# Computational Study of the Michaelis Complex Formation and the Effect on the Reaction Mechanism of Cruzain Cysteine Protease

Kemel Arafet, Katarzyna Świderek, and Vicent Moliner*

Departament de Química Física i Analítica, Universitat Jaume I, 12071 Castelló, Spain

S *Supporting Information*

ABSTRACT: Cruzain, a cysteine protease of the papain family, is essential in the development of the protozoan *Trypanosoma cruzi*, the etiologic agent of Chagas disease, making it an attractive target for developing new drugs. The present paper is aimed at the study of the catalytic mechanism of the cruzain by first exploring the different protonation states of the active site Cys25 and His159 in the Michaelis complex and the effect on the full catalytic mechanism of this enzyme. The exploration of the equilibrium between these two states has been performed with alchemical free energy perturbation methods with molecular mechanics (MM) force fields and by generating the free energy surfaces in terms of the potential of mean force computed at two levels of theory: AM1d/MM and M06-2X/6-31+G(d,p):AM1d/MM. Alternative mechanisms for the acylation step have been identified on the free energy surfaces and the results suggest the existence of three new reaction mechanisms starting from the peptide binding to the apoenzyme in its neutral Cys25S/His159 dyad state. The mechanism starting with the protonation of the nitrogen atom of the peptide followed by the attack of Cys25S⁻ was revealed as the most favorable one, but it can be competitive with its counterpart mechanism initiated in the Cys25S⁻/His159H⁺ ion pair Michaelis complex. Analysis of energetic and average geometries will allow continuing improvement of our knowledge on this enzyme at the molecular level, which can be crucial to the design of new inhibitors based on the structures of the transition states (transition states analogues) or stable intermediates.
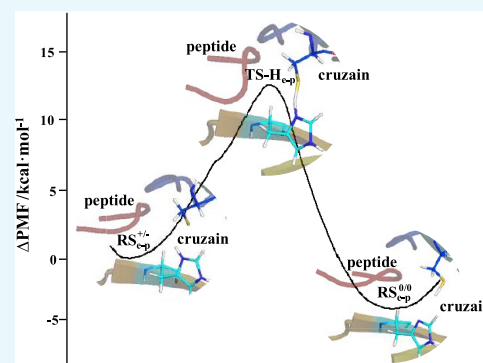
INTRODUCTION

Cysteine proteases of the papain family are essential enzymes to the life cycle of several parasitic protozoa, making them attractive targets for developing new drugs.[1] For instance, rhodesian is expressed by the protozoa *Trypanosoma brucei rhodesiense*, which is responsible for the African sleeping sickness;[2] falcipain-2 is a drug target against Malaria.[3] Cruzain cysteine protease is expressed by *Trypanosoma cruzi*, the etiologic agent of Chagas disease,[1] which concerns 8 millions of people living mainly in endemic Latin American countries.[4] This important class of enzymes are involved in the hydrolysis of peptides and proteins (see Scheme 1) which explains why they are commonly implicated in severe human disease progression.[5] The high catalytic activity of cysteine proteases has been related to the nucleophilic character of their active site, enhanced by the formation of CysS⁻/HisH⁺ ion pair.[6] Nevertheless, a debate is open about the timing of the formation of this reactive complex.

Polgar, in the spectrophotometric determination of pH-dependence of the reaction catalyzed by the cysteine protease papain, concluded that the ion pair formation is responsible for the high reactivity of the cysteine proteases.[7,8] Shafer and co-workers, in different experimental studies on papain, show that the tautomeric relationship between neutral and ionic dyads is such that the thiolate:imidazolium dyad is favored.[9,10] Later Polgar and co-workers, in a kinetic study with the picornain 3C

cysteine protease, proposed that this enzyme does not have a catalytic ion pair.[11] In 2008, Engels and co-workers showed that the ion pair was about 7.2 kcal mol⁻¹ lower in energy than the neutral form based on molecular dynamics (MD) simulations on cathepsin B, a cysteine protease of the papain family.[12] Moreover, according to their calculations, the enzyme environment would stabilize the ion pair by more than 23.8 kcal mol⁻¹ related to the gas phase, and by more than 9.6 kcal mol⁻¹ compared with the ion pair in aqueous solution. However, Albeck and co-workers using the quantum mechanical (QM) methods and a continuum model to mimic the environment, studied the protonation state of the Cys25–His59 (papain numbering) catalytic dyad in free papain and concluded that both residues are protonated at neutral pH.[13] It is nevertheless difficult to envision how the fully protonated form could manage catalysis. Later, the same authors in a study of the inhibition mechanism of papain by peptidyl aldehydes suggested that the ion pair is not important for the catalytic mechanism of the papain family of cysteine proteases.[14] Grazioso and co-workers, based on the generation of potential energy surfaces (PES) of the mechanism for falcipain-2 inhibition by α,β-unsaturated benzo[1,4]diazepin-2-

Scheme 1. Peptide Hydrolysis Catalyzed by Cysteine Proteases
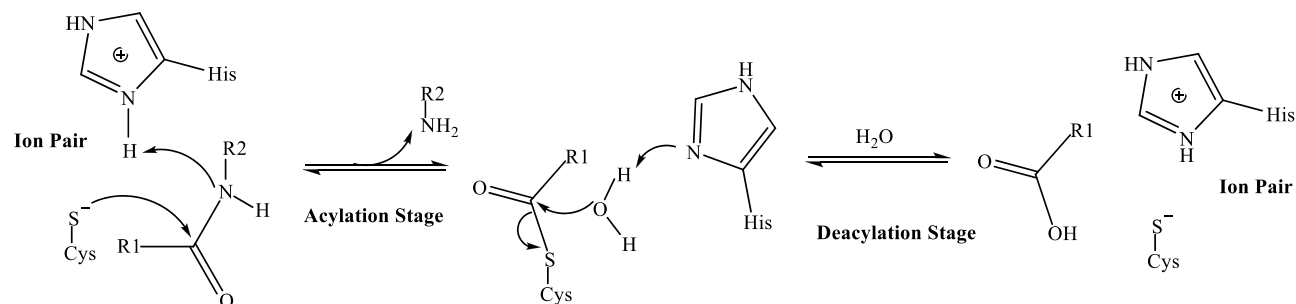
## Scheme 2. Proposed Full Reaction Mechanisms of the Cruzain[a]
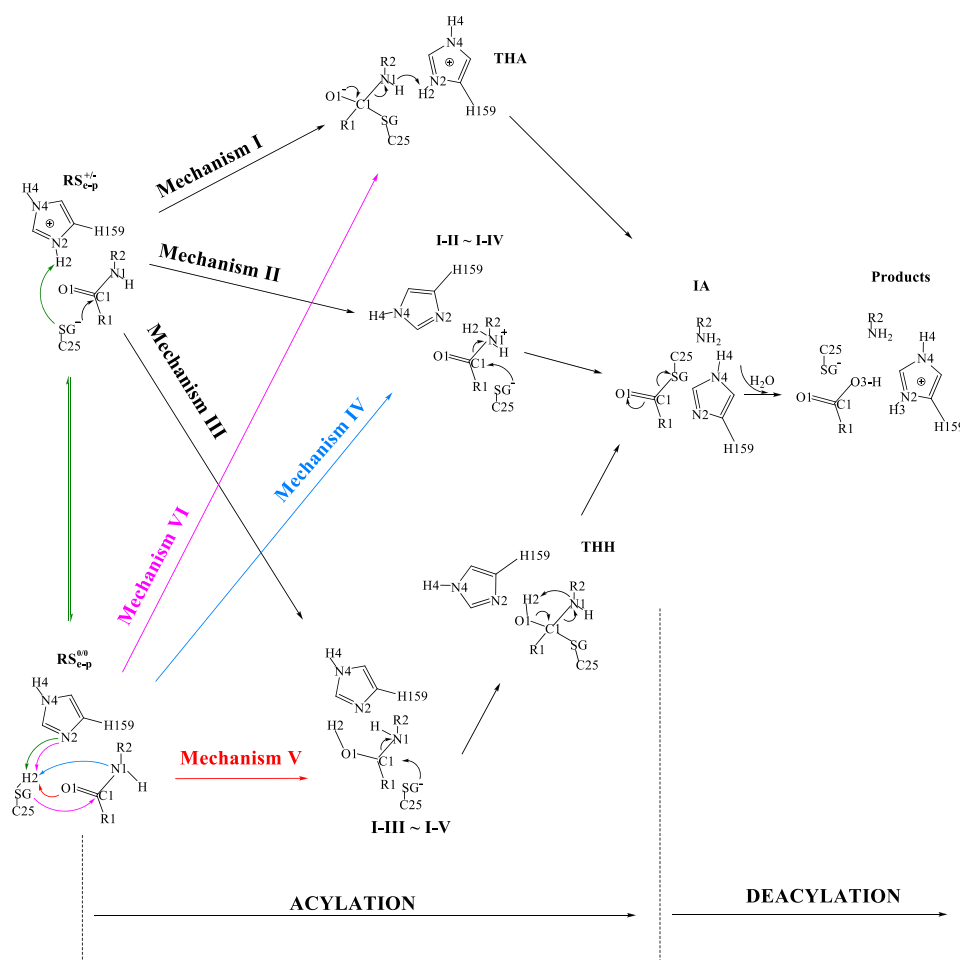
[a]The mechanisms I, II, and III are reported in ref 18.

one methyl ester using hybrid QM/molecular mechanics (MM) methods, proposed that the activation of Cys42 by His174 (falcipain-2 numbering) is not the rate-limiting step and occur without important change in the relative energy of the system.[15] Zhan and co-workers[16] studied the hydrolysis of *N*-acetyl-Phe-Gly-4-nitroanilide catalyzed by papain based on pseudobond first-principles with QM/MM potentials, showing that the ion pair is slightly more stable than the neutral form. More recently, based on the kinetic studies of cruzain, Zhai and Meek proposed that both catalytic cysteine and histidine residues are neutral in the free enzyme and the formation of the ion pair only takes place upon substrate binding.[17]

A QM/MM study of the catalytic mechanism of the cruzain cysteine protease, based on MD simulations using semi-empirical AM1d Hamiltonian and the M06-2X functional to describe the QM sub-set of atoms, was recently carried out in our laboratory.[18] In this case, it was assumed that a substrate was bound to the Cys25S[−]/His59H[+] ion pair $(RS_{e-p}^{+/-})$ as a starting point (see mechanisms I, II, and III in Scheme 2). Our results indicated that the acylation stage takes place in two steps; first, the proton from a histidine residue (His159) is transferred to the N1 atom of the peptide, which is followed by Cys25 attack to the carbonyl carbon atom (mechanism II in Scheme 2). After acylation, deacylation takes place through a
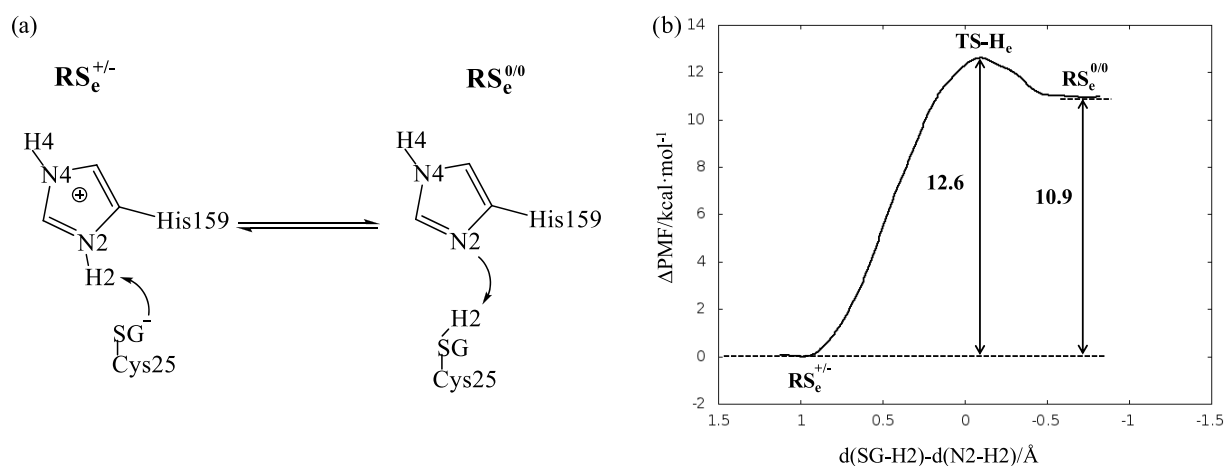
Figure 1. (a) Schematic representation of the equilibrium between the neutral and ion pair dyad in the absence of the peptide in the active site; and (b) the corresponding one-dimensional potential of mean force (1D-PMFs) computed at the M06-2X/6-31+G(d,p):AM1d/MM level.
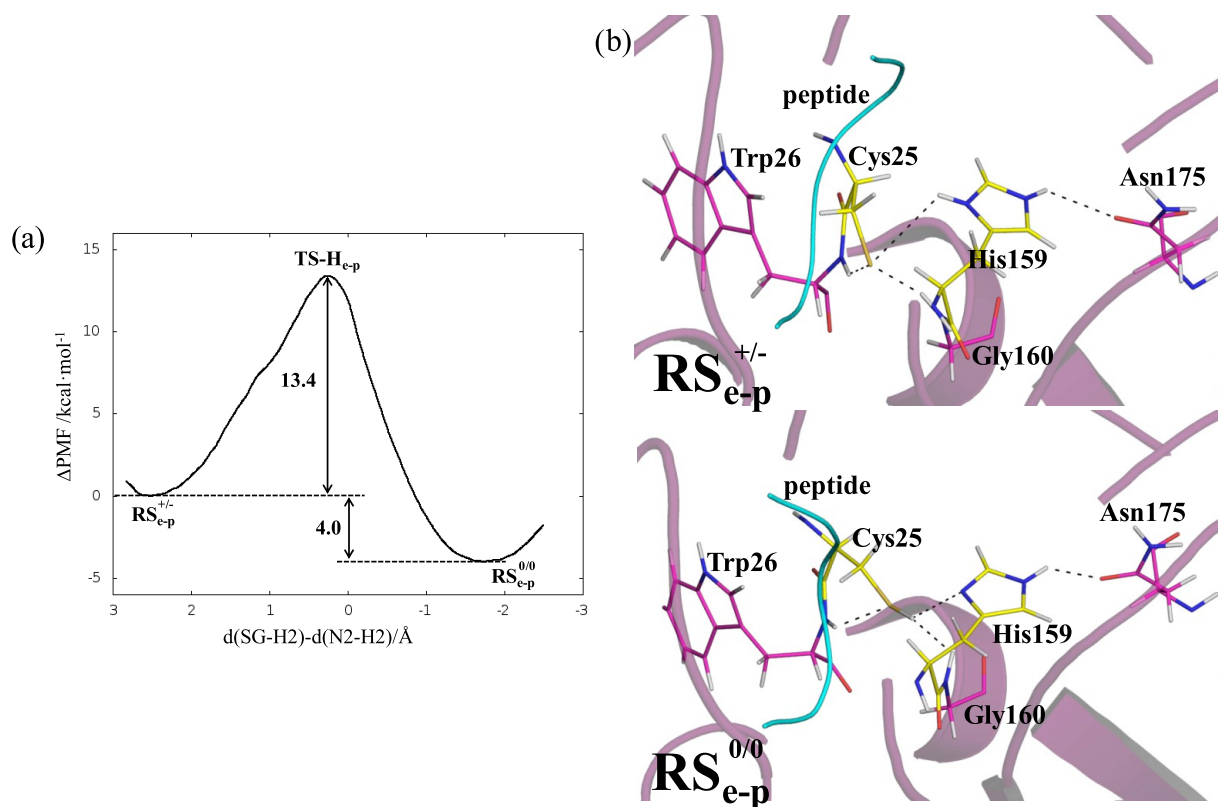
Figure 2. (a) 1D-PMFs for the proton transfer from Cys25 to His159 computed at the M06-2X/6-31+G(d,p):AM1d/MM level in the peptide−enzyme complex model. (b) Snapshots of representative structures of the ion pair and the neutral dyad. Peptide is represented as a ribbon.

concerted mechanism being the rate-limiting step of the full catalytic process.

We herein continue with the study of the catalytic mechanism of the cruzain by exploring the importance of active site protonation states, which is indeed a crucial aspect of enzymology. Our simulations will try to answer the key question of whether the reaction begins with the thiol being thiolate or not. In other words, the equilibrium between the two possible Michaelis complexes, the neutral Cys25/His59 dyad, $RS_{e-p}^{0/0}$, and the $RS_{e-p}^{+/-}$ ion pair, was explored. Then, the first stage of the full catalytic mechanism of this enzyme, the acylation step, was explored from the former. As observed in Scheme 2, the acylation step is herein proposed to take place

through three new possible mechanisms starting from the neutral $RS_{e-p}^{0/0}$ state, mechanisms IV, V, and VI. Mechanism IV implies the direct formation of an intermediate where the neutral Cys25SH transfers the proton to the nitrogen atom of the scissile peptide bond of the substrate (intermediate IV, equivalent to intermediate II). The acylation step through the formation of a protonated thiohemiketal intermediate, THH, is explored in mechanism V, after a preliminary protonation of the carbonyl group of the peptide. The last explored mechanism, mechanism VI, involves the formation of a transient thiohemiketal (THA) intermediate. Free energy surfaces (FESs) corresponding to every possible chemical step will be generated and analyzed. Moreover, in this paper we

are also exploring the relative stability of the ion pair and the neutral forms of the protein−peptide complex which can be crucial to determine the most favourable reaction path. The analysis of energetics and average geometries along the reaction progress allow extracting the conclusions, that closes the present paper.

■ RESULTS AND DISCUSSION

We started our study by exploring the equilibrium between the neutral and ion pair dyad in the absence of the peptide in the active site (see Figure 1a).

As explained in the previous section, the FES for this step was computed as a 1D-PMF using the antisymmetric combination of the distances defining the proton transfer from Cys25 to His159, but also by classical alchemical free energy perturbation (FEP) methods. Figure 1b shows the energy profile corrected at the M06-2X/6-31+G(d,p):AM1d/MM level (the 1D-PMF at the AM1d/MM level is presented in Figure S3 in the Supporting Information). As noticed, the ion pair form of the free enzyme, $RS_e^{+/-}$, is more stable than the neutral form, $RS_e^{0/0}$, with a very low barrier for the formation of the Cys25/His159 ion pair (ca. 2 kcal mol^{-1}). This result is in qualitatively good agreement with classical alchemical calculations that predicts $RS_e^{+/-}$ to be 9.5 kcal mol^{-1}, more stable than $RS_e^{0/0}$. It must be kept in mind that the PMF is computed with QM/MM methods, whereas the full system is treated by MM in the alchemical FEP calculations. In any case, both methods suggest that the peptide would bond to the most populated state of the free protein, i.e., the cavity where the ion pair is already formed. The acylation step would then start from this state, $RS_{e-p}^{+/-}$, as assumed in our previous study of the cruzain catalyzed reaction.[15] Representative snapshots of the ion pair and the dyad is presented in Figure S4 in the Supporting Information.

Nevertheless, the exploration of the free energy surface corresponding to the transfer of the proton from Cys25 to His159, shown in Figure 2a, indicates that when the peptide binds the protein, the stability of Michaelis complex is slightly shifted in favor of $RS_{e-p}^{0/0}$, and thus this step has to be taken into consideration in the exploration of the full reaction mechanism. Analysis of the average structures of the two states (see Figure 2b) reveals that the donor and acceptor inter-atomic distance, SG(C25) and N2(H159), are 3.17 ± 0.15 and 3.97 ± 0.21 Å, in the $RS_e^{0/0}$ and $RS_{e-p}^{0/0}$, respectively (see Table S1 in the Supporting Information for the list of key inter-atomic distances), which means that no large displacements are required in the active site for the transfer of the proton between these two residues in either of the free enzyme or the enzyme−peptide complex. Of note, the free energy barrier from $RS_{e-p}^{+/-}$ to $RS_{e-p}^{0/0}$, 13.4 kcal mol^{-1}, is very close to the value deduced from the rate constants measured by Zhai and Meek that oscillate between 14.0 and 14.8 kcal mol^{-1}, depending on the peptide used in the experiments.[17]

Acylation Stage through Mechanism IV. The next step of our study was the exploration of the acylation stage from the conformation of the active site with Cys25 and His159 as neutral species, $RS_{e-p}^{0/0}$. The results are compared with the results obtained in our previous study,[18] where the starting state of the acylation step catalyzed by cruzain was assumed to be the ion pair (see Scheme 2). The FES for this step, shown in Figure 3 as a two-dimensional PMF (2D-PMF) (see the corresponding 2D-PMF AM1d/MM in Figure S5 of the Supporting Information), describes the acylation stage as a
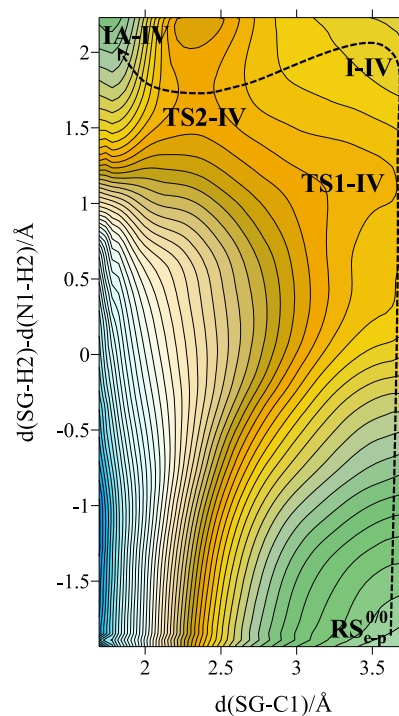

Figure 3. M06-2X/6-31+G(d,p):AM1d/MM FES for the acylation step through mechanism IV. Iso-energetic lines are displayed every 1.5 kcal mol^{-1}.

stepwise mechanism through an intermediate I−IV. As deduced from the FES and reflected in the structures displayed in Figure 4 (a list of key inter-atomic distances of the involved states are reported in Table S2 in the Supporting Information), the proton of Cys25 is transferred directly to the N1 atom of the peptide (1.01 ± 0.02 Å in I−IV), but the bond between thiolate and the C1 atom of the peptide is not formed (SG−C1 distance equal to 3.60 ± 0.03 Å). From this intermediate, the acylation step is completed by the formation of the covalent bond between the sulfur atom of Cys25 and the C1 atom of the peptide (1.73 ± 0.02 Å), concomitant with the breaking of its N1−C1 bond (3.38 ± 0.19 Å).

Figure 5 displays the free energy profile obtained for the acylation step catalyzed by cruzain through mechanism IV and compared with the most favorable reaction path obtained previously in our laboratory through mechanism II starting from the ion pair reactant complex, $RS_{e-p}^{+/-}$.[18] Interestingly, when adding the relative energy between $RS_{e-p}^{0/0}$ and $RS_{e-p}^{+/-}$, presented in Figure 2a, to the acylation step obtained through mechanism IV, it appears that the two mechanisms can be considered as competitive. The highest energy transition state (TS) of mechanism IV (21.3 kcal mol^{-1}) is just 2.1 kcal mol^{-1} lower than the highest energy TS of mechanism II (23.4 kcal mol^{-1}) but, due to the stabilization of $RS_{e-p}^{0/0}$, the later presents a step with the highest energy barrier (23.7 kcal mol^{-1}). Superposition of representative snapshots of the intermediates I and IA obtained from the exploration of mechanisms II and IV are presented in Figure S6 in the Supporting Information.

Acylation Stage through Mechanism V. The exploration of mechanism V from the $RS_{e-p}^{0/0}$ reactant complex has been based on the generation of the FES presented in Figure 6. The 2D-PMF at the AM1d/MM level is deposited in the Supporting Information as Figure S7. As observed in the figure, a protonated thiohemiketal intermediate, THH-V, is
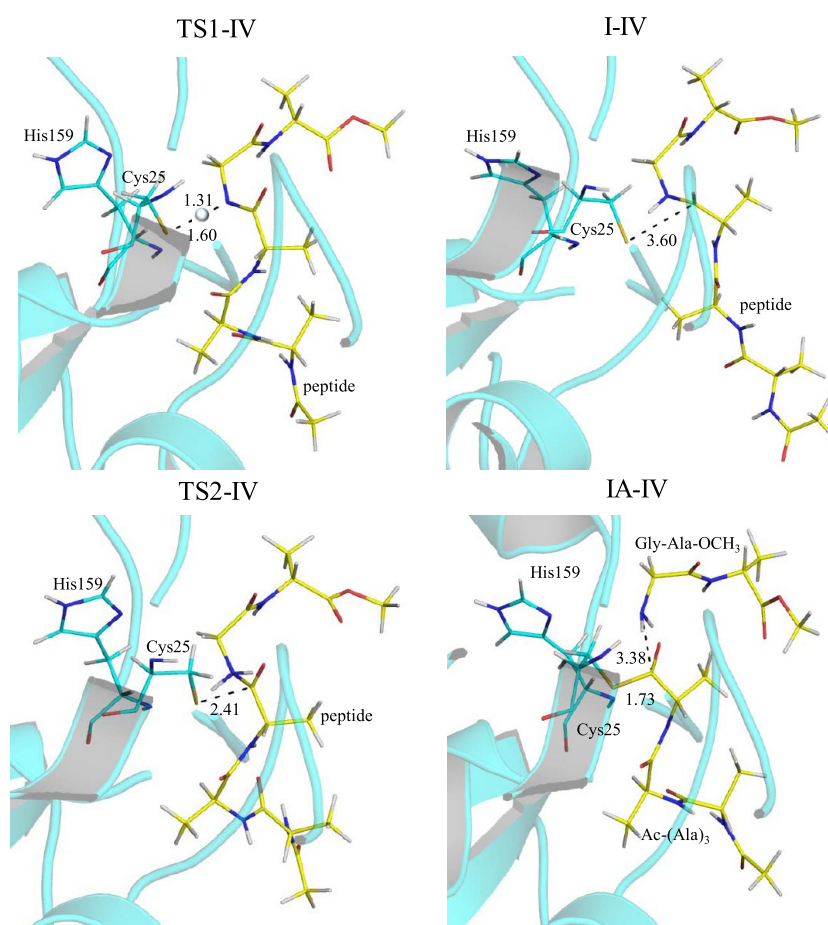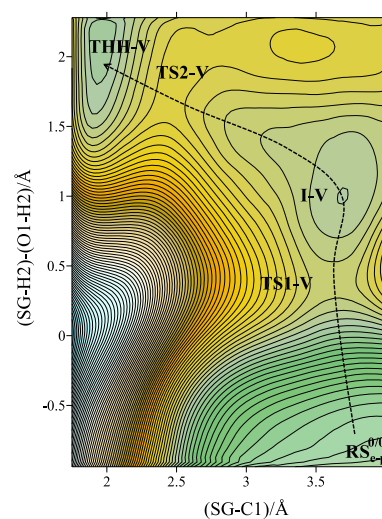
Figure 4. Representative snapshots of the key states located along mechanism IV. Distances are given in Å.
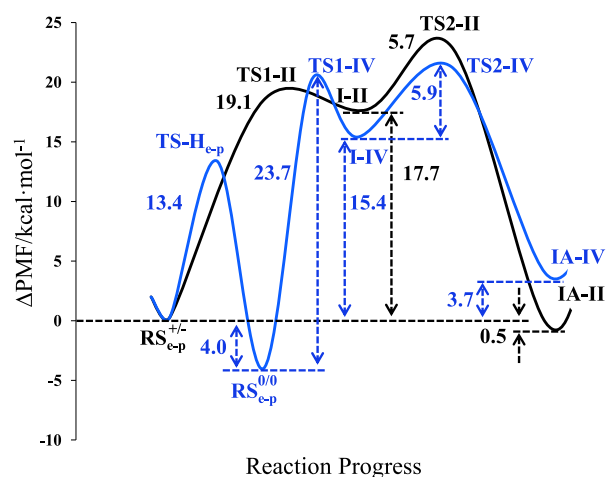

Figure 5. Free energy profile corresponding to the acylation stage through mechanism II obtained in ref 18 (black line) and mechanism IV (blue line). Both profiles derived from FESs computed at the M06-2X/6-31+G(d,p):AM1d/MM level.

Figure 6. M06-2X/6-31+G(d,p):AM1d/MM FES for the formation of the THH-V intermediate. Iso-energetic lines are displayed every 1.5 kcal mol^{-1}.

formed in a stepwise manner through an intermediate, I–V. This stable intermediate is characterized by a deprotonated Cys25 and a protonated carbonyl oxygen O1 atom of the peptide (see Scheme 2, Figure 7 and the list of inter-atomic distances reported in Table S3 of the Supporting Information). From this structure, the active thiolate attacks the C1 atom of the peptide, thus producing the THH-V intermediate. A stable

intermediate before the formation of THH-V was already detected in our previous study when the acylation step of the cruzain reaction was considered to start from the ion pair $RS_{e-p}^{+/-}$.[18] The comparison of the average structures confirms that both intermediates, I–V and I–III, correspond to the same state (see representative snapshots presented in Figure S8 in the Supporting Information). The evolution from THH-

TS1-V

I-V

TS2-V

THH-V
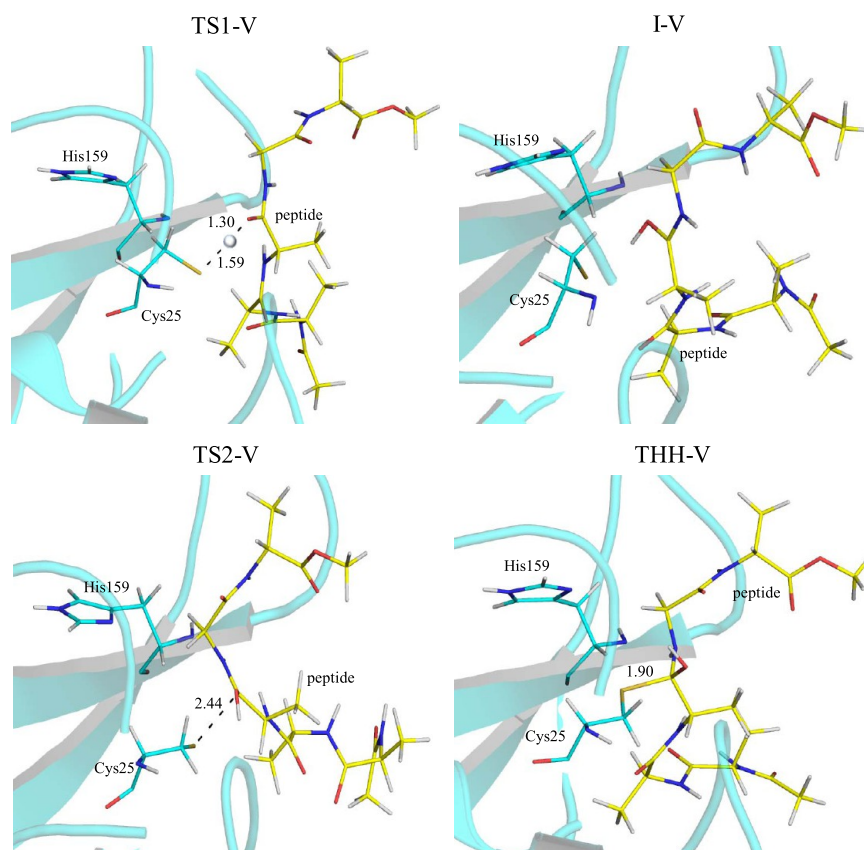
Figure 7. Representative snapshots of the key states located along mechanism V. Distances are given in Å.

V to the final intermediate of the acylation step, IA, rendered a free energy barrier of 32.7 kcal mol^{-1},[18] thus precluding the feasibility of this mechanism.

The resulting free energy profiles from $RS_{e-p}^{+/-}$ to the formation of the THH intermediate through mechanisms III and V, depicted in Figure 8, indicate that the new mechanism V could be competitive with mechanism III. Nevertheless, taking into account the high energy barrier required for the transformation between THH and the final product of the
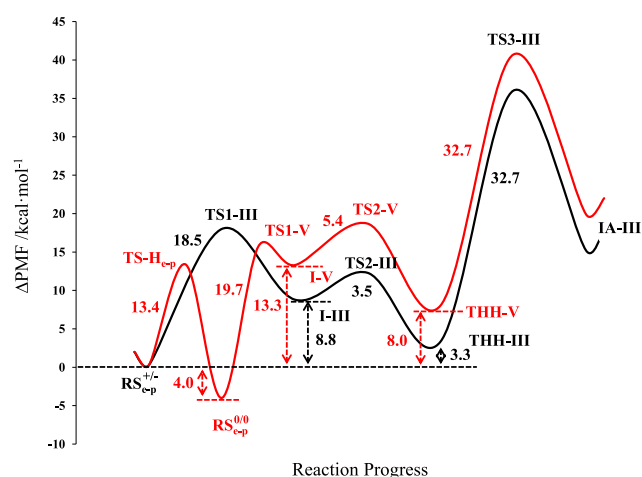

Figure 8. Free energy profile corresponding to the formation of IA intermediate through mechanism III obtained from ref 18 (black line) and mechanism V (red line). Both profiles derived from FESs computed at the M06-2X/6-31+G(d,p):AM1d/MM level.

acylation step of the full reaction catalyzed by cruzain (32.7 kcal mol^{-1}, as reported in our previous study[18]), the present results confirm the reaction mechanism through a THH intermediate as unfeasible.

Acylation Stage through Mechanism VI. Finally, an acylation mechanism through a thiohemiketal intermediate, denoted as THA, was explored from the $RS_{e-p}^{0/0}$ reactant complex. This mechanism was explored in our previous study from the ion pair $RS_{e-p}^{+/-}$ but the FES showed that the formation of THA required a significantly high energy (ca. 48 kcal mol^{-1} higher in energy than the reactant state).[18] The instability of this kind of thiohemiketal tetrahedral intermediate was already observed in our previous study of the inhibition of cruzain.[19] In the present study, the FES, generated with the SG−C1 distance and the antisymmetric combination of distances defined for the hydrogen transfer between Cys25 and His159, $d(SG−H2)−d(N2−H2)$ as reaction coordinates (see Figure S9 in the Supporting Information), corroborates that a THA intermediate does not represent a feasible stable intermediate (the free energy difference between THA and $RS_{e-p}^{0/0}$ was 54.1 kcal mol^{-1} at the M06-2X/6-31+G(d,p):AM1d/MM level). Accordingly, this mechanism was discarded as a competitive reaction path for the acylation step.

■ CONCLUSIONS

Free energy surfaces for the acylation step in cruzain have been obtained in terms of 1D- and 2D-PMFs with hybrid QM/MM potentials. The results, together with the estimation of the relative energy between the two possible conformations of the enzyme active site, the Cys25S^{-}/His159H^{+} ion pair dyad, and the neutral Cys25SH/His159 form, obtained from alchemical
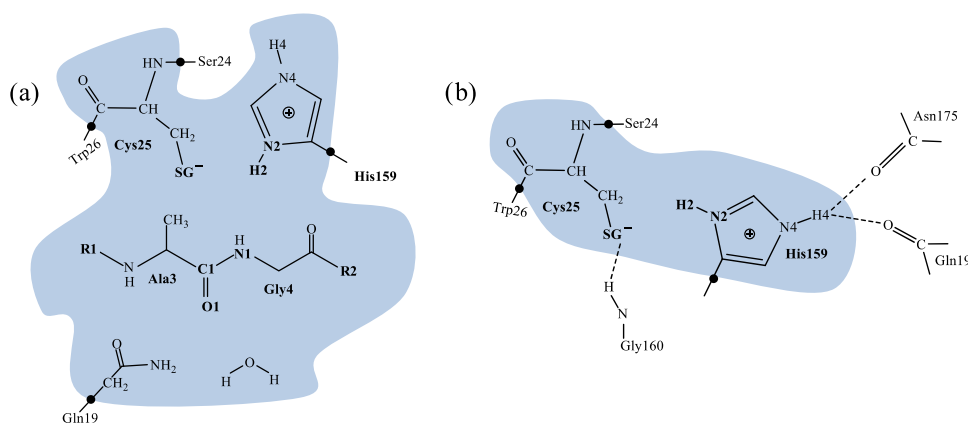
Figure 9. Details of the atoms of the active site treated quantum mechanically (blue regions) in (a) the peptide−protein complex, where R1 correspond to CH$_3$CO−(Ala)$_2$ and R2 correspond to Ala−CH$_3$, and (b) the free protein. Black dots represent the hydrogen link atoms.

FEP methods, have allowed obtaining a picture of the full catalytic process.

The QM/MM free energy surface associated with the transfer of the proton from Cys25H to His159 in the absence of peptide in the cavity of the enzyme, in agreement with the classical FEP calculations, reveals that the equilibrium is displaced towards the more stable ion pair dyad, Cys25$^-$/His159H$^+$, RS$_{e\text{-}p}^{+/-}$. Nevertheless, when the peptide is bound in the active site the differences in energy between the two possible Michaelis complex conformations are reduced, slightly shifted to the stabilization of the neutral Cys25SH/His159 form, RS$_{e\text{-}p}^{0/0}$, by ca. 4 kcal mol^{-1}. The activation-free energy from RS$_{e\text{-}p}^{+/-}$ to RS$_{e\text{-}p}^{0/0}$ is close to the value deduced from the rate constants reported by Zhai and Meek.[17] Consequently, RS$_{e\text{-}p}^{0/0}$ must be considered as a possible Michaelis complex in the evaluation of the possible mechanisms of the acylation stage catalyzed by cruzain. In the present paper we are presenting three mechanisms not explored in our previous study, where the ion pair dyad was considered as the starting point of the acylation stage,[18] by considering the neutral form, RS$_{e\text{-}p}^{0/0}$, as the starting point.

The complete free energy landscape suggests that the most favorable reaction path for the acylation step initiated from the neutral Michaelis complex proceeds through a direct proton transfer from Cys25 to the nitrogen atom of the scissile bond of the peptide (see mechanism IV in Scheme 2). Then, the sulfur atom of the activated thiolate would attack the carbon atom of the peptide, thus provoking the C−N breaking bond. Interestingly, when taking into account the relative energy between the two reactant complexes, the acylation step obtained through this mechanism IV appears to be competitive with the previously explored mechanism II (see Scheme 2). The highest energy TS of mechanism IV is just 2.1 kcal mol^{-1} lower than the highest energy TS of mechanism II but, due to the stabilization of the protein−peptide reactant complex when Cys25 and His159 residues are neutral, the later presents a step with the highest energy barrier. Alternative mechanisms involving the formation of a protonated or unprotonated tetrahedral thiohemiketal intermediates have been explored but they must be discarded according to the associated high energy barriers.

In all, the present results suggest that the reaction can also take place through the formation of a protein−peptide complex with the catalytic cysteine and histidine in their neutral forms. The fact that the newly explored mechanism shows activation-free energy barriers similar to previously published ones that started from the ion pair Michaelis complex opens new frontiers to future developments in designing new drugs (mimicking for instance new identified TSs) for Chagas disease.

■ COMPUTATIONAL METHODS

Our first molecular model corresponding to the peptide−protein complex consists of one chain of 215 amino acids (cruzain) and the peptide Ac−Ala−Ala−Ala−Gly−Ala−OCH$_3$. This model was constructed from the X-ray crystal structure of cruzain from *T. cruzi* with PDB code 1AIM[20] and 2.0 Å resolution, that contains Bz−Tyr−Ala−CH$_2$F in the active site. The covalent inhibitor Bz−Tyr−Ala−CH$_2$F was replaced by the peptide Ac−Ala−Ala−Ala−Gly−Ala−OCH$_3$. Taking into account the low substrate specificity of this protein,[21−23] this peptide was chosen because it was already used in our previous study of the catalytic mechanism from the ion pair dyad, thus facilitating the comparison between mechanisms.[18] In addition, it presents computational advantages due to its small size. The selected bond cleavage was the Ala3−Gly4 peptide bond (see Figure 9a). The cruzain cysteine protease specificity for the residue−residue peptide bond has been recently investigated by Zhai and Meek.[17] Their experimental kinetic studies suggest that a larger charge residue at the P1 position is preferred for catalysis, whereas a hydrophobic residue is more favorable than a charged one at the P2 position. The missing hydrogen atoms of the X-ray structure were introduced at pH 7 using the fDYNAMO library,[24] within the pK_a values of the titratable residues calculated within the empirical PROPKA 3.1 program.[25] The protonation state of His159 and Cys25 was studied in detail by exploring the free energy surfaces, as detailed below. Eleven counterions (Na$^+$) were placed into optimal electrostatic positions (those where the potential reaches maximum negative values) around the enzyme (further than 10.5 Å from any atom of the system and 5 Å from any other counterion, using a regular grid of 0.5 Å), to obtain the electroneutrality of the system. Once the system was complete, it was solvated by a 79.5 Å side cubic box of water molecules.

The QM region, first described with the AM1d semi-empirical Hamiltonian,[26] contains the full peptide, residue Cys25, the imidazole ring of His159, the side chain of Gln19, and one water molecule (88 atoms, see Figure 9a). This is the Hamiltonian already employed in the exploration of mecha-

nisms I, II, and III reported in our previous study.[18] Recently, Lameira and co-workers[27] in a study combining experiments and computational modeling, employed the same Hamiltonian to compute the free energy surface of the inhibition of cruzain cysteine protease by dipeptidyl nitriles, with results in accordance with the available experimental data. The rest of the system, protein and water molecules, was described by the OPLS-AA[28] and TIP3P[29] force fields, respectively. Hydrogen link atoms were used to saturate the valence of the QM−MM frontier bonds.[30] All residues further than 25 Å from the C1 atom of the peptide were kept frozen during the simulations (43 115 atoms from a total of 50 029). A force switching function with a cut-off distance in the range of 14.5−16 Å and periodic boundary conditions were employed and used to treat the nonbonding interactions. The fDYNAMO library[24] was used for the QM/MM calculations. The reactants, products, intermediates, and transition state structures were optimized by means of a micro−macro iteration scheme.[31]

Once the PESs of every chemical step were explored, free energy surfaces (FESs), in terms of the potential of mean force (PMF), were generated using umbrella sampling and the weighted histogram analysis method to recover the probabilities.[32,33] The error associated with this method, when properly carried out, is usually accepted to be around 1 kcal mol^{-1}.[34] The harmonic umbrella sampling force constants was 2500 kJ mol^{-1} Å^{-2}. A plot showing the proper overlapping of the sampling in consecutive windows is deposited in the Supporting Information (Figure S1). 20 ps of equilibration and 40 ps of production, with a time step of 1 fs, were used in every window of the PMFs. To obtain averaged geometrical parameters, 200 ps of AM1d/MM MD simulations of the windows corresponding to the different states were performed.

Because no X-ray structure has been solved for the free protein, a second molecular model was prepared in which the peptide was removed from the active site of the first model. Twenty nanoseconds of classical MD simulation were required to equilibrate the new system employing NAMD parallel molecular dynamics code.[35] The structure of the apoenzyme was confirmed to be equilibrated by the analysis of the time evolution of the root-mean-square-deviation (RMSD) (see Figure S2). The MD simulations were carried out at 300 K using NVT ensemble and the Langevin−Verlet integrator using a time step of 1 fs. The protein atoms were treated with the CHARMM[36] force field and the water molecules were treated with the TIP3P[37] force fields; both as implemented in NAMD.[35] For further QM/MM calculations in this new model, fDYNAMO library[24] was employed with the QM region containing just the residue Cys25 and the imidazole ring of His159 (19 atoms, as depicted in Figure 9b). The equilibrium between the neutral and the ion pair dyad for the free enzyme (model 2, Figure 9b) was explored by generating a monodimensional AM1d/MM PMFs (1D-PMF) with the antisymmetric combination of the bond-forming and bond-breaking distances, $d(SG−H2)−d(N2−H2)$ as the reaction coordinate. In addition, calculations based on classical alchemical free energy perturbation methods (FEP) were also carried out to study this transformation of the active site without the peptide. NAMD was employed for carrying out these calculations (see the Supporting Information for details).

Model 1 (Figure 9a) was employed to study the equilibrium between the neutral and the ion pair dyad in the peptide−protein complex, as well as for the exploration of the acylation step, by generating the free energy surfaces of every single chemical step. Similar to the study in the apoenzyme, a monodimensional AM1d/MM PMFs (1D-PMF) was generated with the antisymmetric combination of the bond-forming and bond-breaking distances, $d(SG−H2)−d(N2−H2)$ as the reaction coordinate. In the case of mechanism IV (starting from the $RS_{e-p}^{0/0}$ complex as depicted in Scheme 2), a two-dimensional PMF (2D-PMF) was computed at the AM1d/MM level using as ζ_1 the bond-forming distance between Cys and carbon of scissile peptide bond, $d(SG−C1)$, and as ζ_2 the antisymmetric combination of distances defined for the hydrogen transfer between Cys25 and nitrogen of the same peptide bond, $d(SG−H2)−d(N1−H2)$. A total of 2187 simulation windows were used to generate this free energy surface. For the exploration of mechanism V (starting from the $RS_{e-p}^{0/0}$ complex as depicted in Scheme 2), a 2D-PMF was computed where the SG−C1 distance was defined as ζ_1, and the antisymmetric combination of two distances defined for the hydrogen transfer, ζ_2 but this time from Cys25 to oxygen of the carbonyl group, $d(SG−H2)−d(O1−H2)$. A total of 2989 simulation windows were required to generate this 2D-PMF. Finally, for the exploration of mechanism VI (starting from the $RS_{e-p}^{0/0}$ complex as depicted in Scheme 2), a 2D-PMF was computed where the SG−C1 distance was defined as ζ_1, and as ζ_2 the antisymmetric combination of distances defined for the hydrogen transfer between Cys25 and His159, $d(SG−H2)−d(N2−H2)$. This required a series of 1519 simulation windows.

The QM sub-set of atoms in the QM/MM FESs were restricted to low-level because of the computer limitations. To improve the level of theory, an interpolated correction scheme with a higher level Hamiltonian, developed in our laboratory,[38] was applied as explained in detail in our previous paper.[18] The M06-2X functional[39] with the standard 6-31+G(d,p) basis set,[40] following Truhlar and co-workers suggestions,[39,41] was used employing the Gaussian 09 program,[42] combined with the fDYNAMO library.[24]

■ ASSOCIATED CONTENT

ⓈSupporting Information

Computational details for the alchemical FEP calculations; overlapping of the sampling in consecutive windows during the PMF AM1d/MM; RMSD computed along the classical MD simulation for the backbone atoms of the protein; FESs computed at the AM1d/MM level; key average inter-atomic distances for the states located along the equilibrium between the ion pair and the neutral Cys25/His159 dyad in the absence and presence of the peptide; representative snapshots of the different states located along the different explored steps (PDF)

■ AUTHOR INFORMATION

Corresponding Author
*E-mail: moliner@uji.es. Tel: +34964728084.

ORCID ⓘ
Katarzyna Świderek: 0000-0002-7528-1551
Vicent Moliner: 0000-0002-3665-3391

Notes
The authors declare no competing financial interest.

REFERENCES

(1) McKerrow, J. H.; Engel, J. C.; Caffrey, C. R. Cysteine protease inhibitors as chemotherapy for parasitic infections. *Bioorg. Med. Chem.* **1999**, *7*, 639−644.

(2) Barrett, M. P.; Burchmore, R. J. S.; Stich, A.; Lazzari, J. O.; Frasch, A. C.; Cazzulo, J. J.; Krishna, S. The trypanosomiases. *Lancet* **2003**, *362*, 1469−1480.

(3) Renslo, A. R.; McKerrow, J. H. Drug discovery and development for neglected parasitic diseases. *Nat. Chem. Biol.* **2006**, *2*, 701−710.

(4) http://www.who.int/chagas/disease/en/,2018.

(5) Scott, C. J.; Taggart, C. C. Biologic protease inhibitors as novel therapeutic agents. *Biochimie* **2010**, *92*, 1681−1688.

(6) Brocklehurst, K. Specific covalent modification of thiols-applications in the study of enzymes and other biomolecules. *Int. J. Biochem.* **1979**, *10*, 259−274.

(7) Polgár, L. Mode of activation of catalytically essential sulfhydryl group of papain. *Eur. J. Biochem.* **1973**, *33*, 104−109.

(8) Polgár, L. Mercaptide-imidazolium ion-pair-reactive nucleophile in papain catalysis. *FEBS Lett.* **1974**, *47*, 15−18.

(9) Lewis, S. D.; Johnson, F. A.; Shafer, J. A. Potentiometric determination of ionizations at active-site of papain. *Biochemistry* **1976**, *15*, 5009−5017.

(10) Lewis, S. D.; Johnson, F. A.; Shafer, J. A. Effect of Cysteine-25 on the ionization of Histidine-159 in papain as determined by proton nuclear magnetic-resonance spectroscopy - Evidence for a His-159-Cys-25 ion-pair and its possible role in catalysis. *Biochemistry* **1981**, *20*, 48−51.

(11) Sárkány, Z.; Szeltner, Z.; Polgar, L. Thiolate-imidazolium ion pair is not an obligatory catalytic entity of cysteine peptidases: The active site of picornain 3C. *Biochemistry* **2001**, *40*, 10601−10606.

(12) Mladenovic, M.; Junold, K.; Fink, R. F.; Thiel, W.; Schirmeister, T.; Engels, B. Atomistic insights into the inhibition of cysteine proteases: First QM/MM calculations clarifying the regiospecificity and the inhibition potency of epoxide- and aziridine-based inhibitors. *J. Phys. Chem. B* **2008**, *112*, 5458−5469.

(13) Shokhen, M.; Khazanov, N.; Albeck, A. Challenging a paradigm: Theoretical calculations of the protonation state of the Cys25-His159 catalytic diad in free papain. *Proteins: Struct., Funct., Bioinf.* **2009**, *77*, 916−926.

(14) Shokhen, M.; Khazanov, N.; Albeck, A. The mechanism of papain inhibition by peptidyl aldehydes. *Proteins: Struct., Funct., Bioinf.* **2011**, *79*, 975−985.

(15) Grazioso, G.; Legnani, L.; Toma, L.; Ettari, R.; Micale, N.; De Micheli, C. Mechanism of falcipain-2 inhibition by alpha,beta-unsaturated benzo 1,4 diazepin-2-one methyl ester. *J. Comput.-Aided Mol. Des.* **2012**, *26*, 1035−1043.

(16) Wei, D. H.; Huang, X. Q.; Liu, J. J.; Tang, M. S.; Zhan, C. G. Reaction Pathway and Free Energy Profile for Papain-Catalyzed Hydrolysis of N-Acetyl-Phe-Gly 4-Nitroanilide. *Biochemistry* **2013**, *52*, 5145−5154.

(17) Zhai, X.; Meek, T. D. Catalytic Mechanism of Cruzain from *Trypanosoma cruzi* As Determined from Solvent Kinetic Isotope Effects of Steady-State and Pre-Steady-State Kinetics. *Biochemistry* **2018**, *57*, 3176−3190.

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

(19) Arafet, K.; Ferrer, S.; Moliner, V. First Quantum Mechanics/Molecular Mechanics Studies of the Inhibition Mechanism of Cruzain by Peptidyl Halomethyl Ketones. *Biochemistry* **2015**, *54*, 3381−3391.

(20) Gillmor, S. A.; Craik, C. S.; Fletterick, R. J. Structural determinants of specificity in the cysteine protease cruzain. *Protein Sci.* **1997**, *6*, 1603−1611.

(21) Judice, W. A. S.; Cezari, M. H. S.; Lima, A. P. C. A.; Scharfstein, J.; Chagas, J. R.; Tersariol, I. L. S.; Juliano, M. A.; Juliano, L. Comparison of the specificity, stability and individual rate constants with respective activation parameters for the peptidase activity of cruzipain and its recombinant form, cruzain, from *Trypanosoma cruzi*. *Eur. J. Biochem.* **2001**, *268*, 6578−6586.

(22) Salvati, L.; Mattu, M.; Policelli, F.; Tiberi, F.; Gradoni, L.; Venturini, G.; Bolognesi, M.; Ascenzi, P. Modulation of the catalytic activity of cruzipain, the major cysteine proteinase from *Trypanosoma cruzi*, by temperature and pH. *Eur. J. Biochem.* **2001**, *268*, 3253−3258.

(23) Castro, H. C.; Abreu, P. A.; Geraldo, R. B.; Martins, R. C. A.; dos Santos, R.; Loureiro, N. I. V.; Cabral, L. M.; Rodrigues, C. R. Looking at the proteases from a simple perspective. *J. Mol. Recognit.* **2011**, *24*, 165−181.

(24) Field, M. J.; Albe, M.; Bret, C.; Proust-De Martin, F.; Thomas, A. The Dynamo library for molecular simulations using hybrid quantum mechanical and molecular mechanical potentials. *J. Comput. Chem.* **2000**, *21*, 1088−1100.

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

(26) Nam, K.; Cui, Q.; Gao, J.; York, D. M. Specific reaction parametrization of the AM1/d Hamiltonian for phosphoryl transfer reactions: H, O, and P atoms. *J. Chem. Theory Comput.* **2007**, *3*, 486−504.

(27) Dos Santos, A. M.; Cianni, L.; De Vita, D.; Rosini, F.; Leitão, A.; Laughton, C. A.; Lameira, J.; Montanari, C. A. Experimental study and computational modelling of cruzain cysteine protease inhibition by dipeptidyl nitriles. *Phys. Chem. Chem. Phys.* **2018**, *20*, 24317−24328.

(28) Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.

(29) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.

(30) Field, M. J.; Bash, P. A.; Karplus, M. A combined quantum-mechanical and molecular mechanical potential for molecular-dynamics simulations. *J. Comput. Chem.* **1990**, *11*, 700−733.

(31) Martí, S.; Moliner, V.; Tuñón, I. Improving the QM/MM description of chemical processes: A dual level strategy to explore the potential energy surface in very large systems. *J. Chem. Theory Comput.* **2005**, *1*, 1008−1016.

(32) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. The weighted histogram analysis method for free-energy calculations on biomolecules. 1. The method. *J. Comput. Chem.* **1992**, *13*, 1011−1021.

(33) Torrie, G. M.; Valleau, J. P. Non-physical sampling distributions in monte-carlo free-energy estimation - Umbrella Sampling. *J. Comput. Phys.* **1977**, *23*, 187−199.

(34) Field, M. J. *A Practical Introduction to the Simulation of Molecular Systems*, 2nd ed.; Cambrige University Press: Cambrige, U.K., 2007.

(35) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781−1802.

(36) MacKerell, A. D.; Banavali, N.; Foloppe, N. Development and current status of the CHARMM force field for nucleic acids. *Biopolymers* **2000**, *56*, 257−265.

(37) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.

(38) Ruiz-Pernia, J. J.; Silla, E.; Tuñón, I.; Martí, S.; Moliner, V. Hybrid QM/MM potentials of mean force with interpolated corrections. *J. Phys. Chem. B* **2004**, *108*, 8427−8433.

(39) Zhao, Y.; Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, non-covalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **2008**, *120*, 215−241.

(40) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. Ab Initio Molecular Orbital Theory. John Wiley and Sons: New York, 1986.

(41) Lynch, B. J.; Zhao, Y.; Truhlar, D. G. Effectiveness of Diffuse Basis Functions for Calculating Relative Energies by Density Functional Theory. *J. Phys. Chem. A* **2003**, *107*, 1384−1388.

(42) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, revision A.1, 2009.