JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC Article

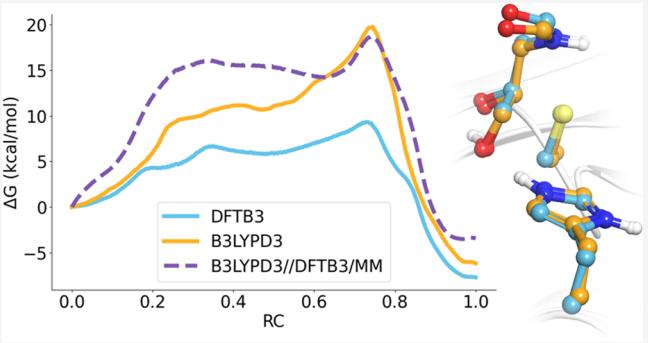
Testing Affordable Strategies for the Computational Study of Reactivity in Cysteine Proteases: The Case of SARS-CoV-2 3CL Protease Inhibition

Carlos A. Ramos-Guzmán, José Luis Velázquez-Libera, J. Javier Ruiz-Pernía,* and Iñaki Tuñón*

Cite This: https://doi.org/10.1021/acs.jctc.2c00294 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Cysteine proteases are an important target for the development of inhibitors that could be used as drugs to regulate the activity of these kinds of enzymes involved in many diseases, including COVID-19. For this reason, it is important to have methodological tools that allow a detailed study of their activity and inhibition, combining computational efficiency and accuracy. We here explore the performance of different quantum mechanics/molecular mechanics methods to explore the inhibition reaction mechanism of the SARS-CoV-2 3CL protease with a hydroxymethyl ketone derivative. We selected two density functional theory (DFT) functionals (B3LYP and M06-2X), two semiempirical Hamiltonians (AM1d and PM6), and two tight-binding DFT methods (DFTB3 and GFN2-xTB) to explore the free energy landscape associated with this reaction. We show that it is possible to obtain an accurate description combining molecular dynamics simulations performed using tight-binding DFT methods and single-point energy corrections at a higher QM description. The use of a computational strategy that provides reliable results at a reasonable computational cost could assist the in silico screening of possible candidates during the design of new drugs directed against cysteine proteases.



1. INTRODUCTION

Cysteine proteases represent one of the four main groups of peptide bond hydrolases. These enzymes are characterized by the use of a cysteine side chain as the nucleophile in charge of the proteolysis reaction. Cysteine proteases are found in all forms of life, including viruses, bacteria, protozoa, fungi, plants, and animals.[1,2] In human cells, cysteine proteases mediate a wide variety of processes, from the bulk digestion of proteins to bone resorption and apoptosis.[3,4] These enzymes are also present in many infectious agents, such as viruses, bacteria, and parasites, making them an attractive target for the development of cysteine protease inhibitors to be used as drugs.

The reaction mechanism of these enzymes involves two basic steps: acylation and deacylation (see Figure 1).[5] Their active sites contain a catalytic dyad formed by a cysteine and a histidine. While the former performs the nucleophilic attack, the latter can act as a proton acceptor/donor during the reaction mechanism. The nucleophile is activated after deprotonation that leads to the formation of the thiolate anion. This anion attacks the electrophilic carbon of the reactive bond, while the leaving amino group is protonated, resulting in the cleavage of the peptide bond. This leads to the release of the first reaction product and the formation of a thioester intermediate. The hydrolysis of this complex produces the second reaction product and the recovery of the initial state of the enzyme. Different variants of this basic reaction mechanism have been described for different enzymes.[6−8] Main differences are associated with the protonation state of the catalytic dyad (both residues being neutral or forming an ion pair), the participation of the catalytic histidine in the proton transfer from the cysteine to the substrate and the concerted or stepwise nature of the basic steps.

The interest in the study of the reactivity and inhibition of cysteine proteases has increased since the COVID outbreak caused by the SARS-CoV-2 virus. This virus uses the transcription machinery of infected cells to translate its genomic material in two large polyproteins that must be cleaved to produce functional proteins for the new generation of viruses. This cleavage process is carried out by two cysteine proteases: the main or 3CL protease, in charge of cleaving the polyprotein in 11 of a total of 14 positions, and the papain-like protease (PPL).[9] The 3CL protease is an interesting target for

Received: March 24, 2022

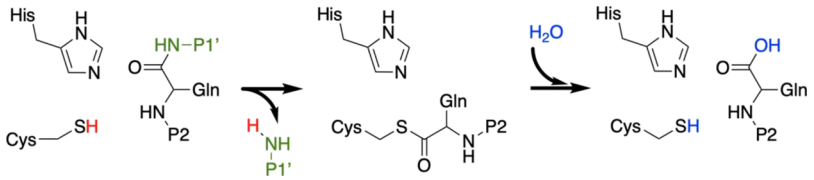


Figure 1. Schematic representation of the reaction mechanism in cysteine proteases showing the acylation and deacylation steps.

the development of inhibitors to be used as antivirals because this enzyme uses a recognition sequence not employed in any of the known human proteases (it cleaves the peptide bond after a Gln residue) and because it is well-conserved among different variants of the virus (the Omicron 3CL protease contains only one mutation, far from the active site, with respect to the wild type presenting similar catalytic properties).[10] Pfizer already announced the development of two inhibitors of the SARS-CoV-2 3CL protease: PF-00835231, a hydroxymethyl ketone derivative,[11] and PF-07321332 (or Nirmatrelvir), a nitrile-derived compound present in Paxlovid (see Figure 2).[12] These compounds are peptidomimetics

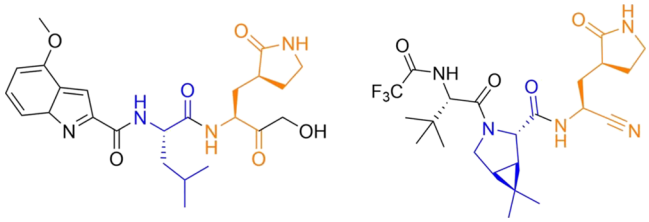


Figure 2. PF-00835231, left, and PF-07321332, right. P2 is highlighted in blue and P1 in orange.

containing a warhead that is able to react with the nucleophilic cysteine, forming a stable acylenzyme complex with the 3CL protease of SARS-CoV-2. Both inhibitors present a γ-lactam ring at the P1 position and a hydrophobic group at P2, mimicking the natural substrate of the enzyme.

In general, cysteine protease inhibitors reduce the quantity of free enzymes, which can help to regulate some cellular processes or to interrupt the replication cycle of infectious agents. Atomistic details of reactions in cysteine protease reactivity can provide valuable information for the development of efficient and specific covalent inhibitors for them. This knowledge can be obtained from a combination of experimental techniques (such as kinetic and structural studies) and computational approaches. Proper simulations of enzymatic reactions must be carried out, including the interactions of the reactive system with a dynamical environment formed by the protein and the solvent. These simulations are nowadays affordable using hybrid quantum mechanics/molecular mechanics (QM/MM) approaches, where a small part of the system is described using QM methods, while the rest is treated using MM.[13,14] The bottleneck in the use of these methodologies is the need to repeat a QM calculation at each step of the simulation, which typically restricts current applications to low-cost QM methods, such as wave-function-based semiempirical Hamiltonians or density functional theory (DFT) tight-binding descriptions. However, the reaction mechanism of cysteine proteases (see Figure 1) involves the deprotonation of a sulfur atom and the nucleophilic attack of the corresponding anion on a carbonyl group, which can compromise the performance of low-cost QM approaches. Very recently, we explored the reaction mechanism of the natural substrate and several inhibitors with the 3CL protease of SARS-CoV-2 using higher-level DFT treatments based on the B3LYP and M06-2X functionals, with the 6-31+G* basis set and including D3 dispersion corrections.[8,15−18] These simulations provided results in excellent agreement with experimental estimations but at a considerable computational cost, which limits the amount of configurations that can be sampled and/or the size of the QM region.

In this work, we have used our previous experience to explore the performance of different low-cost QM descriptions, based on semiempirical and tight-binding treatments, for the description of SARS-CoV-2 3CL protease reactivity. We will show that the combination of an adequate low-level treatment and single-point energy corrections at higher levels can provide results in very good agreement with more expensive DFT/MM simulations. These results pave the way for trustable simulations of the reactivity of cysteine proteases within affordable computational costs, a desirable goal for the incorporation of QM/MM simulations in future studies for the design of new and more efficient inhibitors of cysteine proteases, including those of SARS-CoV-2. In principle, similar strategies for the optimization of QM/MM protocols can be envisaged for the study of different enzymes. These approaches should be particularly interesting to evaluate the impact of modifications in the design of inhibitors against relevant pharmacological targets, as in the case of the SARS-CoV-2. Note that related strategies have been already successfully used to develop fast computational QM/MM assays for the study of class A β-lactamases[19] and fatty acid amide hydrolase.[20]

2. METHODS

In order to analyze the performance of different QM/MM schemes in the analysis of 3CL protease inhibition with PF-00835231, we followed the same computational strategy reported in our previous study.[18] The noncovalent enzyme−inhibitor (EI) complex was built using the PDB structure 6XHM.[11] Maestro[21] and PROPKA[22] were used for the H-bond assignments and to determine the most likely protonation states of the titratable residues at pH 7.4. Regarding the parameters employed to describe the hydroxymethyl inhibitor, they were obtained by following the nonstandard residue parameterization procedure implemented in Amber using the Antechamber[23] program from the AmberTools18[24] package. The restrained electrostatic potential method[25] using the HF/6-31G* level was used to define the atomic charges of the inhibitor, while the standard residues were described using the ff14SB forcefield.[26] The necessary information required to carry out the simulations with the PF-00835231 inhibitor can be found in Supporting Information (frcmod and prepin files). The system was formed by the EI complex with enough Na^+ ions to neutralize the electrostatic charge of the system and a water box, in which the solute's closest atom to the boundaries of the box is at a distance of at

https://doi.org/10.1021/acs.jctc.2c00294
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

least 12 Å. For that procedure, the program *tleap* from the AmberTools18 suite was chosen. The inhibitor was present in both monomers of the enzyme.

We first carried out a classical molecular dynamics (MD) simulation of the EI complex. The standard procedure employed to simulate the system started by a series of minimizations using 500 steps of the steepest descent algorithm, followed by steps of the gradient conjugate method. From this minimized structure, the system was heated using a linear ramp from 0 to 300 K. The first 60 K was raised during 10 ps running Sander in CPU, followed by a 100 ps heating ramp from 60 to 300 K using the Amber19 GPU version of pmemd.[27,28] During this procedure, the heavy atoms of the backbone are restrained using a harmonic restraint 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 7.5 ns. The Berendsen barostat was used to control the pressure, while the Langevin thermostat was chosen to keep the temperature controlled. During this simulation time, the force constant is decreased every 1.25 from 15 to 0 kcal·mol^{-1} $Å^{-2}$ (in 3 unit steps) until the system runs restraint-free during the last 1.25 ns of the equilibration. Then, the system was simulated for 1 μs in the *NVT* ensemble. The SHAKE[29] algorithm was used to constraint the distances of the bonds between heavy atoms and hydrogens. The time step was 2 fs. The electrostatic short-range interactions were evaluated using a 10 Å cutoff, while the long-range electrostatic interaction was evaluated using the particle mesh Ewald.[30,31]

To study the reaction for the formation of the covalent EI complex, a QM/MM description was employed. In these calculations, the B3LYP[32,33] and M062X[34] functionals with a 6-31+G* basis set and D3 dispersion corrections[35] were selected to describe the QM region at the DFT level.[36,37] PM6[38] Hamiltonian was selected among wave function semiempirical descriptions, while GFN2-xTB[39] and DFTB3[40] methods were used as DFT tight-binding methods. The Sander−Gaussian interface[41] was employed to run the QM/MM simulations using a modified version of the AmberTools18 code[42,43] with Gaussian16[44] to perform the DFT calculations. The QM region included the side chain of the residues in the catalytic dyad (His41 and Cys145), backbone of residue P1, and the hydroxymethyl P1' fragment of the inhibitor. The boundary between the QM/MM regions was described using the link atom approach. For all the QM−MM interactions, the cutoff radius used was 15 Å.

In the chemical reaction under study, several degrees of freedom change simultaneously from reactants to products. For this reason, we selected a methodology that help us to explore the associated multidimensional free energy surface (FES). In order to solve the curse of dimensionality, the adaptive string method (ASM)[45] was chosen to find the minimum free energy path (MFEP). In this method, the MFEP can be traced on a multidimensional FES of arbitrary dimensionality defined by a selected set of collective variables (CVs) defined by the bond lengths of all the bonds being formed, broken, or whose formal order changes during the process, as it is the case of the carbonyl bond (see below). Once the MFEP is located, the path is used to define a Path Collective Variable (Path-CV, denoted as s), which is a function that increases monotonically, while the system moves along this path from reactants to products. This makes the Path-CV an adequate reaction coordinate for free energy calculations. As a single reaction coordinate is used, the computational cost is independent of the number of CVs describing the chemical events.

To locate the MFEP, a string that connects the regions of reactants and products is defined. In our case, the string is formed by a set of 96 replicas of the system (string nodes) evenly separated along the path. Every replica is a QM/MM restrained MD simulation. At every simulation step, each node is moved to a region of a lower energy according to their free energy gradient while keeping them equidistant along the string. This series of steps is repeated until the string converges to its MFEP displaying a RMSD below a 0.1 $amu^{1/2}$ Å for at least 1 ps. Replica exchange attempts between nodes were made every 50 steps to increase the convergence. The converged path was then used to define the reaction coordinate (s) to trace the corresponding reaction free energy profile along the MFEP. For the system in which the QM region was described using a DFT level of theory, every node run for 10 ps and the sampled values were integrated with WHAM[46] to obtain the free energy profile along the Path-CV. For the semiempirical calculations, this integration was made over longer simulation times of 80−100 ps. Details about simulation times and the computational cost of each method are provided in Table S1. The force constant values employed to bias the ASM simulations were obtained on-the-fly.[45] This help us to obtain a homogeneous probability density distribution of the reaction coordinate. During the string simulations, the masses of the interchanging protons are set equal to 2 amu, and the time step employed was of 1 fs. Because the total length of the reaction coordinate s was different for each QM/MM method, we present the results as a function of a renormalized reaction coordinate ($RC = (s - s_{min})/(s_{max} - s_{min})$) that takes values between 0 and 1 in all the cases.

3. RESULTS

3.1. QM/MM Descriptions. The catalytic reaction in 3CL protease includes the nucleophilic attack of Cys145 to the electrophilic carbon atom of the substrate and a proton transfer from His41 to the substrate.[8,15−18] The nucleophilic attack requires the activation of Cys145 by means of a proton abstraction carried out by the catalytic His41, resulting in the formation of an ion pair (IP: Cys141-S^- His41H^+). When a peptide substrate is present, this proton is then transferred directly from His41 to the amino group of the target peptide bond.[8] When an inhibitor is present, its protonation is mediated by a water molecule that occupies the position of the amino group in the natural substrate.[15−17] In the case of PF-00835231, the hydroxyl moiety occupies the position of the amino group, and then the proton transfer from His41 to the substrate takes place through the hydroxyl group (see Figure 3).[18,47]

The results obtained for the determination of the minimum free energy path (MFEP) for 3CL protease inhibition by PF-00835231 using different descriptions for the QM subsystem are presented in Figure 4. All of them agree in the timing of the key events during this reaction: the reaction is initiated with the proton abstraction from Cys145 by His41 (resulting in the IP formation), followed by the nucleophilic attack and then the proton transfer from His41 to the carbonyl oxygen atom of the substrate, using the hydroxyl group as a proton shuttle. However, different methods disagree in several details.

Both DFT and MM descriptions (using the B3LYP and M06-2X functionals, both with D3 corrections) provide an

https://doi.org/10.1021/acs.jctc.2c00294
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

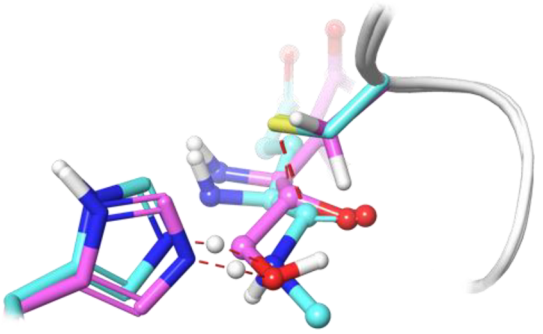


Figure 3. Comparison of the TS structures for the reaction of the 3CLpro with its natural substrate in cyan and the PF-00835231 in pink. Catalytic dyad in tube representation.[18]

almost identical geometrical description of the reaction mechanism, as indicated by the evolution of the CVs along the MFEP (see Figure 4c). Proton transfer from Cys145 to His41 clearly precedes the nucleophilic attack, as reflected in the value of the Sγ−C distance, around 2.7 Å when the proton is at equal distances from the donor and acceptor. The nucleophilic attack is then completed reducing the Sγ−C distance below 2 Å, which is followed by the proton transfer from His41 to the inhibitor. Regarding the proton relay mechanism, the proton transfer from the hydroxyl group to the carbonyl oxygen atom slightly precedes the proton transfer from His41 to the hydroxyl group. The main difference between both DFT descriptions is found in the free energy profiles (see Figure 4a). The M06-2X functional predicts a more stable IP intermediate than B3LYP (7.0 vs 10.7 kcal mol^{-1}, see Table 1). In fact, at the B3LYP level, the IP appears as a flat free energy region (see Figure 4a, black line), while at the M06-2X level, the IP is a clear free energy minimum (see Figure 4a, gray line). M06-2X also results in a smaller activation free energy (16.3 vs 19.7 kcal mol^{-1}, see Table 1) and a more exergonic process (−12.0 vs −6.2 kcal mol^{-1}). The reaction transition state (TS) at the B3LYP level and the rate-limiting TS at the M06-2X level appear with very similar values of the CVs (see Figure 5). PDB files with all the TS structures are provided as Supporting Information. At the TS, the Sγ−C bond (Sγ−C) is almost completely formed, displaying a bond

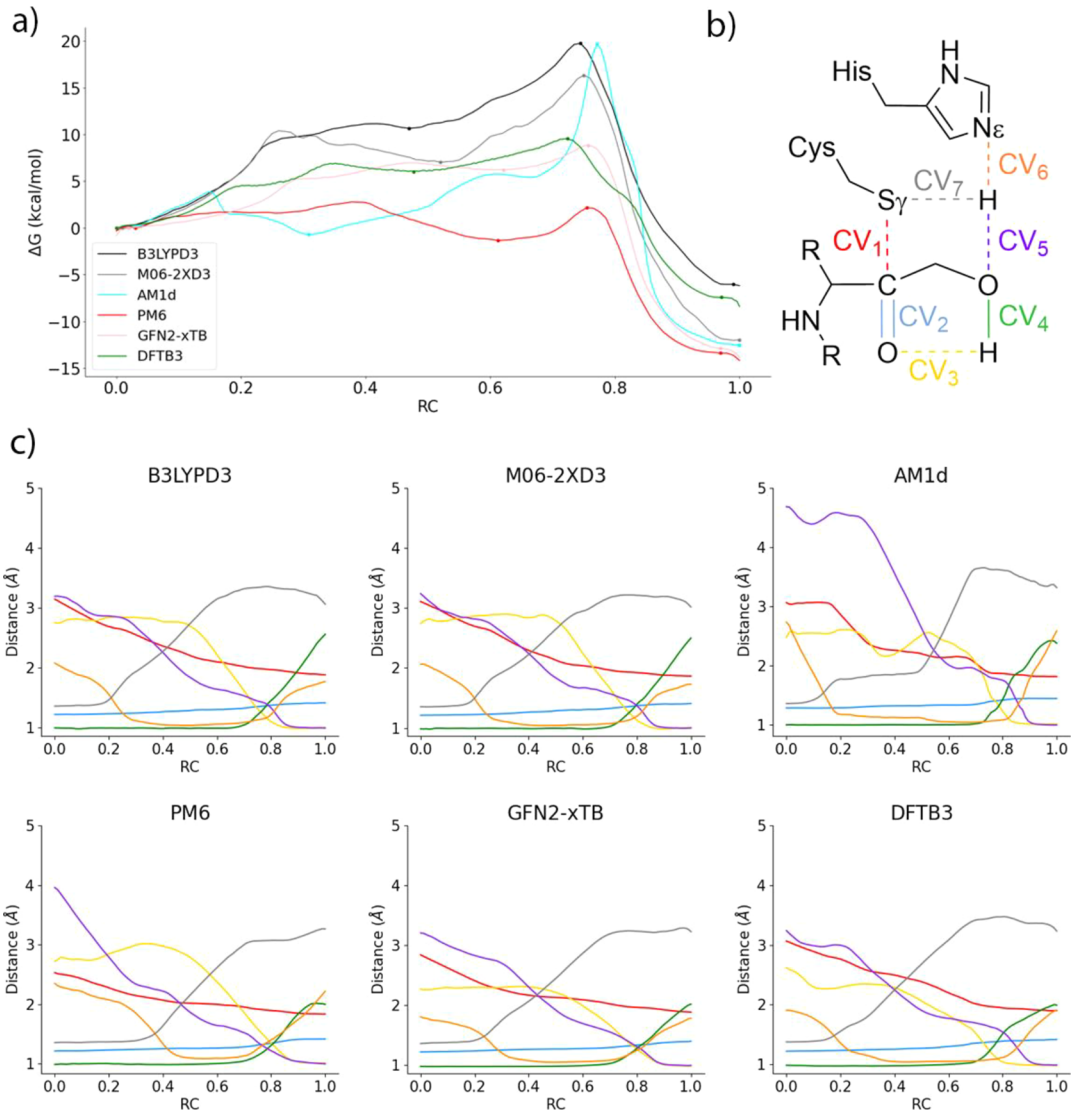


Figure 4. (a) Free energy profiles obtained with different QM levels for the inhibition of 3CLpro with PF-00835231; (b) representation of the CVs used to obtain the MFEPs; and (c) evolution of the CVs along the string nodes for the converged MFEPs.

https://doi.org/10.1021/acs.jctc.2c00294
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 1. Free Energies (in kcal·mol^{-1}) for Different Species Appearing along the Reaction Path Relative to Reactants (R) Using B3LYPD3, M06-2XD3, GFN2-xTB, DFTB3, AM1d, and PM6 Levels for the QM Region

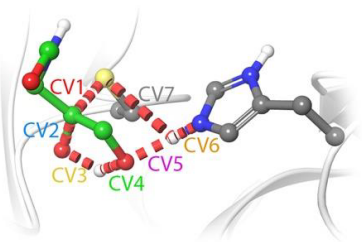
	R	IP	TS	P
B3LYPD3	0.0	10.7	19.8	−6.2
M06-2XD3	0.0	7.0	16.3	−12.0
AM1d	0.0	−0.7	19.7	−12.5
PM6	0.0	−1.3	2.2	−13.4
GFN2-xTB	0.0	6.2	8.8	−12.9
DFTB3	0.0	6.0	9.6	−7.4

length of 1.99/1.94 Å, at the B3LYP/M06-2X level. The two proton transfers, from His41 to the hydroxyl group and from this to the carbonyl oxygen atom, are not very advanced at the TS. The His41Nε−H distance has been lengthened up to 1.12/1.10 Å, while the distance of this proton to the hydroxyl oxygen atom of the inhibitor is 1.47/1.50 Å. Regarding the proton transfer from the hydroxyl group to the carbonyl oxygen, the distances of the proton to the donor and acceptor atoms of 1.16/1.11 and 1.39/1.43 Å, respectively, are just slightly more advanced than the transfer from His41.

Although the free energy profiles are not dramatically different, it is of interest to discuss the reliability of both functionals to characterize the energetics of the reaction under study. One must consider that we are interested in a multistep process and that it can be difficult to find a functional describing equally well all the steps. The B3LYP functional is one of the best functionals to characterize the energetic of the proton transfer between cysteine and histidine when compared to higher-level electronic structure methods.[48] On the other hand, it is also known that this functional can present problems to describe correctly enolate and carbanion intermediates.[49,50] However, this limitation is not much important in the system studied because the nucleophilic attack takes place together with the proton transfer from His41 to the substrate. In this case, an enolate is not fully formed since the negative charge is compensated by the incoming proton. In fact, the Sγ−C distance presents an almost identical evolution along the MFEP with the two functionals presented here (see Figure 4c). Our previous results with other inhibitors also indicate that B3LYP correctly describes the nucleophilic attack of a thiolate to a Michael acceptor if this is accompanied by a proton transfer.[15] On the experimental side, there are no direct measurements of the rate constant for the inhibition of 3CL protease with PF-00835231, and then, the experimentally derived activation free energy is not known. However, the free energy barrier derived from the experimental rate constant for a closely related aldehyde inhibitor, GC373, at 30° is 21.1 kcal·mol^{-1},[18,51] a value close to our B3LYP estimation for PF-00835231. The agreement also obtained between the theoretical and experimental values of the activation free energy for the reaction of the natural substrate[8] and the N3 inhibitor[15] with 3CL protease provides further support to the validity of the B3LYP functional in the study of these processes.

We also located the MFEPs corresponding to two density functional tight-binding methods in combination with the MM potential (GFN2-xTB/MM and DFTB3/MM). The geometrical descriptions obtained from these two methods agree well with the DFT/MM results previously obtained. The main difference with respect to the DFT/MM results, which does not affect the description of the chemical steps, is the formation of a hydrogen bond between the hydroxyl group and the carbonyl oxygen atom of the substrate at the reactant state of the process at the GFN2-xTB level (note the smaller value of CV3 in Figure 4c for the reactant state when using GFN2-xTB). The evolution of the bond-breaking and bond-forming distances follows the same timing than that observed when B3LYP and M06-2X functionals are employed. The reaction is triggered by the abstraction of the proton of Cys165 by His41, followed by the formation of the Sγ−C bond and the protonation of the substrate mediated by the hydroxyl group. GFN2-xTB shows a shorter Sγ−C distance when the proton is being transferred to His41, about 2.4 Å, while DFT and DFTB3 calculations predict a Sγ−C distance of about 2.7 Å. GFN2-xTB also predicts a slightly more concerted character for the proton transfers between His41, the hydroxyl group, and the carbonyl oxygen atom of the substrate, as observed in the evolution of the respective CVs (CV3, CV4, CV5, and CV6) in Figure 4c. The shape of the free energy profiles obtained at GFN2-xTB and DFTB3 levels (see Figure 4a) also agrees qualitatively with the DFT results, showing a TS corresponding to the proton transfer from His41 to the substrate, although the barriers are substantially smaller (8.8 and 9.6 kcal·mol^{-1}, using GFN2-xTB and DFTB3, respectively, see Table 1). Both tight-binding methods also agree with B3LYP results in describing the IP as a metastable species presenting a flat free energy landscape around (see pink and green lines in Figure 4a).

The two semiempirical methods used in this work (AM1d and PM6) also agree in the general description of the reaction mechanism for 3CL protease inhibition with PF-00835231. However, the MFEPs located using these QM descriptions present noticeable differences with respect to the DFT/MM results. At the PM6 level, the Sγ−C distance at the reactant state is substantially shorter than that with other methods, about 2.5 Å; while using other QM levels, the distance is closer or longer than 3.0 Å (see red lines in Figure 4c). Using PM6,



	CV1	CV2	CV3	CV4	CV5	CV6	CV7
B3LYPD3	1.99	1.35	1.39	1.16	1.47	1.12	3.34
M06-2XD3	1.94	1.34	1.43	1.11	1.50	1.10	3.21
AM1d	1.88	1.38	1.57	1.14	1.81	1.05	3.63
PM6	1.92	1.31	1.60	1.13	1.37	1.24	3.06
GFN2-xTB	1.97	1.32	1.34	1.20	1.36	1.19	3.24
DFTB3	2.01	1.36	1.49	1.06	1.69	1.06	3.42

Figure 5. Representation of the rate-limiting TS for 3CLpro inhibition (left) and values (in Å) of all the CVs corresponding to the MFEP at the reaction TS, obtained with different QM approaches (right).

https://doi.org/10.1021/acs.jctc.2c00294
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

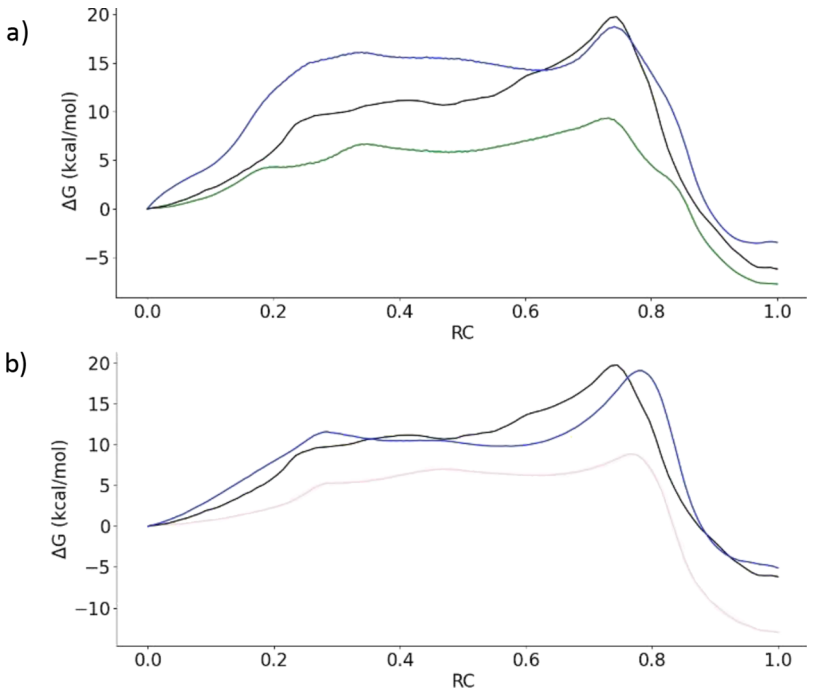


Figure 6. Free energy profiles for 3CLpro inhibition with PF-00835231 obtained with different QM/MM methods: (a) B3LYPD3/MM (in black), DFTB3/MM (in green), and the corrected B3LYPD3//DFTB3/MM profile using eq 1 (in blue). (b) B3LYPD3/MM (in black), GFN2-xTB/MM (in pink), and the corrected B3LYPD3//GFN2-xTB/MM profile using eq 1 (in blue).

when the proton of the Cys145 side chain is being abstracted by His41, Sγ−C is almost completely formed; while in other cases, the bond is formed only after the deprotonation of the Sγ atom. Regarding the proton transfer from His41 to the substrate, the PM6 method predicts a reversed ordering for the proton relay mechanism: the proton transfer from His41 to the hydroxyl group precedes the proton transfer from this group to the carbonyl oxygen atom of the substrate (see the evolution of CV3, CV4, CV5, and CV6 in Figure 4c). Instead, the AM1d Hamiltonian predicts the same ordering than the DFT/MM results. However, this last method fails to describe the formation of the Sγ−C bond. First, the proton transfer from Cys145 to His41 is decoupled from the nucleophilic attack; the Sγ−C distance starts to decrease only once the sulfur atom has been deprotonated (see red line in Figure 4c). Second, the Sγ−C bond is only fully formed once the carbonyl oxygen atom has been protonated, reflecting the inaccurate description of the carbanionic species with this semiempirical Hamiltonian. The energetic description is not much better. With both semiempirical Hamiltonians, the IP form is predicted to be a minimum even more stable than the reactants, with relative free energies of −1.3 and −0.7 kcal·mol^{-1} (see Table 1) with respect to the reactants, at the PM6 and AM1d levels, respectively. Instead, the AM1d Hamiltonian predicts a much more reasonable activation free energy. The free energy differences of the rate-limiting TSs with respect to the reactants are 2.2 and 19.7 kcal·mol^{-1}. In all cases, the reaction is clearly exergonic.

3.2. Dual-Level Approach. According to our previous results, the DFTB3/MM and GFN2-xTB/MM levels provide a geometrical description of the 3CL protease inhibition process close to that obtained at higher, but computationally much more expensive, levels. In geometrical terms, the MFEPs obtained using DFTB3 and GFN2-xTB are close to those obtained using B3LYP and M06-2X functionals, while they can be obtained at a fraction of their computational cost. However, as discussed above, these tight-binding methods clearly underestimate the activation free energy and the free energy cost required to form the ionic pair form of the catalytic dyad.

The previous observation suggests a computational scheme to treat the inhibition of 3CL protease, and hopefully other chemical reactions in cysteine proteases, based on the combination of tight-binding/MM simulations with a posteriori single-point energy corrections at a higher DFT or ab initio level. This combination requires much less computational effort than direct DFT/MM or ab initio/MM schemes, allowing us to explore longer simulation times and/or larger QM regions if required. In this dual-level treatment, the free energy profile obtained at the DFTB3/MM or GFN2-xTB/MM levels is corrected according to the following scheme:

$$\Delta G_{\mathrm{corr}}(s) = \Delta G_{\mathrm{LL}}(s) + \mathrm{Spl}[\Delta E_{\mathrm{LL}}^{\mathrm{HL}}(s)] \quad (1)$$

where s is the Path-CV used to obtain the free energy profile along the MFEP, Spl denotes a spline function, and $\Delta E_{\mathrm{LL}}^{\mathrm{HL}}(s)$ is a correction term taken as the difference between a high-level (HL) energy calculation for the QM subsystem at the coordinate s and the low-level (LL) result. In our case, we corrected the DFTB3/MM and GFN2-xTB/MM values (our LL methods) by means of single-point energy calculations at the B3LYPD3/MM level (the HL method). The structures selected to calculate the correction were obtained after LL optimizations at different s values corresponding to the final positions of the string nodes. The optimizations along the Path-CV were done iteratively following forward and backward directions until the convergence of consecutive energy profiles. For each node of the string, we carried optimizations combining the steepest descent and the conjugated gradient optimization algorithms, as implemented in Amber. In these optimizations, harmonic restraints are applied along the Path-

https://doi.org/10.1021/acs.jctc.2c00294
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

CV and orthogonally, relaxing both the QM and MM regions. We refer to these correction schemes as HL//LL/MM.

The resulting free energy profiles are presented in Figure 6. The corrected B3LYPD3//DFTB3/MM free energy profile correctly reproduces the free energy height of the rate-limiting TS at the B3LYPD3/MM level, predicting an activation free energy of 18.7 kcal·mol^{-1}, to be compared with the reference value of 19.7 kcal·mol^{-1}. The main difference with respect to the B3LYPD3/MM free energy profile is the overestimation of the free energy associated with the IP, probably due to an underestimation of the magnitude of the QM−MM interactions stabilizing the charge separation at the lower level. Instead, the corrected GFN2-xTB/MM free energy profile reproduces better the relative free energy of the IP when compared to the B3LYPD3/MM profile and also the activation free energy (19.1 vs 19.7 kcal·mol^{-1}). Regarding the geometry of the TS, the MFEP values reported in Table 2

Table 2. Values (in Å) of All the CVs Corresponding to the MFEP at the TSs Obtained with the Different QM Levels, Including B3LYPD3, Tight-Binding Methods (DFTB3 and GFN2-xTB), and Correction Schemes (B3LYPD3//DFTB3 and B3LYPD3//GFN2-xTB)

	CV1	CV2	CV3	CV4	CV5	CV6	CV7
B3LYPD3	1.99	1.35	1.39	1.16	1.47	1.12	3.34
DFTB3	2.01	1.36	1.49	1.06	1.69	1.06	3.42
B3LYPD3//DFTB3	1.97	1.37	1.25	1.20	1.64	1.08	3.44
GFN2-xTB	1.97	1.32	1.34	1.20	1.36	1.19	3.24
B3LYPD3//GFN2-xTB	1.96	1.32	1.30	1.23	1.33	1.21	3.24

show that this simple correction scheme does not necessarily result in a TS geometry closer to the target (the one obtained using the B3LYP functional), although the initial geometries provided by the DFTB3 and GFN2-xTB methods were already quite close and then are difficult to improve. The formation of the S−C bond is almost equally well described in all the methods presented in Table 2. Regarding the hydroxyl-mediated proton transfer from His41 to the substrate, the correction scheme provides an improved value over the DFTB3 description, while it is slightly too advanced when corrections are applied to the GFN2-xTB MFEP.

According to these results, the B3LYPD3//DFTB3/MM and B3LYPD3//GFN2-xTB/MM combinations seem to be interesting low-cost strategies to study the reaction mechanism for 3CL protease inhibition, and other reactions taking place in cysteine proteases, provided that the energy results are corrected *a posteriori*. Although the activation free energy would be clearly underestimated, more accurate results can be readily obtained adding single-point energy corrections at higher QM levels, not necessarily restricted to DFT calculations. Note that due to the accuracy of the DFTB3 and GFN2-xTB methods to describe the geometry of the structures appearing along the MFEP, this correction could be limited to the reactant and TS if we were only interested in the activation free energy. This provides an efficient and agile strategy to test the impact of new inhibitors' designs on the kinetics of the formation of the covalent complex, facilitating the in silico screening of putative drugs.

4. CONCLUSIONS

Cysteine proteases are an important group of enzymes in charge of proteolysis reactions using the side chain of a catalytic cysteine as the nucleophile forming a covalent bond with the electrophilic carbon atom of the target peptide bond. Because of the presence of sulfur and of intermediate charged species, the electronic description of the reactions taking place in the active sites of these enzymes may be difficult to describe correctly at a reasonable computational cost. This problem becomes even more critical considering that enough sampling of the protein environment must be considered to properly take into account all the rearrangements needed to accommodate the charge transfer processes taking place during these chemical reactios. That is, one simultaneously needs an accurate electronic description combined with long enough simulation times.

In this work, we have compared the performance of different QM/MM schemes in the study of the 3CL protease inhibition by PF-00835231, a hydroxy methyl ketone compound that forms a covalent bond in the enzymatic active site between the electrophilic carbon atom of the ketone group of the inhibitor and the Sγ atom of the catalytic cysteine. We employed the string method to locate the MFEP using different descriptions for the QM subsystem: two DFT functionals with D3 corrections (B3LYP and M06-2X), two tight-binding methods (DFTB3 and GFN2-xTB), and two semiempirical Hamiltonians (AM1d and PM6). The analysis of the evolution of the CVs used to obtain the MFEP (the distances of all those bonds being broken, formed, or whose bond order changes during the reaction) demonstrates that the two DFT functionals provide a very similar geometrical description of the reaction, although some differences are found regarding the stability of the ion pair, the activation, and the reaction free energies. The use of the two tight-binding methods to describe the QM region results in a geometrical description of the reaction in very good agreement with the DFT results, in particular for the DFTB3 method. Instead, these methods produce activation free energies clearly underestimated with respect to the DFT values. Finally, the MFEPs obtained using the two semiempirical Hamiltonians present larger discrepancies with respect to the DFT results. These methods also overestimate the stability of the ionic pair form of the catalytic dyad.

The results presented here suggest that one can efficiently study the reactivity of the SARS-CoV-2 3CL protease, and probably other cysteine proteases, exploring the reaction mechanism using DFTB3/MM or GFN2-xTB/MM free energy calculations and then correct the activation free energies through single-point calculations using higher-level QM methods. Another possibility would be the use of reweighting schemes to introduce correction to the QM description using short, and then easily affordable, simulations at the reactants and TSs or to add energy corrections predicted by machine learning schemes. In any case, the possibility of having a methodology capable of providing reliable results at a reasonable computational cost paves the way for the use of QM/MM methods in the design of new drugs.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.2c00294.

https://doi.org/10.1021/acs.jctc.2c00294
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Convergence and sampling times for each MFEP and link to the files containing information for classical simulations and TS configurations (PDF)

AUTHOR INFORMATION

Corresponding Authors

J. Javier Ruiz-Pernía − *Departamento de Química Física, Universitat de Valencia, Burjassot, Valencia 46100, Spain;* orcid.org/0000-0002-4640-0419; Email: j.javier.ruiz@uv.es

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

Authors

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

José Luis Velázquez-Libera − *Departamento de Química Física, Universitat de Valencia, Burjassot, Valencia 46100, Spain; Departamento de Bioinformática, Facultad de Ingeniería, Centro de Bioinformática, Simulación y Modelado (CBSM), Universidad de Talca, Talca 3460000, Chile;* orcid.org/0000-0002-4941-6619

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.2c00294

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors acknowledge financial support from Feder funds and the Ministerio de Ciencia, Innovación (project PGC2018-094852-B-C22) and Consellería de Innovación, Universidades, Ciencia y Sociedad Digital, Generalitat Valenciana (GVCOV19/Decreto180/2020). They acknowledge Barcelona Supercomputing Center (BSC) for awarding them access to MareNostrum and the staff from BSC for the technical support. They 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. J.L.V.-L. acknowledges the Computational Center at the University of Talca, whose cluster was obtained with the grant CONICYT-FONDEQUI-PEQM160063. J.L.V.-L. also acknowledges Powered@NLHPC; this research was partially supported by the supercomputing infrastructure of the NLHPC (ECM-02). The authors also deeply acknowledge Dr. Kirill Zinovjev for assistance in the adaptation of the code and helpful discussions for the use of the string method.

REFERENCES

(1) Sharma, A.; Padwal-Desai, S. R.; Ninjoor, V. Intracellular Hydrolases of Aspergillus, Parasiticus and Aspergillus, Flavus. *Biochem. Biophys. Res. Commun.* **1989**, *159*, 464−471.

(2) Mizuno, K.; Nakamura, T.; Takada, K.; Sakakibara, S.; Matsuo, H. A Membrane-Bound, Calcium-Dependent Protease in Yeast α-Cell Cleaving on the Carboxyl Side of Paired Basic Residues. *Biochem. Biophys. Res. Commun.* **1987**, *144*, 807−814.

(3) Chapman, H. A.; Riese, R. J.; Shi, G.-P. Emerging Roles For Cysteine Proteases In Human Biology. *Annu. Rev. Physiol.* **1997**, *59*, 63−88.

(4) Vasiljeva, O.; Reinheckel, T.; Peters, C.; Turk, D.; Turk, V.; Turk, B. Emerging Roles of Cysteine Cathepsins in Disease and Their Potential as Drug Targets. *Curr. Pharm. Des.* **2007**, *13*, 385−401.

(5) Brocklehurst, K.; Willenbrock, F.; Sauh, E. Chapter 2 Cysteine Proteinases. In *Hydrolytic Enzymes*; Elsevier: New York, 1987; pp 39−158.

(6) 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*, 212−224.

(7) Arafet, K.; Ferrer, S.; Moliner, V. Computational Study of the Catalytic Mechanism of the Cruzain Cysteine Protease. *ACS Catal.* **2017**, *7*, 1207−1215.

(8) 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.

(9) 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.

(10) Sacco, M. D.; Hu, Y.; Gongora, M. V.; Meilleur, F.; Kemp, M. T.; Zhang, X.; Wang, J.; Chen, Y. The P132H Mutation in the Main Protease of Omicron SARS-CoV-2 Decreases Thermal Stability without Compromising Catalysis or Small-Molecule Drug Inhibition. *Cell Res.* **2022**, 498.

(11) 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.

(12) 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. J. C.; Lee, J. J. C.; 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.; Steppan, 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* **2021**, *374*, 1586−1593.

(13) Senn, H. M.; Thiel, W. QM/MM Methods for Biomolecular Systems. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198−1229.

(14) van der Kamp, M. W.; Mulholland, A. J. Combined Quantum Mechanics/Molecular Mechanics (QM/MM) Methods in Computational Enzymology. *Biochemistry* **2013**, *52*, 2708−2728.

(15) 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*, 3489−3496.

(16) 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*, 4157−4168.

(17) 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.

(18) 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. Chem., Int. Ed.* **2021**, *60*, 25933−25941.

(19) Hirvonen, V. H. A.; Hammond, K.; Chudyk, E. I.; Limb, M. A. L.; Spencer, J.; Mulholland, A. J.; van der Kamp, M. W. An Efficient Computational Assay for β-Lactam Antibiotic Breakdown by Class A β-Lactamases. *J. Chem. Inf. Model.* **2019**, *59*, 3365−3369.

(20) Lodola, A.; Capoferri, L.; Rivara, S.; Tarzia, G.; Piomelli, D.; Mulholland, A.; Mor, M. Quantum Mechanics/Molecular Mechanics

Modeling of Fatty Acid Amide Hydrolase Reactivation Distinguishes Substrate from Irreversible Covalent Inhibitors. *J. Med. Chem.* **2013**, *56*, 2500−2512.
(21) *Schrödinger Release 2021-1: Maestro*; Schrödinger, LLC: New York, NY. 2021.
(22) Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical p K a Predictions. *J. Chem. Theory Comput.* **2011**, *7*, 525−537.
(23) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graphics Modell.* **2006**, *25*, 247−260.
(24) Cavaleri, F. Paradigm Shift Redefining Molecular, Metabolic and Structural Events in Alzheimer's Disease Involves a Proposed Contribution by Transition Metals. Defined Lengthy Preclinical Stage Provides New Hope to Circumvent Advancement of Disease- and Age-Related Ne. *Med. Hypotheses* **2015**, *84*, 460−469.
(25) 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.
(26) 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.
(27) 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.
(28) 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.
(29) 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.
(30) 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.
(31) 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.
(32) Becke, A. D. Density-functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.
(33) 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.
(34) Zhao, Y.; Truhlar, D. G. The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent Interactions, Excited States, and Transition Elements: Two New Functionals and Systematic Testing of Four M06-Class Functionals and 12 Other Function. *Theor. Chem. Acc.* **2008**, *120*, 215−241.
(35) 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.
(36) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. Development and Use of Quantum Mechanical Molecular Models. 76. AM1: A New General Purpose Quantum Mechanical Molecular Model. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.
(37) Thiel, W.; Voityuk, A. A. Extension of MNDO to d Orbitals: Parameters and Results for the Second-Row Elements and for the Zinc Group. *J. Phys. Chem.* **1996**, *100*, 616−626.
(38) Stewart, J. J. P. Optimization of Parameters for Semiempirical Methods V: Modification of NDDO Approximations and Application to 70 Elements. *J. Mol. Model.* **2007**, *13*, 1173−1213.
(39) Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-XTB—An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *J. Chem. Theory Comput.* **2019**, *15*, 1652−1671.
(40) Gaus, M.; Cui, Q.; Elstner, M. DFTB3: Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method (SCC-DFTB). *J. Chem. Theory Comput.* **2011**, *7*, 931−948.
(41) Andreas, W. G. An Extensible Interface for QM/ MM Molecular Dynamics Simulations with AMBER. *J. Comput. Chem.* **2014**, 95−108.
(42) 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.
(43) Zinovjev, K. String-Amber https://github.com/kzinovjev/string-amber (accessed Jun 24, 2020).
(44) 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, Jr., J. A.; 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.
(45) Zinovjev, K.; Tuñón, I. Adaptive Finite Temperature String Method in Collective Variables. *J. Phys. Chem. A* **2017**, *121*, 9764−9772.
(46) 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.
(47) Martí, S.; Arafet, K.; Lodola, A.; Mulholland, A. J.; Świderek, K.; Moliner, V. Impact of Warhead Modulations on the Covalent Inhibition of SARS-CoV-2 M pro Explored by QM/MM Simulations. *ACS Catal.* **2022**, *12*, 698−708.
(48) 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.
(49) Smith, J. M.; Jami Alahmadi, Y.; Rowley, C. N. Range-Separated DFT Functionals Are Necessary to Model Thio-Michael Additions. *J. Chem. Theory Comput.* **2013**, *9*, 4860−4865.
(50) Awoonor-Williams, E.; Isley, W. C.; Dale, S. G.; Johnson, E. R.; Yu, H.; Becke, A. D.; Roux, B.; Rowley, C. N. Quantum Chemical Methods for Modeling Covalent Modification of Biological Thiols. *J. Comput. Chem.* **2020**, *41*, 427−438.
(51) 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.

https://doi.org/10.1021/acs.jctc.2c00294
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX