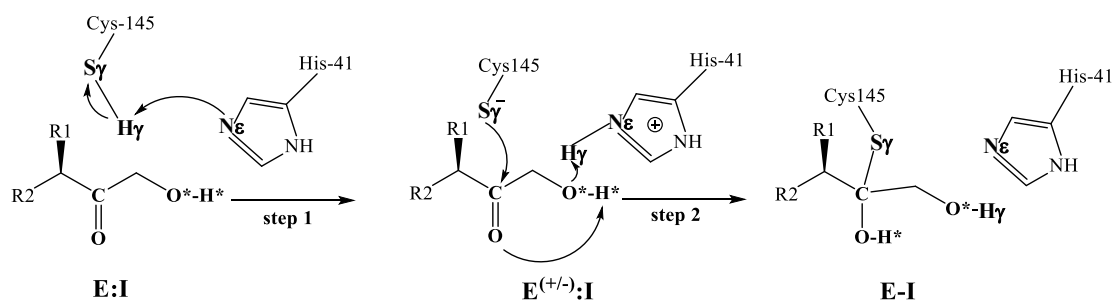
Scheme 3. Proposed mechanism of SARS-CoV-2 M^pro cysteine protease inhibition by **B4**. R1 and R2 represent different substituents, as shown in Scheme 1.

The present study is focused on the computational study of the mechanism of inhibition of M^pro by **B3** and **B4**. The reaction mechanisms for each inhibitor were initially explored by nudged elastic band calculation of the minimum energy paths. Then, two free energy-based methodologies, such as the umbrella sampling (US) and free energy perturbation (FEP) methods, both at density functional theory level combined with classical force fields, were employed to explore the full inhibition process.

RESULTS AND DISCUSSION

The first step in our study was to carry out a deep analysis of the interactions established between the two studied compounds and the active site of M^pro in the initial **E:I** state. Figure 1 shows a schematic representation of these interactions, Figure 2 reports the average interaction energies (electrostatic plus Lennard-Jones) between M^pro residues and inhibitor fragments (see Supporting Information for details). A list of relevant interatomic distances is deposited in the Supporting Information. Analysis of the results confirms the formation of a stable reactant Michaelis complex in both cases, with a similar pattern of interactions. Keeping in mind that the difference between **B3** and **B4** is restricted to the warhead, and in both cases the interactions with the S1' take place through hydrogen bond interactions with the carbonyl oxygen next to P1 that is common in both inhibitors, the results appear as reasonable. Thus, this carbonyl group is interacting with the oxyanion hole located in S1' formed by G143, S144 and C145. In addition, there are some not direct interactions that also stabilize the P1' fragment, such as L27, N28, G146 and S147. The specific favorable interactions between the lactam ring on P1 and S1 are almost equivalent in both inhibitors, mainly through interactions with P140, N142, H163 and E166. The backbone atoms of the residues of the P2 site are the responsible of the interactions with Q189, H164, D187 and M165. Finally, it is worth mentioning the unfavorable interaction with R40, which is ca. 9 Å from P1, thus corresponding to an electrostatic interaction.

The conformation adopted by both compounds in the active site of M^{pro} can be compared with the X-ray crystal structures of related complexes. Thus, the cocrystal structure of the covalent adduct of PF-00835231 bound to SARS CoV-2 3CLpro (PDB code 6XHM)[5] that, as commented above is like **B4**, shows protein-ligand distances in S1, S2 and S3 equivalent to the ones shown in Figure 1b. These similarities are also observed when comparing the distances between the key atoms involved in the inhibition reaction, $S\gamma^{C45}$ – $C^{C=O}$, $S\gamma^{C45}$ – $N\epsilon^{H41}$ and $N\epsilon^{H41}$ – $O*^{OH}$: 1.86, 3.71 and 3.80 Å, respectively, in the crystal structure, and 2.90, 3.13 and 2.34 Å, respectively, in the **B4**. Regarding the **B3**, despite no X-ray structure is available for SARS CoV-2 M^{pro} complexed with a structure comparable to **B3**, there is a cocrystal structure of the covalent adduct 2 of Hoffman and co-workers bound to SARS CoV-1 3CLpro (PDB code 6XHO).[4] Analysis of these structure provide similar conclusions regarding the interactions between the different sub-sites of the active site of the related proteins, CoV-1 and CoV-2 M^{pro}, and the corresponding compounds in the X-ray structure of SARS CoV-1 3CLpro and **B3** in SARS CoV-2 M^{pro}. Obviously, the absence of the carbonyl group at α position of P1' in **2** explains the lack of the interactions with the oxyanion hole of the S1' site. Nevertheless, the comparison of the inter-atomic distances that are related with the inhibition reaction, $S\gamma^{C45}$ – C_β, $S\gamma^{C45}$ – $N\epsilon^{H41}$, and $N\epsilon^{H41}$ – $C\alpha$, also shows similar values: 1.76, 3.96 and 3.23 Å, respectively, in the crystal structure, and 3.29, 3.29 and 4.09 Å, respectively, in **B3**. Obviously, these comparison must be done with caution because the X-ray structures correspond to the protein-inhibitor covalent complex (**E-I** in our schemes 2 and 3) while the **B3** and **B4** structures analyzed at this point correspond to the initial reactant complex **E:I**. Thus, differences observed in the distances defining the attack of the sulfur atom of C145 to the corresponding carbon atom of the inhibitor ($C^{C=O}$ or C_β for **B3** and **B4**, respectively) are as expected. Anyway, the good overlapping of the X-ray structures and the equilibrated **E:I** reactant complex support the quality of our initial state structures (see Figures S10 and S11 in the Supporting Information).

Once confirmed that **E:I** complex represents an stable reactant complex, in both cases, the inhibition reaction was studied according to the general mechanisms proposed in Scheme 2 and 3 for the reaction with **B3** and **B4**, respectively.
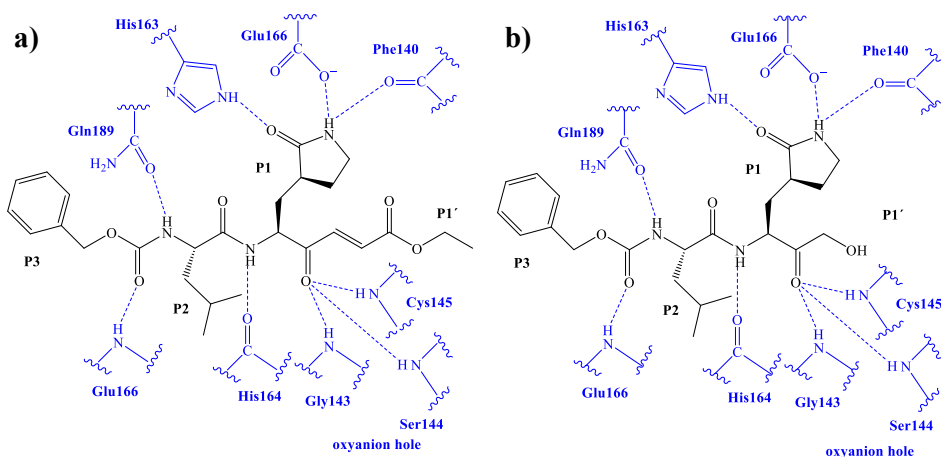
Figure 1. Details of the H-bond interactions between the inhibitor and the active site of the SARS-CoV-2 M^{pro} from QM/MM MD simulations of **B3** (a) and **B4** (b) inhibitors in the **E:I** state.
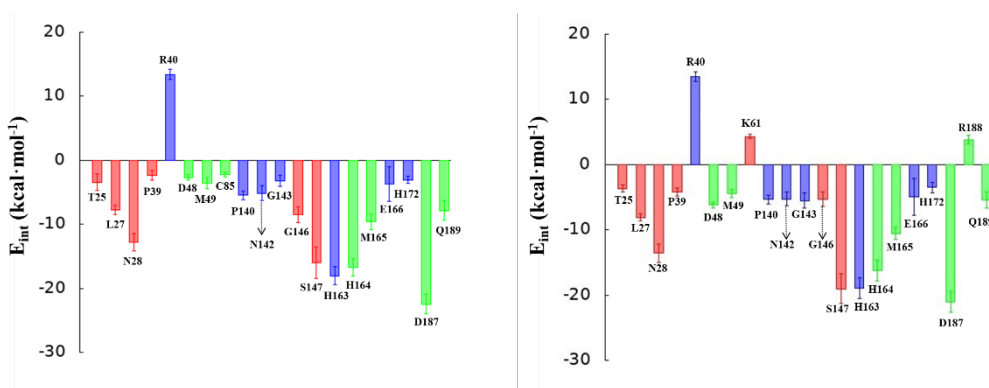

Figure 2. Main average interaction energies (electrostatic plus Lennard-Jones) between residues of Chain-A and each fragment of the inhibitor **B3** (a) and **B4** (b) computed at **E:I** state. Results obtained as an average over 1000 structures from the AM1/MM MD simulations. The red bars corresponds to the P1`:::S1` interactions, the blue bars corresponds to the P1:::S1 interactions, and the green bars corresponds to the P2:::S2 interactions.
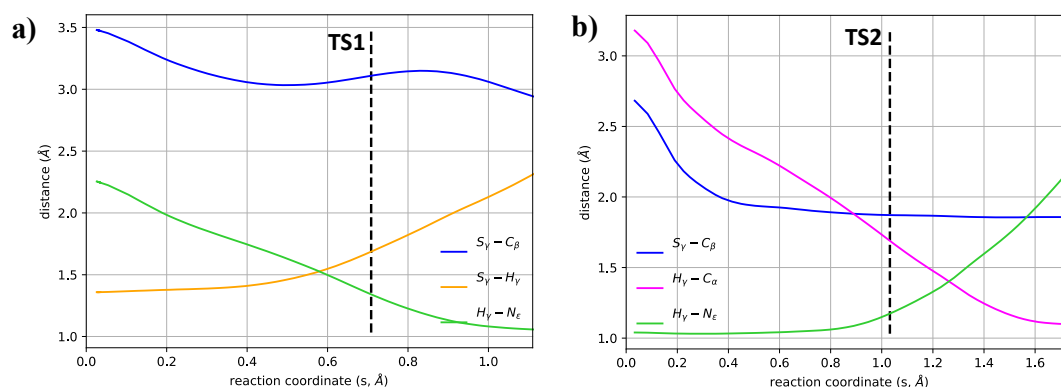
Inhibition of SARS CoV-2 M^{pro} with B3

As shown in Scheme 2, after the C145 is activated by a proton transfer to H41, thus forming the ion pair complex $E^{(+/-)}$**:I**, the covalent complex is formed by the nucleophilic attack of the sulfur atom of C145 to the C_β atom of the **B3** inhibitor. Then, the reaction is completed by the transfer of the proton from the protonated H41 to either the C_α atom, to render the **E-I** final covalent adduct, or to the carbonyl oxygen atom then ending in **E-I'**. Exploration of both mechanisms by M06-2X/6-31+G(d,p)/MM FEP calculations revealed that the formation of the former (i.e. the *direct addition* mechanism) is both thermodynamically and kinetically favored with respect to the formation of **E-I'** (see Figure S4 in Supporting information). Thus, while the reaction that renders the **E-I** product is strongly exergonic (-16.2 kcal·mol^{-1}), the energy of **E-I'** product appears to be 15.2 kcal·mol^{-1} higher than the initial reactants state, **E:I**. These differences in the

reaction energies are also associated, as mentioned, with significant differences in activation energies; 13.7 and 21.0 kcal·mol⁻¹ to form **E-I** and **E-I'**, respectively. Consequently, the much more computationally demanding M06-2X/6-31+G(d,p)/MM US method was applied only to the exploration of the mechanism rendering the **E-I** final product. The resulting free energy profile for the covalent inhibition of SARS CoV-2 M^{pro} with **B3** is depicted in Figure 3, while the evolution of the selected bond distances along the PMF is shown in Figure 4.
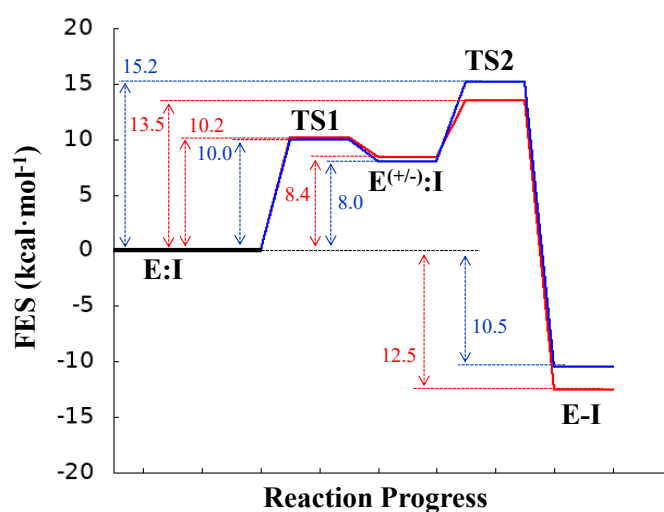

Figure 3. M06-2X/6-31+G(d,p)/MM free energy profiles obtained with umbrella sampling MD for the inhibition mechanism of SARS-CoV-2 M^{pro} cysteine protease by **B3** (red line) and **B4** (blue line) inhibitors at 310 K.

Figure 4. Evolution of the selected bond distances along the PMF of the SARS-CoV-2 M^{pro} inhibition with **B3**. a) Formation of the ion pair **E$^{(+/-)}$:I**. b) Formation of the final **E-I** covalent complex. Vertical dashed lines represent the position of the optimized TS structures.

According to our results, the full reaction mechanism of the inhibition of SARS-CoV-2 M^{pro} cysteine protease by **B3** take place in two steps. First the proton from C145 is transferred to H41 with an activation energy barrier of 10.2 kcal·mol⁻¹. The resulting ion

pair complex, $E^{(+/-)}$:**I**, a zwitterion species that according to previous studies is well described by the M06-2X functional here,[24] is clearly less stable than the initial complex in which both residues of the C145/H41 dyad are in their neutral states (by ca. 8 kcal·mol^{-1}). This result agrees with our previous computational studies of the proteolysis reaction and the inhibition reaction with different inhibitors.[14,15] Thus, despite the quantitative energetic difference between the ion pair and the neutral form appears to be dependent on the substrate, the neutral dyad must be considered as the starting state of the reaction catalyzed by M^{pro}. As shown in Figure 5a, the proton transfer from C145 to H41 is associated with a slight approach of the sulfur atom of the former to the nucleophilic atom of the substrate (from 3.5 to 2.8 Å). Then, the covalent bond formation between C145 and the C_β atom of the substrate takes place concertedly with the proton transfer from the protonated H41 to the C_α atom of the substrate to reach the final **E-I** covalent complex. This step, which is the rate limiting step of the process with a free energy barrier of 13.5 kcal·mol^{-1} measured from the reactants **E:I** complex, appears to be a very asynchronous process (see Figure 5b). The transition state, **TS2**, defined as the maximum of the PMF but also confirmed by optimizing and characterizing a representative structure at M06-2X/6-31+G(d,p)/MM level (see Figure 6 and Tables S3 and S4), is characterized by Sγ-C_β bond formation in a very advanced stage of the process (1.89 Å) but a proton transfer in an early stage of the reaction Hγ-C_α distance of ca. 1.70 Å. This concerted character was also confirmed by tracing the IRC down to the ion pair intermediate and the product from the optimized **TS2**, which in fact was used to generate the free energy profile with the FEP method described above (see Supporting Information). From a technical point of view, it is important to note that both methods, US and FEP, render the same description of the process with just slightly quantitative energetic differences (see Figure 3 vs Figure S4). Finally, the analysis of the average interaction energies (electrostatic plus Lennard-Jones) between residues of M^{pro} and each fragment of the inhibitor **B3** computed at the TS2 shows that the pattern of interactions does not significantly change from the one obtained in the **E:I** complex (see Supporting Information vs Figure 2a).
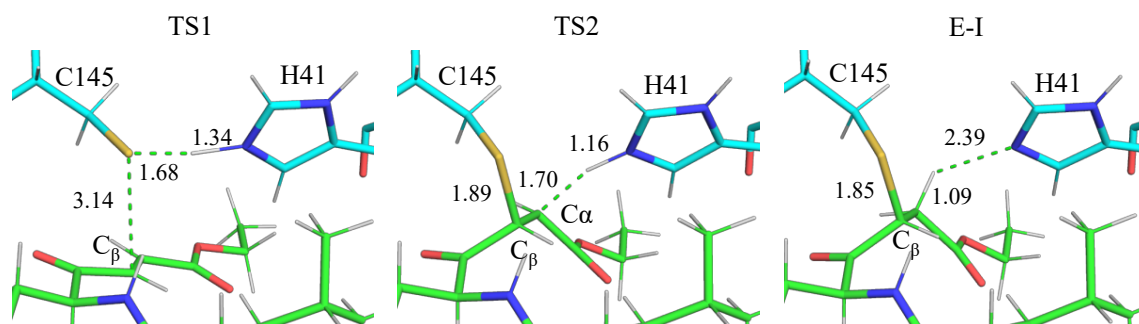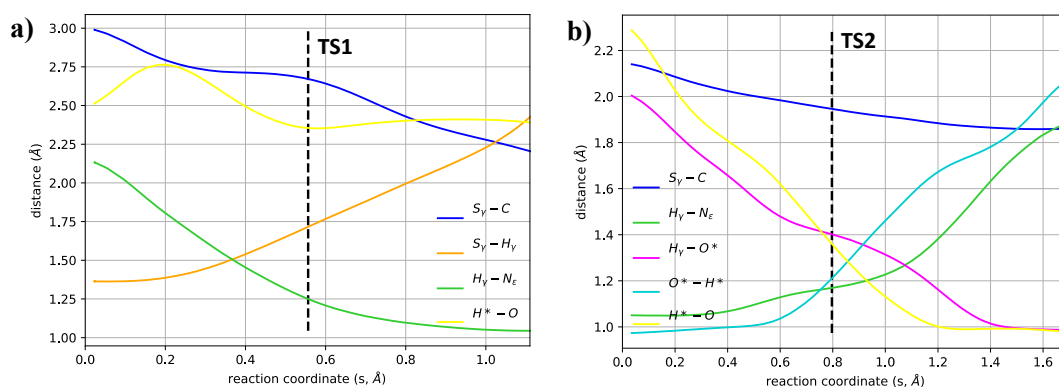
Figure 5. Detail of M06-2X/6-31+G(d,p)/MM optimized structures of the states in the inhibition process of M^{pro} by **B3**. Carbon atoms of the inhibitor are shown in green while those of the catalytic residues Cys145 and His41 are in blue. Key distances are in Å.

Inhibition of SARS CoV-2 M^{pro} with B4

As shown in Figure 3 and schematically depicted in Scheme 3, the inhibition process with **B4** is equivalent to that obtained with **B3**. Thus, the generation of a transient ion pair intermediate **E$^{(+/-)}$:I** by a proton transfer from C145 to H41 precedes the formation of the covalent complex between C145 and the carbonyl carbon atom of **B4**. The first step is virtually the same as in the case of the inhibition with **B3**, confirming that once again, the neutral reactant complex is favored with respect to the ion pair dyad.[14-19] Evolution of the key distances with the progression of the reaction shows that the proton transfer from C145 to H41 is also associated with an approach of the former to the carbonyl carbon atom, from 3.0 to 2.2 Å thus generating a more reactive conformation (Figure 6a). Then, the second step, which as in the case of **B3**, represents the rate-limiting step of the full inhibition process, involves the acylation of the protein together with the proton transfer from the protonated H41 to the carbonyl oxygen atom of the inhibitor with an energy barrier, 15.2 kcal·mol^{-1}. This barrier is slightly higher than that obtained for **B3**, 13.5 kcal·mol^{-1}. This difference of ~2 kcal·mol^{-1} is also observed in the reaction energies, being the reaction with **B3** slightly more exergonic (–12.5 kcal·mol^{-1}), than the reaction with **B4** (–10.5 kcal·mol^{-1}). A recent computational study of the inhibition mechanisms of rhodesain cysteine protease by a dipeptidyl enoate carried out in our laboratory confirmed that the inhibition process can take place through the active site cysteine attack on either the Cβ atom or to the carbonyl carbon atom of the inhibitor in an exergonic process with a low activation energy barrier.[25] Interestingly, and as revealed by the evolution of the interatomic distances monitored in Figure 6b, the proton transfer from the positively charged H41 to the carbonyl oxygen atom does not take place directly but through the hydroxyl group of the substrate. Thus, the proton Hγ is transferred from the Nε of H41 to the oxygen atom of the hydroxyl group, O*, simultaneous with the proton

transfer, H*, from this hydroxyl oxygen atom to the carbonyl oxygen atom, O, of the substrate. A similar mechanism has been found for the inhibition reaction of M^{pro} with PF-00835231 using similar QM/MM methods but with the B3LYP functional.[19] Our activation free energy is 4.5 kcal·mol^{-1} lower than the one obtained in that study, which could be due to chemical and/or methodological differences. In this regard, it has been previously noted that limitations of B3LYP for describing thio-Michael additions.[24] It is also worth mentioning that our ion pair intermediate is clearly a stable minimum in the free energy surface, while a metastable ion pair catalytic dyad is formed by the proton transfer from Cys145 to His41 in that work. The transition state of this second step, defined as the maximum of the PMF but also confirmed by optimizing a representative structure at M06-2X/6-31+G(d,p)/MM level (see Figure 7 and Tables S5 and S6) and tracing the IRC down to the ion pair and the final product complex, support the proposed mechanism. The role played by the terminal hydroxyl group of B4 in the proton transfer from H41 to the carbonyl oxygen atom of the inhibitor agrees with the results of Hoffman and co-workers who measured a drop of potency observed in different HMKs when the terminal hydroxyl group was substituted by other groups.[5]

Figure 6. Evolution of the selected bond distances along the PMF of the SARS-CoV-2 M^{pro} inhibition by **B4**. a) Formation of the ion pair $\mathbf{E^{(+/-)}:I}$. b) Formation of **E-I** covalent complex. Vertical dashed lines represent the position of the optimized TS structures.
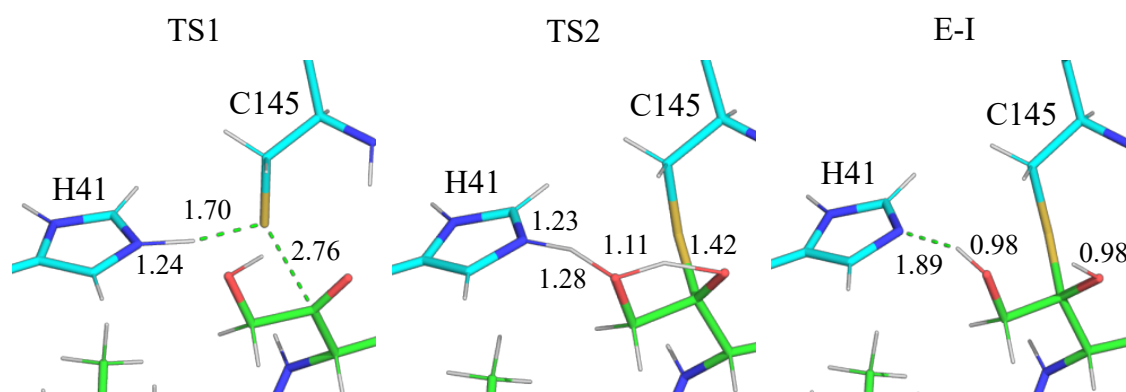
Figure 7. M06-2X/6-31+G(d,p)/MM optimized structures of the important states in the inhibition process of M^{pro} by **B4**. Carbon atoms of the inhibitor are shown in green while those of the catalytic residues Cys145 and His41 are in cyan. Key distances are in Å.

CONCLUSIONS

We report a detailed computational study of the inhibition of SARS-CoV-2 M^{pro} with two proposed covalent (peptidyl) inhibitors: **B3** and **B4**. Both inhibitors share the same recognition part, which is equivalent to the one proposed in our previous study,[14] but differ in the activated carbonyl derivative warhead motifs. The results provide information on the warhead effect in proposed SARS-CoV-2 M^{pro} covalent inhibitors. The full inhibition processes have been explored with two DFT/MM based methods: FEP methods starting from optimized TSs, and US relying on the nudged elastic band and the calculation of the minimum energy paths. There is good agreement between the results derived from these methodologies.

Our results show that the inhibition process with the two compounds takes place by a two-step mechanism, in which the formation of a high energy intermediate (the C145$^-$/H41$^+$ ion pair) precedes the protein-inhibitor covalent bond formation. Analysis of the free energy profiles, the geometries of the states appearing along the reaction path, as well as the interactions established between the inhibitors and the different pockets of the active site on the protein confirms a measurable impact of the warhead in the kinetics and thermodynamics of the process of the second step.

This second step appears to be the rate-limiting step of the process, in both cases, corresponding to the enzyme-inhibitor covalent bond formation, with an activation free energy of 13.5 and 15.2 kcal·mol^{-1} for **B3** and **B4**, respectively. The lower activation free energy of **B3**, together with a higher stabilization of the final covalent product by 2 kcal·mol^{-1}, suggest that future designs should be based on the modification over this kind of warhead introduced in B3. Analysis of the QM-MM interaction energies between the

different residues of the inhibitor and the residues located in the substrate-binding pockets of M^{pro} confirms the predictions assumed during the design of **B3** and **B4,** and the conclusions from geometrical analysis of the structures optimized at the DFT/MM level. In both cases, the interactions between the protein and the inhibitors are dominated by those in the P1:::S1 site, as in our previous studies.[7,8,9] Finally, the good overlapping between the structures of either the reactant complex E:I or the final covalent product E-I obtained with the two tested inhibitors and two cocrystal structures of covalent adduct of similar compounds bound to SARS CoV-2 M^{pro} suggest that **B3** and **B4** can pose in the active site of the protein, and no dramatic geometrical changes occurs in in the protein along the reaction of the inhibition process. The obtained reaction mechanism of **B4** with such low energy barrier, suggests that terminal hydroxyl group is likely an important structural element of its inhibitory activity. This would also mean that modulation of the pKa of this group might represent an effective strategy to improve the potency of this specific class of HMKs.

In summary, our QM/MM study of the inhibition of M^{pro} by two covalent (peptidyl) inhibitors, **B3** and **B4**, which we designed based on medicinal chemistry experience and results derived from our previous computational studies, indicates that **B4** could be used a template to redesign promising candidates as drug leads against COVID-19.

ASSOCIATED CONTENT

Electronic supplementary information (ESI) available:

Computational methods, FF parameters for inhibitors, detail of active site and QM-MM partitioning, M06-2X/6-31+G(d,p)/MM FESs obtained with umbrella sampling and FEP methods, list of average key inter-atomic distances in key states along the reaction path optimized at M06-2X/MM level, protein-substrate non-bonding interaction energies, per residue, cartesian coordinates of the QM sub-set of atoms and full structures (in PDB format) of the rate optimized structures at M06-2X/6-31+G(d,p)/MM level, and figure of inhibitors at **E:I** and **E-I** states overlaying with available related X-ray structures.

REFERENCES:

[1] R. T. Eastman, J. S. Roth, K. R. Brimacombe, A. Simeonov, M. Shen, S. Patnaik, M. D. Hall, *ACS Cent Sci* **2020**, *6*, 672-683.

[2] H. M. Mengist, X. Fan, T. Jin, *Signal Transduction and Targeted Therapy* **2020**, *5*, 67.

[3] G. Zhu, C. Zhu, Y. Zhu, F. Sun, *Current Research in Microbial Sciences* **2020**, *1*, 53-61.

[4] L. Zhang, D. Lin, X. Sun, U. Curth, C. Drosten, L. Sauerhering, S. Becker, K. Rox, R. Hilgenfeld, *Science* **2020**, *368*, 409-412.

[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, B. Taggart, *J. Med. Chem.* **2020**, *63*, 12725-12747.

[6] 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.

[7] F. von Delft, M. Calmiano, J. Chodera, E. Griffen, A. Lee, N. London, T. Matviuk, B. Perry, M. Robinson, A. von Delft, *Nature* **2021**, *594*, 330-332.

[8] 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.

[9] J. Singh, R. C. Petter, T. A. Baillie, A. Whitty, *Nat. Rev. Drug Discovery* **2011**, *10*, 307-317.

[10] E. N. Muratov, R. Amaro, C. H. Andrade, N. Brown, S. Ekins, D. Fourches, O. Isayev, D. Kozakov, J. L. Medina-Franco, K. M. Merz, T. I. Oprea, V. Poroikov, G. Schneider, M. H. Todd, A. Varnek, D. A. Winkler, A. V. Zakharov, A. Cherkasov, A. Tropsha, *Chem. Soc. Rev.* **2021**, DOI:10.1039/D0CS01065K.

[11] M. G. Deshmukh, J. A. Ippolito, C.-H. Zhang, E. A. Stone, R. A. Reilly, S. J. Miller, W. L. Jorgensen, K. S. Anderson, *Structure* **2021**, *29*, 823-833.e825.

[12] C.-H. Zhang, K. A. Spasov, R. A. Reilly, K. Hollander, E. A. Stone, J. A. Ippolito, M.-E. Liosi, M. G. Deshmukh, J. Tirado-Rives, S. Zhang, Z. Liang, S. J. Miller, F. Isaacs, B. D. Lindenbach, K. S. Anderson, W. L. Jorgensen, *ACS Med. Chem. Lett.* **2021**, DOI:10.1021/acsmedchemlett.1c00326.

[13] C.-H. Zhang, E. A. Stone, M. Deshmukh, J. A. Ippolito, M. M. Ghahremanpour, J. Tirado-Rives, K. A. Spasov, S. Zhang, Y. Takeo, S. N. Kudalkar, Z. Liang, F. Isaacs, B. Lindenbach, S. J. Miller, K. S. Anderson, W. L. Jorgensen, *ACS Central Science* **2021**, *7*, 467-475.

[14] K. Arafet, N. Serrano-Aparicio, A. Lodola, A. J. Mulholland, F. V. González, K. Świderek, V. Moliner, *Chem. Sci.* **2021**, *12*, 1433-1444.

[15] K. Świderek, V. Moliner, *Chem. Sci.* **2020**, *11*, 10626-10630.

[16] C. A. Ramos-Guzmán, J. J. Ruiz-Pernía, I. Tuñón, *ACS Catalysis* **2020**, *10*, 12544-12554.

[17] H. T. H. Chan, M. A. Moesser, R. K. Walters, T. R. Malla, R. M. Twidale, T. John, H. M. Deeks, T. Johnston-Wood, V. Mikhailov, R. B. Sessions, W. Dawson, E. Salah, P. Lukacik, C. Strain-Damerell, C. D. Owen, T. Nakajima, K. Świderek, A. Lodola, V. Moliner, D. R. Glowacki, M. A. Walsh, C. J. Schofield, L.

Genovese, D. K. Shoemark, A. J. Mulholland, F. Duarte, G. M. Morris, *BioRxiv* **2021**, DOI:10.1101/2021.06.18.446355.

[18] 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, J. C. Gumbart, *Chem. Sci.* **2021**, *12*, 1513-1527.

[19] C. A. Ramos-Guzmán, J. J. Ruiz-Pernía, I. Tuñón, *ChemRxiv* **2021**, DOI:10.26434/chemrxiv.13340939.v1.

[20] D. W. Kneller, G. Phillips, K. L. Weiss, S. Pant, Q. Zhang, H. M. O'Neill, L. Coates, A. Kovalevsky, *J. Biol. Chem.* **2020**, *295*, 17365-17373.

[21] A. Lodola, D. Callegari, L. Scalvini, S. Rivara, M. Mor, *Methods Mol Biol* **2020**, *2114*, 307-337.

[22] I. Kastrati, M. I. Siklos, E. L. Calderon-Gierszal, L. El-Shennawy, G. Georgieva, E. N. Thayer, G. R. J. Thatcher, J. Frasor, *J. Biol. Chem.* **2016**, *291*, 3639-3647.

[23] A. C. Rosa, E. Benetti, M. Gallicchio, V. Boscaro, L. Cangemi, C. Dianzani, G. Miglio, *Proteomics* **2019**, *19*, 1800301.

[24] J. M. Smith, Y. Jami Alahmadi, C. N. Rowley, *J. Chem. Theory Comput.* **2013**, *9*, 4860-4865.

[25] K. Arafet, F. V. González, V. Moliner, *Chemistry – A European Journal* **2021**, *27*, 10142-10150.