

Peptide Hydrolysis in Thermolysin: Ab Initio QM/MM Investigation of the Glu143-Assisted Water Addition Mechanism

Jochen Blumberger,^{*,†} Guillaume Lamoureux,^{*,‡} and Michael L. Klein

Center for Molecular Modeling and Department of Chemistry, University of Pennsylvania, 231 S. 34th Street, Philadelphia, Pennsylvania 19104-6323

Received March 31, 2007

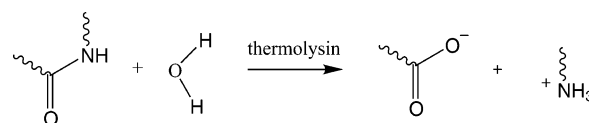
Abstract: Thermolysin (TLN) is one of the best-studied zinc metalloproteases. Yet the mechanism of action is still under debate. In order to investigate the energetic feasibility of the currently most favored mechanism, we have docked a tripeptide to the active site of TLN and computed the free energy profile at the quantum mechanics/molecular mechanics level of theory. The mechanism consists of three distinct steps: (i) a Zn-bound water molecule is deprotonated by Glu143 and attacks the carbonyl bond of the substrate; (ii) Glu143 transfers the proton to the amide nitrogen atom; (iii) the nitrogen atom is protonated and the peptide bond is irreversibly broken. The free energy barriers for steps i and iii have almost equal heights, 14.8 and 14.7 kcal/mol, respectively, and are in good agreement with the effective experimental activation barrier obtained for similar substrates, 12.1–13.6 kcal/mol. Transition state stabilization for nucleophilic attack is achieved by formation of a weak coordination bond between the substrate carbonyl oxygen atom and the Zn ion and of three strong hydrogen bonds between the substrate and protonated His231 and two solvent molecules. The transition state for the nucleophilic attack (step i) is more tightly bonded than the enzyme–substrate complex, implying that TLN complies with Pauling's hypothesis regarding transition-state stabilization. Glu143, at first unfavorably oriented for protonation of the amide nitrogen atom, displayed large structural fluctuations that facilitated reorganization of the local hydrogen-bond network and transport of the proton to the leaving group on the nanosecond time scale. The present simulations give further evidence that Glu143 is a highly effective proton shuttle which should be assigned a key role in any reaction mechanism proposed for TLN.

1. Introduction

Thermolysin (TLN) is an extracellular zinc endoprotease of bacterial origin that catalyzes peptide-bond hydrolysis specif-

ically on the N-terminus side of large hydrophobic residues (Scheme 1).^{1,2}

Scheme 1



The mechanism of action of TLN is assumed to be similar for all families of the thermolysin clan including the important peptidases of higher organisms such as carboxypeptidase A, angiotensin converting enzyme, enkephalinase, collagenase, and neprilysin. The latter, expressed in the

* Corresponding authors. Phone: ++44-(0)1223-763872 (J.B.); 514-848-2424 ext. 5314 (G.L.). fax: ++44-(0)1223-336362 (J.B.); 514-848-2868 (G.L.). e-mail: jb376@cam.ac.uk (J.B.); glamoure@alcor.concordia.ca (G.L.).

† Present address: Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, U.K.

‡ Present address: Department of Chemistry and Biochemistry, Concordia University, 7141 Sherbrooke Street West, Montréal H4B 1R6, Canada.

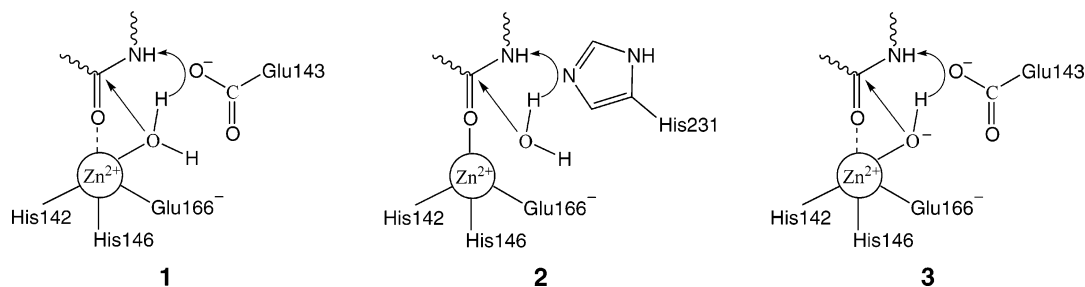


Figure 1. Three mechanisms for peptide bond hydrolysis in thermolysin: (1) Glu143-assisted addition of a water molecule, (2) His231-assisted addition of a water molecule, and (3) Glu143-assisted addition of a hydroxide ion.

kidneys and the central nervous system, is involved in the catabolism of active peptides. Neprilysin inhibitors have attracted great interest as they show antinociceptive and antihypertensive properties.³ Due to its high thermostability, tolerance to organic solvents, and specificity against hydrophobic residues, TLN is also used in industrial processes, for example, as a catalyst for the hydrolysis and synthesis of a precursor of aspartame and in the continuous synthesis of peptide derivatives (see refs in ref 4).

The first crystal structure of TLN was solved in 1972 by Matthews and co-workers,⁵ and a number of TLN⁶ and TLN-inhibitor structures^{7–9} have been reported since then. The Zn²⁺ ion is coordinated with two histidine ligands, His142 and His146; one glutamate residue, Glu166; and one water molecule (see Figure 1). Glu166 can bind monodentate⁵ or bidentate,⁶ and Zn is accordingly either 4- or 5-fold coordinated. Catalytically active second shell residues include Glu143,^{10–13} His231,^{10,14} Asp226, and Tyr157.¹⁵ The histidine and glutamate residues form the “HEXXH + E” zinc-binding motif that is conserved within the thermolysin clan. Analyses of TLN-inhibitor complexes have also given information about the residues involved in substrate recognition and stabilization of the enzyme substrate complex (ES).¹⁶

Despite the wealth of crystallographic^{5–9} and kinetic^{10–15,17–19} data, the reaction mechanism for thermolysin-catalyzed peptide bond cleavage is still under debate.^{1,2} In the “hydroxide” mechanism (mechanism 1 in Figure 1), the reactive nucleophile is generated by deprotonation of the Zn-bound water molecule. Glu143 accepts the proton and transfers it to the amide nitrogen. This mechanism is supported by many crystal structures of enzyme–inhibitor complexes,⁹ and most convincingly by the fact that Glu143 mutation causes an almost total loss of enzymatic activity in neutral proteases^{10,11} and related Zn endopeptidases.^{12,13} The “hydroxide” mechanism was challenged by Mock et al.^{17,18} who found that the logarithm of the catalytic rate constant, $\log k_{\text{cat}}$, increases linearly with increasing pH in the range pH = 5–8 and saturates after a pH of 8.26 is reached. This behavior is indicative of an acidic species with $\text{p}K_{\text{a}} = 8.26$ that actively participates in the reaction in its deprotonated form for the entire range of pH values. Since the “hydroxide” mechanism cannot convincingly account for this observation, Mock suggested that neutral His231 is this particular residue, which in place of Glu143 acts as general base and deprotonates an unbound water molecule (mechanism 2 in Figure 1). This proposal is rather controversial because it cannot explain the dramatic loss of activity of

Glu143 mutants. Thereafter, Lipscomb and Sträter¹ suggested that Mock’s kinetic data can be accounted for by a slight modification of the original “hydroxide” mechanism: not His231 but the Zn-bound water molecule was assigned a $\text{p}K_{\text{a}} = 8.26$, implying that the reactive oxygen moiety in the ES complex is hydroxide rather than water (mechanism 3 in Figure 1).

As implied by its name, thermolysin is resistant to high temperatures. The thermostability is due in part to four Ca²⁺ ions in the interior of the protein that prevent large conformational fluctuations.²⁰ Dynamical effects due to protein motion being intensively discussed in the current literature^{21–23} are expected to be small for thermolysin. This suggests that computer simulations, which are capable of probing the enzyme dynamics on the pico- and nanosecond time scale, can be helpful in supporting or excluding one of the mechanistic interpretations drawn from experiments.

The computation of accurate free energy profiles for enzymatic reactions still represents a major challenge. The less time-consuming quantum mechanics/molecular mechanics (QM/MM) approaches that use standard semiempirical methods for the QM part are not expected to be accurate enough to distinguish between the three mechanistic cases. Indeed, the barrier for the Glu143-assisted peptide bond cleavage (mechanism 1 in Figure 1) was estimated to be more than 40 kcal/mol at the AM1/AMBER level of theory,¹⁶ which is more than twice as large as experimental estimates. Gas-phase modeling of enzymatic reactions with fairly accurate density functionals but rather poor representation of the enzymatic environment is at the other extreme. In the gas-phase model of Pelmenschikov et al.,²⁴ the substrate, all first-shell residues, and certain second-shell residues were treated at the B3LYP level of theory and the enzymatic environment was replaced by a continuum model. The free energy barrier obtained for mechanism 1 was in remarkably good agreement with experimental results, and the measured mutation effects could be reasonably well reproduced. However, the gas-phase modeling did not yield a stable tetrahedral intermediate. Instead, a one-step reaction with a single barrier was reported, which is rather unusual for peptide-bond hydrolysis reactions.

The approach we choose in this work combines the advantages of explicit density functional calculations and full atomistic representation of the enzymatic environment at a finite temperature. Starting from the crystal structure of the apoenzyme, a tripeptide for which kinetic data have been measured is docked to the active site. After selection of a

low-energy docking structure, the energetic feasibility of mechanism 1 is investigated by computing the full free energy profile for the Glu143-assisted peptide-bond cleavage. We find that in the ES complex the substrate is bound to the enzyme merely through hydrogen bonds, but it is not coordinated to Zn. As the reaction proceeds, a tight bond between the substrate and Zn is formed, indicating that the metal ion plays an essential role in transition state stabilization. The free energy profile is computed for three distinct steps: (i) nucleophilic attack, (ii) transfer of the proton to the leaving amide group, and (iii) protonation of the amide nitrogen and break of the peptide bond. The barrier height for step i is in good agreement with experimental results and overestimated by 1.2–2.7 kcal/mol depending on the substrate used in experiments. Our computations give evidence that mechanism 1 is indeed energetically feasible and one possible mechanism of action of thermolysin.

This paper is organized as follows. In section 2, we describe the computational methods used for substrate docking, gas-phase calculations, and classical and QM/MM simulations. In section 3, the energy profile for a gas-phase model of the enzymatic reaction is presented. The BLYP density functional and pseudopotentials used for the full enzymatic reaction are validated by comparison to B3LYP and all-electron calculations. The results from the peptide docking procedure are analyzed, and the choice of a specific structure for the ES complex is explained. Using constrained QM/MM simulation, the free energy profiles for formation of the tetrahedral intermediate (reaction i) and for protonation of the leaving group followed by peptide bond cleavage (reaction iii) are computed. The thermodynamics of the rearrangement of the tetrahedral intermediate (reaction ii) are characterized using biased classical simulation. The paper is concluded in section 4.

2. Computational Methods

Gas-Phase Calculations. Gas-phase calculations were carried out for a minimal model of the active site. The zinc ion is ligated by one water molecule, two imidazoles, and one formate ion, the last three ligands replacing residues His142, His146, and Glu166 of the enzyme. The conserved residue Glu143 is modeled by formate and the substrate by *N*-methylacetamide (NMA). For calculation of the energy profile shown in Figure 2, the distance between Zn and the carbonyl oxygen atom of NMA (ZnO_s) was fixed while all remaining atomic positions were optimized until the default convergence criteria were reached. Geometries for the ES complex, tetrahedral intermediate (TI), and product (P) were obtained from unconstrained optimizations. The structure for transition state 1 (TS1) was approximated from geometry optimization with the distance between the oxygen atom of water and the carbonyl carbon atom of NMA (C_sO_w) fixed to a value that closely corresponds to the maximum of the potential of mean force (PMF) in the enzymatic system, 1.85 Å. Similarly, the structure of transition state 2 (TS2) was obtained by optimization with the distance between the transferred hydrogen atom and the nitrogen atom (H_1N_s) fixed to 1.30 Å. Calculations with the CPMD code²⁵ were carried out using the BLYP density functional,^{26,27} Troullier–Martins

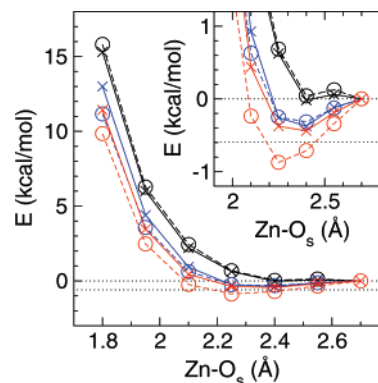


Figure 2. Potential energy profile for *N*-methylacetamide (NMA) binding to a model cofactor of thermolysin in the gas phase. ZnO_s denotes the distance between Zn and the oxygen atom of the carbonyl group of NMA. CPMD/TM/cut70//CPMD/TM/cut70 (black ○), CPMD/TM/cut100//CPMD/TM/cut70 (black ×), BLYP/6-31G(d,p)//BLYP/6-31G(d,p) (blue ○), BLYP/TZVP//BLYP/6-31G(d,p) (blue ×), B3LYP/6-31G(d,p)//BLYP/6-31G(d,p) (red ○), and B3LYP/6-311++G(d,p)//B3LYP/6-31G(d,p) (red ×). All energies are relative to the potential energy at a distance of 2.7 Å. The dotted lines represent the thermal energy $k_B T$ at 300 K. See sections 2 and 3.1 for details.

pseudopotentials,²⁸ and a reciprocal kinetic energy cutoff of 70 Ry. The final energies were refined using a larger cutoff of 100 Ry. For Zn, a pseudopotential similar to that of ref 29 was used. The pseudostates were generated for the $[Ar] 3d^{10}4s^{1.75}4p^{0.25}$ reference configuration. The 3d and 4s electrons were treated as valence and the angular momentum channels of Zn were 3d and 4s with pseudization radii of 1.8 au. See ref 30 for specification of pseudopotentials of second row elements. To test the CPMD pseudopotential calculations, all-electron geometry optimizations were carried out at the BLYP/6-31G(d,p) and B3LYP/6-31G(d,p) levels of theory using the Gaussian program package.³¹ The final potential energies were refined using the 6-311++G(d,p) basis.

Substrate Docking. The tripeptide Gly–Leu–Ala was capped with acetyl (Ace) at the N-terminal residue Gly and methylamide (Mam) at the C-terminal residue Ala. Neutral termini were chosen because charged termini would have created interactions not present in a longer peptide. The structure of the apoenzyme was taken from the Protein Data Bank (PDB), code 1LNF.⁶ Hydrogen atoms were generated using the HBUILD facility of CHARMM,³² and their positions were adjusted to minimize the energy. According to mechanism 1 of Figure 1, His231 was assumed to be protonated and Glu143 to be deprotonated. All other residues were protonated according to the respective protonation state in aqueous solution at pH = 7. The oxygen atom in the first coordination shell of Zn that is closest to Glu143 was modeled as a water molecule. The second first-shell oxygen atom closest to His231 was deleted because it was at the location of the substrate oxygen. Six further crystallographic water molecules interfering with the docking were selectively deleted. The docking of the substrate Ace–Gly–Leu–Ala–Mam into the enzyme pocket was carried out using successive energy minimizations for a four-dimensional represen-

tation of the substrate³³ as implemented in the CHARMM simulation package.³⁴ First, 20 independent conformations of the peptide were extracted from an independent 2 ns classical simulation of the peptide in explicit bulk water (one conformation every 100 ps). The orientation of each was systematically explored by rotating the peptide by 0, 90, 180, and 270° about 12 orientations covering the sphere. Each of these $48 \times 20 = 960$ rotated conformations was used as a starting peptide structure. Second, each peptide structure was superimposed to the 1LNF structure so that the substrate oxygen, O_s, was on top of the assigned zinc-binding site. The docking procedure avoids atomic clashes by embedding the substrate in a four-dimensional space. Each substrate atom *s* is given four spatial coordinates (*x_s*, *y_s*, *z_s*, and *w_s*) that are used to compute the distance with respect to each atom *e* of the enzyme: $r'_{es} = \sqrt{r_{es}^2 + w_s^2}$, where *r_{es}* is the distance in real, three-dimensional space and *w_s* the fourth component. The distance between two atoms *s* and *t* of the substrate is defined by $r'_{ts} = \sqrt{r_{ts}^2 + (w_s - w_t)^2}$. The potential energy was computed from the CHARMM22 empirical force field,³⁵ using the “embedded” *r'* distances instead of the real *r* distances. The force field was slightly modified to account for charge-transfer effects between zinc and its ligands. See ref 36 for the effective charges used. The {*w*} coordinates were restrained to a reference value *w*^{*} by adding the penalty term $u(\{w\}) = k \sum_s (w_s - w^*)^2$ to the potential energy. This ensured that the substrate maintained a relative chemical integrity in real space, and that each substrate atom was driven into the real, three-dimensional enzyme pocket in a controlled manner. A series of successive energy minimizations of the 4*N* coordinates were performed (*N* the number substrate atoms) for decreasing values of *w*^{*} (from 5 to ~0.1 Å, in 11 geometric decrements of ×0.7) while the enzyme, Zn atom, and crystallographic water molecules were maintained rigidly in the 1LNF conformation. A final energy minimization was performed with all *w*'s strictly enforced at 0 Å. The minimization procedure was repeated for a penalty constant *k* of 50, 30, and 10 kcal/mol/Å². In total, 3840 docking trials were performed.

Classical MD Simulations. Classical MD simulations were performed for the ES complex and the tetrahedral intermediate TI. The ES system was constructed by placing the best structure from the docking procedure in a 0.1 M KCl solution contained in a periodic box of dimensions 54 Å × 62 Å × 80 Å. The protein, substrate, and solvent were modeled with the Amber99 force field,³⁷ modifying the atomic charges of the zinc cofactor and its ligands by the same charge increments as in the docking procedure. The ES system was simulated at room temperature and pressure, using a 2 fs integration time step and constraining all bonds involving a hydrogen atom with the SHAKE algorithm.

The initial coordinates for the classical simulation of TI were taken from a free QM/MM simulation of length 2 ps. The simulation protocol is similar to the one of the ES complex. New RESP atomic charges were derived for a model of the TI form of the substrate computed at the B3LYP/6-31G* level (see Table 1). The model used was comprised only of TI in the gas phase: metal–ligand charge-

Table 1. Atomic Charges for the Oxy-Anion Gly(OH[−]) and Leu Backbone Atoms Used in the Classical Simulations of TI and TI'^a

residue	atom	Amber99	TI ^b
Gly(OH [−])	N	−0.4157	−0.38
	H	0.2719	0.23
	CA	−0.0252	−0.50
	HA1	0.0698	0.15
	HA2	0.0698	0.15
	C _s	0.5973	0.94
	O _s	−0.5679	−0.77
	O _w		−0.71
	H ₂		0.32
Leu	N _s	−0.4157	−0.77
	H	0.2719	0.29
	CA	−0.0518	−0.024
	HA	0.0922	0.07
	C	0.5973	0.36
	O	−0.5679	−0.43

^a For comparison, the charges from the original Amber99 force field for Gly and Leu residues are reproduced. All charges are in electrons. The RESP analysis was performed on the B3LYP/6-31G* electronic density of molecule Ace–Gly(OH[−])–Leu–Mam optimized at the B3LYP/6-31G* level. ^b The raw RESP charges are rearranged so that atomic charges of the Leu side chain and of groups Ace and Mam can retain their original values. The charge on the αC of Leu, −0.024e, corresponds to the combined RESP charges of αC and the hydrogen atom present in Gly but not in Leu, minus 0.074e, the total charge of the Amber99 Leu side chain. The total RESP charge on Ace, −0.10e, is transferred to the nitrogen of Gly, and the total RESP charge of Mam, −0.05e, is transferred to the carbonyl carbon of Leu.

transfer effects were not recalculated. To prevent the ligands from detaching from the zinc ion during the MD simulation, energy penalties were applied if any zinc–ligand distance was out of the range observed from QM/MM simulation of the tetrahedral intermediate. While affecting the short time dynamics of the first-shell ligands, the energy penalties are assumed to have a negligible effect on the nanosecond dynamics of second-shell ligands and backbone (which is the main focus of these classical MD simulations of the TI). In addition to free MD simulations, the TI form of the system was simulated using the adaptive biasing force (ABF) method^{38,39} (see section 3.3 for the details on the reaction coordinate).

QM/MM Simulations. The quantum region for QM/MM simulations included the metal ion; first-shell ligands His142, His146, and Glu166; the Zn-bound water molecule; second-shell residue Glu143; and the chemically active part of the substrate. The QM description was terminated at the αC position for protein residues and, for the substrate, at the αC position of Leu and βC position of Ala. The QM/MM boundary atoms were described by monovalent pseudopotentials.⁴⁰ There were 74 QM atoms, and the QM box dimensions were 33.07 × 33.07 × 34.07 au for nucleophilic attack and 35.07 × 35.07 × 35.07 au for protonation of the amide nitrogen atom. All remaining atoms of protein, substrate, and solvent were modeled with the Amber99 force field,³⁷ using the same system composition and atom topology as for the classical simulations described above. The interaction between the QM system and the MM system was computed using a Hamiltonian electrostatic coupling

Table 2. Energy Profile for Amide Bond Break in *N*-methylacetamide (NMA) Binding to a Gas-Phase Model of the Active Site of Thermolysin^a

	TS1 ^b	TI	TS2 ^c	P
BLYP/TM/cut70//BLYP/6-31G(d,p)	22.1	23.9	14.9	−3.1
BLYP/TM/cut100//BLYP/6-31G(d,p)	21.8	23.5	14.3	−3.2
BLYP/6-31G(d,p)//BLYP/6-31G(d,p)	22.3	23.2 ^d	11.6	−1.6
BLYP/6-311++G(d,p)//BLYP/6-31G(d,p)	22.2	24.1	14.3	−1.7
B3LYP/6-31G(d,p)//B3LYP/6-31G(d,p)	23.4	21.8	10.3	−0.6
B3LYP/6-311++G(d,p)//B3LYP/6-31G(d,p)	23.3	22.5	13.0	−0.5

^a TS1, TI, TS2, and P denote optimized model structures for the transition state for nucleophilic attack, tetrahedral intermediate, transition state for protonation of the amide group, and product, respectively. All energies are relative to the energy of the optimized enzyme–substrate complex (ES) and are given in kilocalories per mole. See sections 2 and 3.1 for details. ^b Geometry optimized with distance C_sO_w fixed at 1.85 Å. ^c Geometry optimized with distance H₁N_s fixed at 1.30 Å. ^d Geometry optimized with distance C_sO_w fixed at 1.56 Å.

scheme.⁴¹ QM/MM simulations were carried out with the CPMD code²⁵ using the BLYP density functional, Troullier–Martins pseudopotentials (see above), a reciprocal kinetic energy cutoff of 70 Ry, a fictitious mass of 700 au, and a time step of 5 au (0.1209 fs). After classical equilibration of the ES complex for 1 ns, a configuration was selected and equilibrated on the QM/MM potential energy surface for 5 ps at 300 K. During the first 1.5 ps of equilibration, separate Nosé–Hoover thermostats were used for QM atoms, protein atoms, and solvent atoms (water and ions). Final equilibration was carried out using a chain of Nosé–Hoover thermostats⁴² for the entire system with a target temperature of 300 K. During MD, the bonds between terminating monovalent carbon atoms and the respective QM atoms were fixed at the equilibrium distances taken from classical MD. The PMF for nucleophilic attack was obtained using constrained MD simulations, by taking the average of the Lagrange multiplier constraining distance C_sO_w (denoted as r_1) at a given value. The mean force was computed for 13 windows bridging the equilibrium distances in the ES and TI* complex (* denoting a particular conformation of TI): 3.2, 3.0, 2.8, 2.6, 2.4, 2.2, 2.0, 1.9, 1.8, 1.7, 1.6, 1.5, and 1.4 Å. The initial configuration for each window was selected from an equilibrated configuration of the previous window and the initial snapshot of the first window (3.2 Å) from the free ES simulation. Each constrained MD simulation was equilibrated for about 1–2 ps using again separate thermostats for each subsystem. The next 5 ps were used for calculation of the configurational averages. Except for two windows close to the transition state, the forces averaged over 2.5 and 5 ps were virtually identical, indicating that equilibration and production times were sufficient. At $r = 1.8$ Å, data were averaged over 20 ps, and at $r = 1.7$ Å, the system was equilibrated for 4 ps and averages were taken over 6 ps. Computation of the PMF for protonation of the amide nitrogen atom was carried out similarly using the distance H₁N_s (denoted as r_2) as a constraint. The mean force was calculated for 10 windows with distances constrained to 2.0, 1.8, 1.6, 1.4, 1.35, 1.3, 1.25, 1.2, 1.1, and 1.0 Å.

3. Results and Discussion

3.1. Gas-Phase Hydrolysis. The enzymatic reaction has been modeled in the gas phase as described in section 2. The energies for formation of TS1, TI, TS2, and P are shown in Table 2 for different levels of theory. The main purpose of these calculations is to assess the performance of the BLYP

functional used in QM/MM simulations. Validation relative to correlated wavefunction methods is not feasible because of the large size of the cofactor model. Instead, the BLYP calculations are compared to B3LYP calculations. The latter usually describes second-row chemistry better than BLYP, but unfortunately, it is computationally still too expensive for use in our plane wave code. The absolute values of the gas-phase energies should be interpreted with caution due to the simplicity of the model system chosen.

The energies computed with the plane wave basis set are fairly well converged at a moderate reciprocal space kinetic energy cutoff of 70 Ry (BLYP/TM/cut70, BLYP/TM/cut100). The change in energy is 0.1–0.6 kcal/mol when the cutoff is increased to 100 Ry. Similarly, small is the error due to the pseudopotentials used. The deviation of BLYP/TM/cut100 relative to all electron calculations at the BLYP/6-311++G(d,p) level of theory is not more than 0.6 kcal/mol for any reaction step except for formation of the product (1.5 kcal/mol). Remarkably, the optimizations with the BLYP functional did not give a stable minimum for TI. The optimized energy obtained by constraining the C_sO_w distance to a typical value in the transition state region of TS1, $r_1 = 1.85$ Å, is *lower* than for a typical equilibrium distance of TI, $r_1 = 1.56$ Å, by $22.1 - 23.9 = -1.8$ kcal/mol (see Table 2 and Figure 3 for notation of atoms). The drastic underestimation of energies for stretching of the C_sO_w bond—leading to the instability of TI—is a striking illustration of the self-interaction error of the BLYP density functional. The inclusion of exact exchange at the B3LYP level of theory cures this problem and gives a barrier of $23.3 - 22.5 = 0.8$ kcal/mol for dissociation of the C_sO_w bond.

In the gas-phase model, the second reaction step, protonation of the amide nitrogen atom and dissociation of the amide bond, is predicted to be continuously downhill in energy in contrast with present QM/MM simulations (see section 3.3). Barrierless protonation of the leaving group was also reported for an extended gas-phase model of this reaction.²⁴ The missing barrier for the second step which involves large molecular rearrangements could be due to an oversimplified reaction coordinate which was identified here with the distance H₁N_s. However, also, the more advanced transition state search carried out in ref 24 did not yield a reaction barrier, indicating that barrierless protonation could be an artifact of the simple gas-phase models used. Note that this deficiency disappears when the full enzymatic environment is included in the calculations.

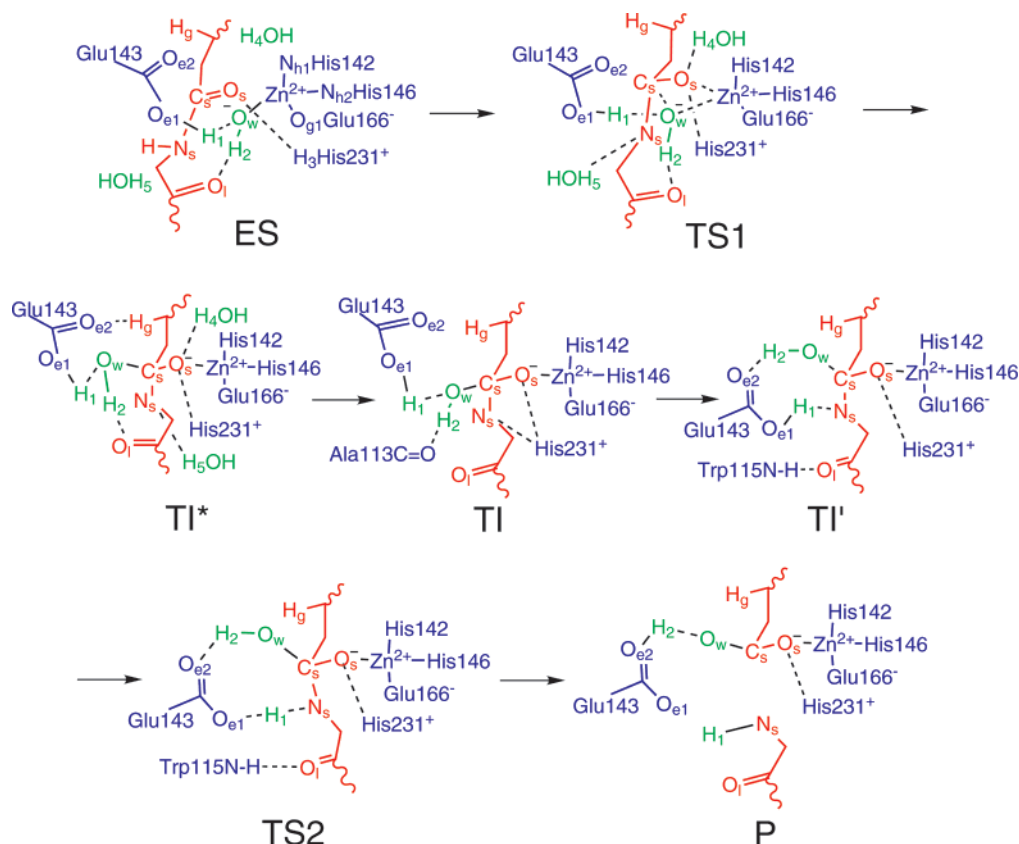


Figure 3. Reaction mechanism for Glu143-assisted peptide bond hydrolysis in thermolysin as obtained in this work. The substrate Ace–Gly–Leu–Ala–Mam is shown in red, catalytic residues of thermolysin in blue, and water molecules in green. Covalent and hydrogen bonds that become broken or formed during hydrolysis are depicted as solid and dashed black lines, respectively. ES denotes the enzyme–substrate complex; TS1 denotes the transition state for nucleophilic attack; TI*, TI, and TI' denote three distinct conformations of the tetrahedral intermediate; TS2 denotes the transition state for protonation of the leaving group; and P denotes the product. See sections 2 and 3.3 for details.

To estimate the error of the electronic structure calculation in QM/MM simulations (carried out at the BLYP/TM/cut70 level), we have compared the energy profile of the gas-phase model reaction obtained at the BLYP/TM/cut70 level to the one obtained at the B3LYP/6-311++G(d,p) level of theory. The barrier for formation of TI appears to be underestimated by $23.3 - 22.1 = 1.2$ kcal/mol relative to B3LYP/6-311++G(d,p), while the reaction energy for TI formation is overestimated by $23.9 - 22.5 = 1.4$ kcal/mol. Accordingly, the energy of TS2 is overestimated by $14.9 - 13.0 = 1.9$ kcal/mol, and the energy of the product P underestimated by $-0.5 + 3.1 = 2.6$ kcal/mol. The error estimates for TS2 and P are probably not as reliable as for TS1 and TI because the gas-phase energy profile for the second reaction step is barrierless and qualitatively different from the one in the enzyme.

3.2. Enzyme–Substrate Complex. Docking Structures. Figure 4A shows a representative set of the energetically most favorable conformations of the substrate Ace–Gly–Leu–Ala–Mam docked to thermolysin. It appears that low-energy docking structures are consistently oriented with the N-terminus of the peptide pointing toward Arg203 (“up”). Although this confornts the choice of an ES conformation for the simulation of reaction 1 of Figure 1, it does not imply that the enzyme is inactive for peptides presented with an N-terminus “down” orientation. By forcing the positions of

both water and substrate oxygen atoms, the system retains a relatively large conformational frustration. The extremities of the tripeptide can adopt many different conformations. The N-terminus of the peptide (at the top of Figure 4A) is either protruding away from the enzyme or buried into a small pocket, depending on which amino group of Arg203 binds the acetyl oxygen atom of Ace. The C-terminus is not forming any specific contacts with the protein and displays a considerable range of conformations.

The lowest-energy structure is shown in Figure 4B. It forms strong electrostatic interactions with residues Arg203, Asn112, and His231, and with the backbone of Trp115 and the hydroxyl group of Tyr157 (which are essential features observed for the ensemble of structures depicted in Figure 4A). The “buried” conformation of the N-terminus allows for an optimal interaction with the δ oxygen of Asn112. This interaction probably does not exist for a longer peptide chain as the latter would not fit into the small pocket. Under the electrostatic influence of positively charged His231 and negatively charged Glu143, the Gly–Leu peptide bond and the substrate water molecule are in an orientation favorable for nucleophilic attack.

Classical Molecular Dynamics Simulation. The conformation of Figure 4B was solvated and simulated with classical MD for 1 ns. In this conformation, the N-terminus of the substrate is stabilized by strong interaction of the acetyl

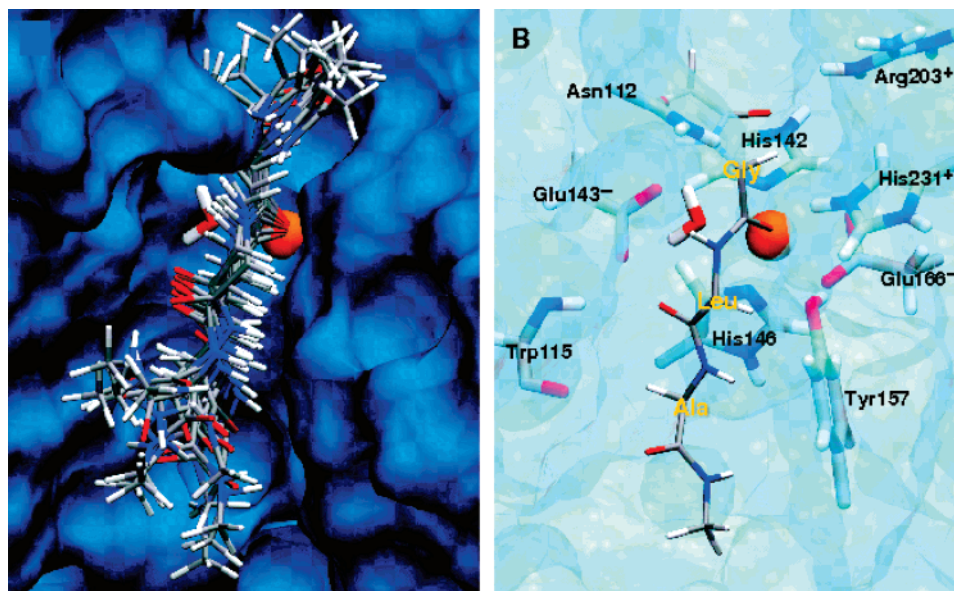


Figure 4. Low-energy conformations of the Ace-Gly-Leu-Ala-Mam tripeptide in the active site of thermolysin (PDB code 1LNF,⁶ Ace = acetyl, Mam = methylamide). The two putative oxygen-binding sites of the 1LNF structure are forcibly occupied by a water molecule and the substrate oxygen. Panel A shows representative docking conformations with a total energy within 24 kcal/mol of the energy minimum. Panel B shows the lowest-energy structure only, along with important residues of the active site. The enzyme is represented by its solvent-accessible structure, and the zinc atom is depicted as an orange sphere. Color code: O, red; N, blue; C, gray; H, white. For clarity, the side chains of the substrate are not shown.

oxygen atom of Ace with charged residue Arg203. Preliminary multi-nanosecond simulations have shown that this interaction was generally stable, except in rare instances where one water molecule from the solvent inserts between Ace and Arg203. To prevent water from occasionally disrupting the position of the substrate, an energy penalty was applied if the Ace oxygen atom separated from the Arg203 nitrogen atoms by more than 3.5 Å.

During the 1 ns MD simulation, the carbonyl oxygen atom of the substrate, O_s , remained coordinated to the Zn ion at an average distance of 2.29 Å, and in strong interaction with protonated His231. Note that simulations at the QM/MM level predict that the binding of O_s to Zn is very weak (see below), indicating that the classical model is slightly overbinding. On the opposite side of the peptide bond to be cleaved, the amide hydrogen atom of Leu was interacting mostly with the carbonyl oxygen atom of Ala113. The carbonyl group of Leu was weakly interacting with the backbone of Trp115 and the amide hydrogen atom of Ala, with the hydroxyl group of Tyr157. The C-terminus of the substrate, exposed to the solvent, had no specific interactions with the protein.

Two minor rearrangements of the ES complex structure were observed. First, the binding mode of Glu166 changed from bidentate to monodentate, presumably because both the water oxygen atom O_w and the substrate carbonyl oxygen atom O_s are bound to Zn (as in the inhibitor complex). Second, the substrate water molecule reoriented to form a hydrogen bond with the carbonyl oxygen atom of Leu, O_l . This vacancy on Glu143 was rapidly filled by the amide hydrogen atom of Gly that left the vicinity of Asn112 to be replaced by a water molecule from the solution and by the polar hydrogen of the backbone of Ala113. Overall, the zinc-ligand distances stayed comparable to the crystal-

Table 3. Selected Bond Lengths during Peptide Bond Hydrolysis in Thermolysin^a

	ES	TS1	TI*	TI'	TS2	P	E(cr) ^b	El(cr) ^c
ZnO _s	3.63 (2.29)	2.15, ^d 2.85 ^e	2.03	1.99	2.01	2.00	2.28	2.17
ZnO _w	1.97 (1.92)	2.58, ^d 2.08 ^e	3.02	2.81	2.77	3.18	2.38	2.59
ZnO _{g1}	2.03 (1.91)	2.05	2.04	2.08	2.08	2.05	2.24	2.04
ZnO _{g2}	3.30 (2.82)	3.19	3.06	2.96	3.03	3.04	2.38	2.92
ZnN _{h1}	2.06 (2.11)	2.08	2.08	2.11	2.10	2.10	1.98	2.09
ZnN _{h2}	2.13 (2.25)	2.13	2.12	2.11	2.10	2.07	1.99	2.11
C _s O _w	3.13 (3.02)	1.80 ^f	1.51	1.45	1.41	1.24		
C _s N _s	1.38	1.44	1.49	1.53	1.54	3.69		
H ₁ O _w	1.40	1.86	1.84	2.61	2.52	4.06		
H ₁ O _{e1}	1.15	1.02	1.02	1.02	1.31	4.64		
H ₁ N _s	4.00	3.85	3.98	2.00 ^f	1.30 ^f	1.00 ^f		
H ₂ O _w	0.99	1.01	1.01	1.00	1.04	1.97		
H ₂ O _l	2.00	1.89	1.77	5.17	5.28	5.40		
H ₂ O _{e2}	3.94	4.24	4.25	1.88	1.59	1.01		
H ₃ O _s	2.17	1.94	1.95	1.88	1.89	2.05		
H ₄ O _s	2.23	1.84	1.84	nb ^g	nb ^g	nb ^g		
H ₅ N _s	2.50	2.16	2.12	nb ^g	nb ^g	nb ^g		
H ₆ O _{g2}	1.81	1.80	1.80	1.82	1.81	1.81		

^a The data are obtained from unconstrained (ES, TI*) and constrained (TS1, TI', TS2, P) QM/MM simulations. Distances in parentheses are obtained from free classical MD simulations. All bond lengths are given in Angstroms. See Figure 3 for notation of atoms and section 2 for simulation details. ^b Crystal structure of thermolysin, PDB code 1LNF.⁶ ZnO_s and ZnO_w represent the distances of one single disordered water molecule or of two water molecules. ^c Crystal structure of thermolysin-inhibitor (ZFPLA) complex, PDB code 4TMN.⁷ ZnO_s and ZnO_w denote the distances between Zn and the two oxygen atoms of the phosphonamidate group. ^d Average over 4 ps after equilibration for 4 ps. ^e Average over the next 12 ps following the trajectory used in footnote d. ^f Distance constrained. ^g Not bonded.

structure values, in a coordination geometry combining characteristics from both the apo structure 1LNF and the inhibitor-bound structure 4TMN (see Table 3).

Residues Arg203 and Tyr157 (see Figure 4B) have a stabilizing effect on the substrate. Indeed, preliminary studies

using NMA as peptide model have revealed that a substrate that is too short does not provide enough secondary interactions to stabilize the orientation of the peptide bond during molecular dynamics. Although this structural instability may be aggravated by the deficiencies in the force field describing zinc ligation, it remains that the plane of the peptide bond to be broken matches the conformation of the reaction site better if the substrate has at least one additional peptide bond on each side.

QM/MM Simulations. The initial structure for QM/MM simulation of ES was taken from classical MD where the carbonyl oxygen atom of the substrate, O_s , was binding tightly to the Zn ion. Within the first picosecond of QM/MM simulation, the carbonyl oxygen atom was expelled from the first coordination shell. The ZnO_s distance fluctuated between 3 and 5 Å, with an average value of 3.6 Å (see Table 3). The remaining ligands (water, His142, His146, and Glu166) were found to bind tightly to the Zn ion, forming a 4-fold coordination sphere. The binding distances of these ligands are reasonably well reproduced with classical MD when compared to the QM/MM results.

Our observation that the substrate is not binding to Zn is not unusual for Zn peptidases. Related QM/MM studies of β -lactamase also showed that the substrate was not directly binding to the Zn atom.^{43,44} In contrast to our result, Pelmeshnikov et al. reported an equilibrium distance $r(ZnO_s) = 2.18$ Å for *N*-methylacetamide binding to a model cofactor for thermolysin (at 0 K) but noted that in the ES complex the interaction between Zn and the substrate via ZnO_s was the weakest along the entire reaction path.²⁴ To support the result of our QM/MM simulations, we have computed the potential energy for the gas-phase model as a function of the ZnO_s distance. As illustrated in Figure 2, the potential energy curve is indeed very flat, with a minimum at $r_m = 2.4$ Å. The weak interaction between Zn and the carbonyl oxygen is indicated by the very small energy required to stretch the ZnO_s bond from the minimum to a separation distance of 2.7 Å, $\Delta E_m = E(2.7 \text{ Å}) - E(r_m) = 0.4$ kcal/mol (BLYP/TZVP), 0.4 kcal/mol (B3LYP/6-311++G(d,p)), and <0.1 kcal/mol (BLYP/TM/cut70), respectively. At room temperature, this energy is smaller than $k_B T$, explaining the large fluctuations observed in QM/MM simulations.

Judging from the measured Michaelis–Menten constants, the binding of Cbz–Gly–Leu–P₂ substrates to thermolysin is slightly exergonic: $K_M = 2.4$ to 20.6 mM.⁴⁵ The corresponding free energies are –2.3 to –3.6 kcal/mol. Our results indicate that the hydrogen bonds formed between the substrate and thermolysin contribute most to the binding free energy whereas the interaction between Zn and the carbonyl oxygen atom is very small.

In the ES complex, the substrate and the water molecule are almost perfectly aligned for nucleophilic attack (see structure ES of Figure 3). The distance between the reactive water molecule and the carbonyl carbon atom C_s is on average 3.13 Å. As suggested by our simulations, the active hydroxide nucleophile is already formed in the ES complex. Hydrogen atom H_1 is observed to switch back and forth frequently between the O_w and O_{e1} of Glu143 (see Figure

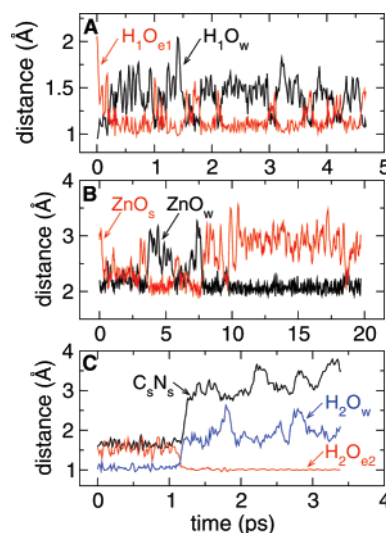


Figure 5. Fluctuations of selected bond lengths at 300 K (A) in the ES complex, (B) first transition state TS1, and (C) shortly after the second transition state TS2 is reached. In A, the fluctuations are obtained from equilibrium QM/MM simulation (no constraints), in B from QM/MM simulation with the distance C_sO_w constrained to 1.8 Å, and in C from QM/MM simulation with the distance H_1N_s constrained to 1.25 Å. See Figure 3 for notation of atoms and sections 2 and 3.3 for further details.

5A). The average bond lengths are 1.40 Å for H_1O_w and 1.15 Å for H_1O_{e1} . The carbonyl group of the substrate is also well preorganized through formation of a hydrogen bond with the protonated histidine His231. Tyr157, which was shown to have a modest catalytic effect,¹⁵ does not form hydrogen bonds with the carbonyl oxygen atom O_s . The hydrogen atom of the phenol side chain is instead tightly bonded to the oxygen atom O_{g2} of Glu166 along the entire reaction path (see distance H_6O_{g2} in Table 3).

3.3. Enzymatic Hydrolysis. The mechanism for the Glu143-assisted hydrolysis of the substrate is summarized in Figure 3. Selected bond lengths for transition states and intermediates are given in Table 3 and shown in Figures 5 and 6. The free energy profiles for the individual reaction steps are illustrated in Figures 6 and 7 and combined into a single profile in Figure 8.

Transition State for Nucleophilic Attack (ES → TS1). Transformation of the ES complex into the tetrahedral intermediate was enforced by constrained QM/MM simulations using the distance C_sO_w as reaction coordinate r_1 . The mean force and the corresponding free energy profile (PMF) are shown in Figure 7A and B. The mean force is close to zero at the average distance in the ES complex and increases almost linearly between 2.8 and 2.2 Å corresponding to a quadratic PMF. At 2.4 Å, hydrogen H_1 is fully transferred to Glu143 and strongly hydrogen-bonded to the reactive hydroxide ion. Remarkably, the average distances between Zn and first-shell ligands for TS1 differ by not more than 0.02 Å from the distances in the crystal structure of the enzyme–inhibitor complex⁷ (see Table 3).

In the transition state region at around 1.8 Å, we observe two binding modes of the substrate (see Figure 5B). In the first 8 ps of QM/MM dynamics, the substrate is bonded to Zn (2.15 Å) while hydroxide is displaced (2.58 Å). This

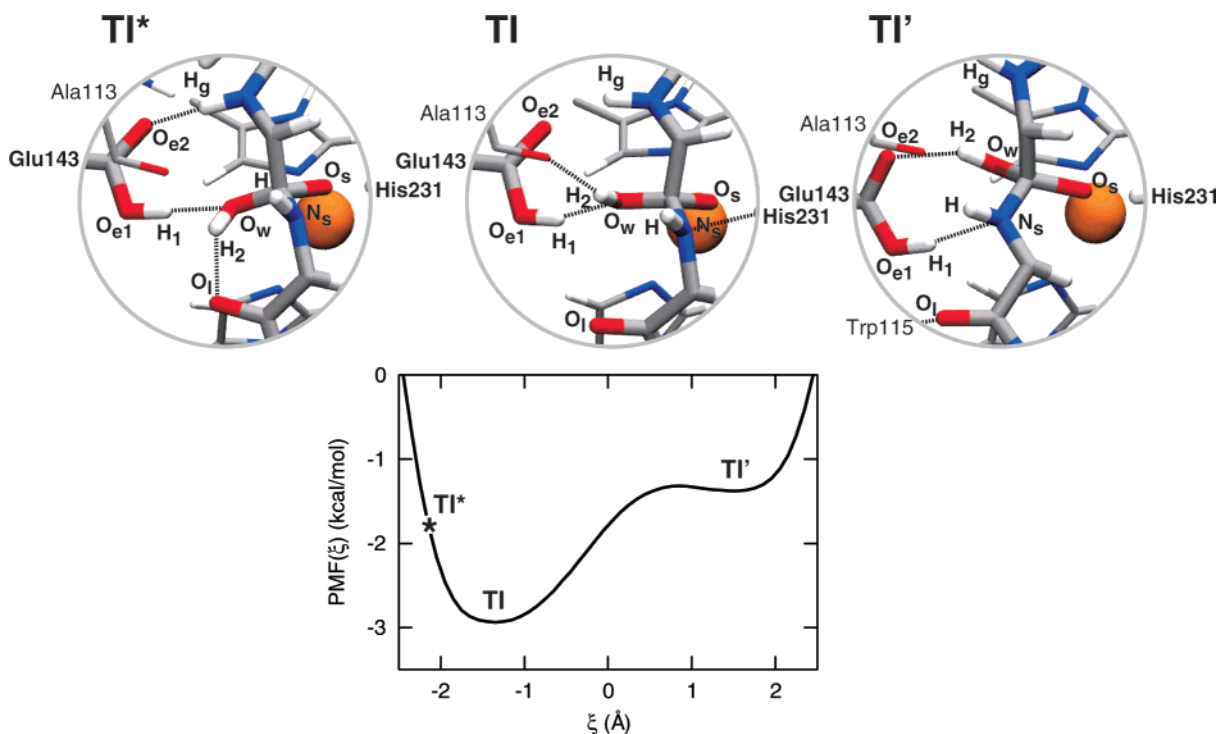


Figure 6. Reorganization of the tetrahedral intermediate described in terms of $\xi = d_{OO} - d_{ON}$, where d_{OO} is the distance between O_I (the carbonyl oxygen of substrate residue Leu) and the center of mass of pair O_wH_2 , and d_{ON} is the distance between the centers of mass of pairs $O_{e1}H_1$ and N_sH . Snapshots of representative conformations along the reorganization pathway are shown as insets. In free energy basin “TI”, Glu143 is interacting with H_g and O_w , and His231 is interacting both with O_s and N_s . In basin “TI’”, Glu143 is interacting with H_2 and N_s , and His231 is interacting with O_s only. Left-most structure (TI*) is a high-energy configuration in which the intrasubstrate hydrogen bond is partially formed. For clarity, the side chains of the substrate are not shown.

binding mode is favorable for reaction to the tetrahedral intermediate, and therefore the mean force is negative (Figure 7A, ○ at $r_1 = 1.8$ Å). During the following 12 ps of dynamics, the binding mode is reversed: the substrate is bonded weakly (2.85 Å), whereas hydroxide is bonded tightly (2.08 Å). In this case, the forward reaction is unfavorable, as indicated by a positive mean force (Figure 7A, × at $r_1 = 1.8$ Å). Evidently, sufficient sampling of the transition between the two binding modes would require even longer simulations (100 ps or more) which is beyond current capabilities. For computation of the PMF, we took the average of the mean force in the two binding modes and considered the difference as statistical uncertainty. The structural rearrangement of the substrate in the transition state region is not affected by the binding mode. Besides formation of the fluctuating ZnO_s bond (3.63 Å in ES, 2–3 Å in TS1), the advancing formation of the oxy-anion is stabilized through the formation of two strong hydrogen bonds, one formed with protonated His231 (2.17 Å in ES, 1.94 Å in TS1) and one with a solvent molecule (2.23 Å in ES, 1.84 Å in TS1). Moreover, the increased negative charge on the amide nitrogen atom is stabilized by the formation of a weak hydrogen bond with a solvent molecule (2.50 Å in ES, 2.16 Å in TS1). In the transition state region, the carbonyl Π bond is partly broken, as indicated by the beginning pyramidalization of carbon and nitrogen atoms.

Tetrahedral Intermediate (TS1 \rightarrow TI*). A decrease of the reaction coordinate from 1.8 to 1.5 Å led to a negative mean force or, equivalently, to a loss of free energy. At $r_1 = 1.5$

Å, the tetrahedral intermediate TI* is formed and separated from the reactant by a free energy barrier of 4.0 kcal/mol (* denotes a particular conformation of the tetrahedral intermediate; see below). The oxy-anion O_s is tightly bonded to Zn (2.03 Å), and the hydrogen bonds with His231 and the solvent molecule are fully formed. The oxygen O_w of the hydroxo group is only weakly bonded to Zn (3.02 Å), while the distances of the remaining first-shell ligands are almost unchanged when compared to those of the ES complex. The formation of TI* can therefore be viewed as a ligand exchange reaction where the Zn-bound nucleophile is replaced by the substrate carbonyl oxygen atom, leaving Zn 4-fold coordinated in the reactant and product states. The tetrahedral intermediate is a stable species in the enzyme as opposed to the model in the gas phase (see section 3.1).

Rearrangement of Tetrahedral Intermediate (TI* \rightarrow TI \rightarrow TI’). The hydrogen-bonding pattern of TI* as obtained from the relatively short QM/MM constrained dynamics is rather unfavorable for subsequent protonation of the leaving group (see structure TI* of Figure 3). The amide nitrogen atom N_s is initially hydrogen-bonded to a solvent molecule, while the nearest acidic protons, H_1 at Glu143 and H_3 at His231, are separated from N_s by as much as 4 Å. Atom H_1 of protonated Glu143 forms a hydrogen bond with O_w , while the carbonyl oxygen atom of Glu143 (O_{e2}) is bonded to the proton of the amide group of Gly, H_g . Glu143 cannot deliver H_1 to N_s because hydrogen atom H_2 is blocking the transfer, the latter forming a strong intramolecular hydrogen bond with O_I , the carbonyl oxygen atom of Leu. The rather large separation

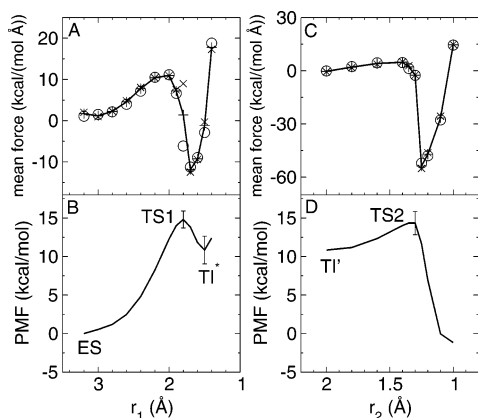


Figure 7. Computed mean force and free energy profile for Glu143-assisted hydrolysis of Ace-Gly-Leu-Ala-Mam catalyzed by thermolysin. (A) Mean force and (B) PMF for nucleophilic attack of the amide bond along the distance C_sO_w , r_1 . (C) Mean force and (D) PMF for protonation of the amide nitrogen atom along the distance H_1N_s , r_2 . The mean force averaged over the first and second halves of the QM/MM trajectory (5 ps/window in total) are denoted by circles (○) and crosses (×), respectively. (Two points in A were averaged over longer trajectories, see discussion in section 3.3.) The average of the two values for the mean force is denoted + and used for the calculation of the PMF shown in B and D. The error bars in B and D denote the difference in free energy obtained when the first and the second halves of the trajectories are used for calculation of the mean force. The PMF in B is set equal to zero for the ES state at $r_1 = 3.2$ Å.

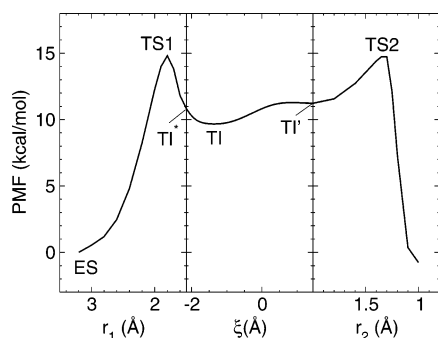


Figure 8. PMF for the full enzymatic reaction obtained by merging the PMFs shown in Figures 6 and 7. See section 3.3 for details.

distance between H_1 and N_s is a consequence of the “up” orientation of the substrate in the active site pocket (see section 3.2 and Figure 4). The backbone atoms between C_s and the N-terminus of the substrate are close to Glu143, while the backbone atoms between C_s and the C-terminus (including the C_sN_s peptide bond) are at the far end. In the inhibitor complex of ref 7, the substrate is aligned in the opposite direction, which facilitates direct protonation of N_s .

In order to probe the stability of the hydrogen-bonding pattern, we have carried out classical MD simulation for the tetrahedral intermediate. Within a few 100 picoseconds of dynamics, major reorganization of the hydrogen-bonding pattern was observed (see Figures 3 and 6). First, the intrasubstrate hydrogen bond H_2O_1 broke, and a new hydrogen bond between H_2 and the carbonyl oxygen atom of

Ala113 was formed, converting TI^* into TI . Second, the hydrogen bond H_gO_{e2} broke, and a new hydrogen bond between H_2 and O_{e2} was formed. This rearrangement of TI into a structure we call TI' brings H_1 into the vicinity of N_s . Due to the large amplitude motion of Glu143, H_1 only forms a transient hydrogen bond with N_s and frequently binds to O_1 or reverts to form a hydrogen bond with O_w as in the TI structure. Figure 6 shows representative conformations of the hydrogen bond network as it rearranges into the reactive conformation TI' . The transition from TI to TI' requires simultaneously a translation of Glu143 and a torsion of the substrate. The wide-range motion of Glu143 is facilitated by the presence of residues Ala113 and Trp115, whose backbone atoms are competing with atoms O_{e2} for hydrogen bonding with atoms H_g of Gly and O_1 of Leu. Our analysis shows that Glu143 is indeed a highly effective proton shuttle, even in cases where the substrate is not favorably aligned for protonation.

The breaking of the intrasubstrate hydrogen bond H_2O_1 ($TI^* \rightarrow TI$) is consistent with ab initio energy minimizations done on a mimic of the isolated substrate. At the B3LYP/6-31G* level, the minimum-energy structure has no intrasubstrate hydrogen bond and is 4 kcal/mol more stable than a structure optimized with atoms O_w and O_1 kept 2.62 Å apart (i.e., forming a hydrogen bond as in the final snapshot of the QM/MM simulation). The breaking happens on the nanosecond time scale in the protein presumably because it requires a reorganization of the whole substrate.

The free energy corresponding to this reorganization is computed using the ABF method^{38,39} for a reaction coordinate ξ equal to $d_{OO} - d_{ON}$. d_{OO} is the distance between atom O_1 and the center of mass of pair O_wH_2 . It describes the breaking of the intrasubstrate hydrogen bond and the gradual exposition of the amide nitrogen N_s to acidic residue Glu143. d_{ON} is the distance between the center of mass of pair $O_{e1}H_1$ and the center of mass of pair N_sH . ξ is minimum in the TI^* structure when Glu143 is in its original, “high” position and when the intrasubstrate hydrogen bond is formed, and it is maximum in the TI' structure when Glu143 is in position to protonate the amide nitrogen. The graph of Figure 6 shows the free energy profile obtained from 20 ns of ABF simulation. Two distinct groups of structures can be identified: TI structures, for which Glu143 binds substrate atoms H_g and O_w , and TI' structures, for which Glu143 binds atoms H_2 and N_s . The TI' structures are 1.5 kcal/mol less stable than structures TI .

The TI^* structures obtained from the relatively short QM/MM simulations do not correspond to a minimum in the long classical simulations. The average value of ξ calculated from a 2 ps free QM/MM simulation of TI^* is -2.15 Å. The ABF free energy at this value is 1.1 kcal/mol higher than the free energy of TI . Despite the coordination restraints preventing the Zn ligands from detaching, a water molecule from the solution inserted into the coordination shell of Zn during the third nanosecond of the ABF simulation. Although this insertion made the metal ion 6-fold coordinated, and slightly distorted its original coordination structure, it has a minor effect on the PMF of Figure 6. Indeed, we have extended the ABF simulation after manually expelling the spurious

Table 4. Computed and Experimental Free Energies for Peptide Bond Hydrolysis Catalyzed by Thermolysin^a

	TS1	TI*	TI	TI'	TS2	P
Antonczak et al. ^b	60		20		22	7
Antonczak et al. ^c	50		1		5	1
Antonczak et al. ^d	20.49		17.00		41.57	−3.91
Pelmenschikov et al. ^e	15.2		6.9		4.3	−8.8
this work ^f	14.8 ± 1.1	10.8 ± 1.8	9.7 ^g	11.2 ^g	14.7 ± 1.5	
this work ^h	16.0		8.3		12.8	
experiment ⁱ	13.6 ^j					
	12.1 ^k					

^a TS1 denotes the transition state for nucleophilic attack; TI*, TI, and TI' denote conformations of the tetrahedral intermediate; TS2 denotes the transition state for breakdown of the tetrahedral intermediate; and P denotes the product. All free energies are given in kilocalories per mole relative to the free energy of the enzyme–substrate complex. See section 2 for details. ^b Energy profile for the nonassisted hydrolysis of formamide estimated from Figure 5 of ref 53. QM/MM, AM1/AMBER. ^c Energy profile for the water-assisted hydrolysis of formamide estimated from Figure 5 of ref 53. QM/MM, AM1/AMBER. ^d Energy profile for the Glu143-assisted hydrolysis of the tripeptide Gly–Phe–Leu, taken from ref 16. QM/MM, AM1/AMBER. ^e Free energy profile for the Glu143-assisted hydrolysis of NMA, taken from ref 24. B3LYP gas-phase optimizations + self-consistent isodensity continuum model (SCI–PCM). ^f Potential of mean force (PMF) for the Glu143-assisted hydrolysis of Ace–Gly–Leu–Ala–Mam, Ace = acetyl, Mam = methylamide. CPMD/AMBER QM/MM, BLYP/TM/cut70 unless stated otherwise. The free energies are taken from Figure 8. ^g Estimated from classical free energy simulation. ^h Energy-corrected PMF. The correction is the energy difference between B3LYP/6-311++G(d,p) and BLYP/TM/cut70 levels of theory obtained for the model reaction in the gas phase (see Table 2). ⁱ Effective activation free enthalpy calculated from measured rate constants k_{cat} using the classical transition state formula. ^j Ref 45. Substrate: Cbz–Gly–Leu–Ala–OH, pH = 7, $k_{\text{cat}}/K_{\text{M}} = 78\,000\text{ s}^{-1}\text{M}^{-1}$, $K_{\text{M}} = 10.6\text{ mM}$. ^k ref 49. Substrate: Fua–Gly–Leu–Ala–OH, $k_{\text{cat}}/K_{\text{M}} = 870\,000\text{ s}^{-1}\text{M}^{-1}$, K_{M} assumed to be the same as for Cbz–Gly–Leu–Ala–OH.

water molecule, and the PMF showed little change. While the free energy difference between structures TI and TI' are probably reliable within less than $k_{\text{B}}T$, the free energy from structures TI* to TI (from $\xi = -2.15$ to -1.35 \AA) could be less reliable because it involves the breaking of an intramolecular hydrogen bond (H_2O_1) for which the empirical force field was not explicitly parametrized.

Protonated residue His231 is forming a strong interaction with substrate atom O_5 during most of the 20 ns of the biased classical MD simulation. For TI structures, His231 occasionally gets at close range to the amide nitrogen N_s , in a position where it could transfer a proton (see structure TI of Figure 6). However, if it were to protonate N_s , His231 would have to get reprotonated for the next catalytic cycle to proceed. Similarly, Glu143 would have to get deprotonated. There is no proton chain connecting His231 with Glu143 that could help restore protonation states. Deprotonation of Glu143 by a solvent molecule also does not seem likely as it is not exposed to the solvent. For this reason, we have not considered His231 as a potential proton donor facilitating the breakdown of TI'.

Transition State for Proton Transfer (TI' \rightarrow TS2). For the QM/MM simulation of the protonation of the amide group, we have taken a TI' structure from the classical simulation, where H_1 of Glu143 forms a strong hydrogen bond with N_s . The proton transfer was enforced using again constrained QM/MM simulation and the distance r_2 between H_1 and N_s as geometrical constraint (see Figure 7C and D). The mean force is zero at 2.0 \AA , increases slightly up to a small maximum value at 1.4 \AA , and vanishes again in the transition state region $1.30\text{--}1.35\text{ \AA}$. The corresponding free energy barrier is 3.5 kcal/mol . At the second transition state TS2, the second proton transfer is about to begin, as indicated by the slightly elongated H_2O_w bond and the stronger interaction between H_2 and O_{e2} . Yet the amide bond is not significantly elongated.

Product Formation (TS2 \rightarrow P). A decrease of the H_1N_s distance from 1.30 to 1.25 \AA leads to an irreversible break

of the amide bond. Virtually at the same time, H_2 is transferred from O_w to O_{e2} (see Figure 5C). The break of the peptide bond is indicated by a sudden jump of the mean force to a large negative value. After the peptide bond breaks, the carboxylate and amine groups separate quickly from one another (Figure 5C). The distance C_sN_s fluctuates between $3\text{--}4\text{ \AA}$ for the remainder of the simulation. The mean force remains negative up until the approximate equilibrium distance of 1.05 \AA is reached.

Evidently, the N-protonated tetrahedral intermediate is not a stable species; full protonation and breakage of the peptide bond occur virtually at the same time. A concerted breakdown of the tetrahedral intermediate was also reported for the hydrolysis of amide bonds in the gas phase^{46,47} and in β -lactamase.⁴³ The reaction coordinate r_2 captures the structural reorganization for TI' \rightarrow TS2 well but cannot describe reversibly the spontaneous break of the peptide bond. Since we are primarily interested in reaction barriers, we have not carried out additional simulations to compute the free energy of the product. This could be done by taking the distance difference between H_1N_s and C_sN_s as a reaction coordinate instead of r_2 .

Full Free Energy Profile. The free energy profile for the full hydrolysis reaction is shown in Figure 8 and summarized in Table 4. The PMFs of Figures 6 and 7B are merged by connecting the minimum TI* of the latter with the free energy at $\xi = -2.15\text{ \AA}$ of the former. This value of ξ is the average computed from a free QM/MM trajectory for TI* of length 2 ps. Similarly, the minimum of the PMF shown in Figure 7D is connected to the minimum denoted TI' in Figure 6.

The free energy barrier for nucleophilic attack is $14.8 \pm 1.1\text{ kcal/mol}$, obtained as the difference in free energy at $r_1 = 1.8\text{ \AA}$ and $r_1 = 3.2\text{ \AA}$. The uncertainty represents the difference in free energy obtained when averages were computed for the first and second half of the constrained trajectories, respectively (denoted by error bars in Figure 7). The major contribution to the statistical error is the uncer-

tainty of the mean force at the transition state (see discussion above). The barrier increases to 16.0 kcal/mol if the underestimation of the gas-phase reaction barrier relative to the B3LYP density functional is taken into account (see section 3.1).

Formation of the tetrahedral intermediate TI reduces the free energy by 5.1 kcal/mol relative to TS1 or by 7.7 kcal/mol if corrected again relative to B3LYP gas-phase calculations. Reorganization of unreactive TI into the reactive TI' conformation is reversible on the time scale of the enzymatic reaction. Protonation of the leaving group leads again to an increase in free energy by 3.5 kcal/mol to 14.7 ± 1.5 kcal/mol. Energy correction relative to B3LYP gives a value of 12.8 kcal/mol. However, as mentioned in section 3.1, the gas-phase reaction that was adopted to calculate the correction energy is not expected to model protonation as well as nucleophilic attack. The difference of the reaction barriers for nucleophilic attack and protonation, 0.1 kcal/mol at the BLYP level of theory, is significantly smaller than the statistical uncertainties, which prevents us from identifying the rate-limiting step of the overall reaction.

The free energy profile presented in Figure 8 does not include nuclear quantum effects. They are generally expected to slightly lower the barriers. For chorismate mutase, for instance, the zero-point energy correction to the reaction barrier was calculated to be -1.5 kcal/mol.⁴⁸ Nuclear tunneling effects are expected to play a minor role for nucleophilic attack since proton transfer from O_w to O_{e1} occurs spontaneously and well before transition state TS1 is reached. However, tunneling effects might be important for protonation of the nitrogen atom. In this case, the second barrier is expected to slightly decrease and the difference between the first and second reaction barrier would be more pronounced.

The largest barrier height of the reaction, 14.8 kcal/mol at the BLYP level of theory, is in good agreement with the experimental activation free energy for hydrolysis of the Gly–Leu peptide bond in Cbz–Gly–Leu–Ala–OH, 13.6 kcal/mol,⁴⁵ and Fua–Gly–Leu–Ala–OH, 12.1 kcal/mol⁴⁹ (pH = 7). When the barrier is corrected for the error of the BLYP functional (relative to B3LYP), +1.2 kcal/mol, and for the zero-point energies, assumed to be between -1 and -2 kcal/mol, our estimate for the barrier is 0.4–2.9 kcal/mol higher than the experimental activation free energies for the two substrates. The experimental values were obtained from the measured rate constants k_{cat} using the classical transition state formula and unity for the transmission coefficient. The latter is typically between 0.5 and 1 for enzymatic reactions.^{50–52} Adopting a conservative value of 0.5, the experimental activation free energies decrease by 0.4 kcal/mol, leading to an increase of the overestimation of the experimental barrier by the same amount. The relatively large variation of experimental rate constants among substrates with different terminal groups implies that part of the discrepancy with experimental results could stem from the use of acetyl and *N*-methylester as terminal groups of the substrate.

4. Conclusion

Starting from the crystal structure of the apoenzyme, a tripeptide was docked to the active site of thermolysin and the free energy profile for peptide bond cleavage computed using state-of-the-art QM/MM calculations. After the nucleophilic attack of deprotonated water, Glu143 and the oxy-anion substrate rearrange into a form where Glu143 is in position to protonate the amide group. Accounting for the free energy of this rearrangement, the barrier heights for nucleophilic attack and protonation of the leaving group are almost identical at the BLYP level of theory and overestimate the experimental activation free energies by 1.2–2.7 kcal/mol depending on the substrate used in the experiment. We expect that the barrier is overestimated by the same energy range if the error of the BLYP density functional (relative to B3LYP), zero-point energies, tunneling effects, and the deviation of the transmission coefficient from unity are included in the calculation.

The QM/MM molecular dynamics simulations carried out in this work indicate that the carbonyl oxygen atom is *not* coordinated to Zn in the ES complex. Stabilization of the transition state is achieved through formation of a weak bond between the carbonyl oxygen atom of the substrate and Zn and formation of a strong hydrogen bond with protonated His231. The formation of the oxy-anion is further stabilized by two hydrogen bonds with solvent molecules. Dynamical effects to catalysis are expected to be small for thermolysin due to the presence of Ca^{2+} , which makes the backbone motions very rigid.²⁰ Thermolysin therefore complies with Pauling's paradigm that enzymes accelerate rates because they bind the transition state better than the substrate and thereby lower the activation barrier.

In this study, we have shown that the Glu143-assisted water addition mechanism, which is best supported by many crystallographic and biochemical studies, is energetically feasible and has a free energy barrier that is slightly higher but still close to the effective barrier calculated from experimental measurements. We have further shown that the conserved residue Glu143 is a highly effective proton shuttle which is capable of transferring the proton to the leaving group even if the substrate is not ideally aligned. However, while intuitive, the mechanism investigated can explain the observed pH dependence of k_{cat} only if Glu143 is identified with the catalytic residue that has a $\text{p}K_{\text{a}} = 8.26$. Although Glu143 is not exposed to the solvent, this value seems rather high. Mock and Stanford¹⁷ suggested that His231 could be this particular residue that in place of Glu143 deprotonates the reactive water molecule (mechanism 2 in Figure 1). Preliminary classical MD simulations show, however, that deprotonated His231 does not form a strong hydrogen bond with a water molecule in plane with the histidine ring as required for deprotonation. Instead, the ϵN atom of His231 is prone to form a hydrogen bond with the substrate. Although we cannot exclude Mock and Stanford's proposal on this basis, we agree with the authors of ref 1 that the His231-based mechanism is rather unlikely. The proposal that a Zn-bound hydroxide ion is the nucleophile rather than water (mechanism 3 in Figure 1) seems to be a more likely alternative that will be investigated in future work.

Acknowledgment. This research was supported by NIH under grant GM 67689. We want to thank Drs. Matteo Dal Peraro and Jérôme Hénin for helpful discussions and Dr. Axel Kohlmeyer for valuable advice on QM/MM molecular dynamics simulations. G.L. is a fellow of the Fonds Québécois de la Recherche sur la Nature et les Technologies (FQRNT). J.B. acknowledges financial support from The Royal Society for a University Research Fellowship. High-performance computing resources were provided by the Pittsburgh Supercomputing Center.

References

- (1) Lipscomb, W. N.; Sträter, N. *Chem. Rev.* **1996**, 96, 2375.
- (2) Parkin, G. *Chem. Rev.* **2004**, 104, 699.
- (3) Roques, B. P.; Noble, F.; Dauge, V.; Fournie-Zaluski, M. C.; Beaumont, A. *Pharmacol. Rev.* **1993**, 45, 87.
- (4) Marukami, Y.; Chiba, K.; Oda, T. A. H. *Biotechnol. Bioeng.* **2001**, 74, 406.
- (5) Matthews, B.; Jansonius, J. N.; Colman, P. M.; Schoenborn, B. P.; Dupourque, D. *Nature* **1972**, 238, 37.
- (6) Holland, D. R.; Hausrath, A. C.; Juers, D.; Matthews, B. *Protein Sci.* **1995**, 4, 1955.
- (7) Holden, H. M.; Tronrud, D.; Monzingo, A. F.; Weaver, L.; Matthews, B. *Biochemistry* **1987**, 26, 8542.
- (8) Bartlett, P. A.; Marlowe, C. K. *Biochemistry* **1987**, 26, 8553.
- (9) Matthews, B. *Acc. Chem. Res.* **1988**, 21, 333.
- (10) Toma, S.; Campagnoli, S.; De Gregoriis, E.; Gianna, R.; Margarit, I.; Zamai, M.; Grandi, G. *Protein Eng.* **1989**, 2, 359.
- (11) Kubo, M.; Mitsuda, Y.; Takagi, M.; Imanaka, T. *Appl. Environ. Microbiol.* **1992**, 58, 3779.
- (12) Devault, A.; Nault, C.; Zollinger, M.; Fournie-Zaluski, M.-C.; Roques, B. P.; Crine, P.; Boileau, G. *J. Biol. Chem.* **1988**, 263, 4033.
- (13) Corbeil, D.; Milhiet, P.-M.; Simon, V.; Ingram, J.; Kenny, A. J.; Boileau, G.; Crine, P. *FEBS Lett.* **1993**, 335, 361.
- (14) Beaumont, A.; O'Donohue, M. J.; Paredes, N.; Rousselet, N.; Assicot, M.; Bohuon, C.; Fournie-Zaluski, M.-C.; Roques, B. P. *J. Biol. Chem.* **1995**, 270, 16803.
- (15) Marie-Claire, C.; Ruffet, E.; Tiraboschi, G.; Fournie-Zaluski, M.-C. *FEBS Lett.* **1998**, 438, 215.
- (16) Antonczak, S.; Monard, G.; Ruiz-Lopez, M. F.; Rivail, J.-L. *J. Mol. Model.* **2000**, 6, 527.
- (17) Mock, W. L.; Stanford, D. J. *Biochemistry* **1996**, 35, 7369.
- (18) Mock, W. L.; Aksamavati, M. *Biochem. J.* **1994**, 302, 57.
- (19) Inouye, K.; Lee, S.-B.; Nambu, K.; Tonomura, B. *J. Biochem.* **1997**, 122, 358.
- (20) Feder, J.; Garrett, L. R.; Wildi, B. S. *Biochemistry* **1971**, 10, 4552.
- (21) Benkovic, S. J.; Hammes-Schiffer, S. *Science* **2003**, 301, 1196.
- (22) Schramm, V. L. *Curr. Opin. Struct. Biol.* **2005**, 15, 604.
- (23) Olsson, M. H. M.; Parson, W. W.; Warshel, A. *Chem. Rev.* **2006**, 106, 1737.
- (24) Pelmentschikov, V.; Blomberg, M. R. A.; Siegbahn, P. E. M. *J. Biol. Inorg. Chem.* **2002**, 7, 284.
- (25) CPMD, version 3.10; The CPMD consortium, MPI für Festkörperforschung and the IBM Zurich Research Laboratory: Zurich, Switzerland, 2005. <http://www.cpmc.org> (accessed Jul 2007).
- (26) Becke, A. D. *Phys. Rev. A: At., Mol., Opt. Phys.* **1988**, 38, 3098.
- (27) Lee, C.; Yang, W.; Parr, R. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, 37, 785.
- (28) Troullier, N.; Martins, J. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1991**, 43, 1993.
- (29) Rothlisberger, U. *ACS Sym. Ser.* **1998**, 712, 264.
- (30) Blumberger, J.; Klein, M. L. *Chem. Phys. Lett.* **2006**, 422, 210.
- (31) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; 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.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision C.01; Gaussian, Inc.: Wallingford, CT, 2004.
- (32) Brünger, A. T.; Karplus, M. *Proteins* **1988**, 4 (2), 148,156.
- (33) Crippen, G. M. *J. Comput. Chem.* **1982**, 3 (4), 471,476.
- (34) Brooks, B. R.; Brucoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, 4, 187–217.
- (35) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kucsera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, M.; Reiher, W. E., III; Roux, B.; Schlenkerich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kucsera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, 102 (18), 3586–3616.
- (36) Dal Peraro, M.; Spiegel, K.; Lamoureux, G.; De Vivo, M.; DeGrado, W. F.; Klein, M. L. *Struct. Biol.* **2007**, 157, 444–453.
- (37) Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, T. E., III; Wang, J.; S. Ross, W. S.; Simmerling, C.; Darden, T.; Merz, K. M.; Stanton, R. V.; Cheng, A.; Vincent, J. J.; Crowley, M.; Tsui, V.; Gohlke, H.; Radmer, R.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G. L.; Singh, U. C.; Weiner, P.; Kollman, P. A. *AMBER 7*; University of California: San Francisco, 2002.
- (38) Darve, E.; Pohorille, A. *J. Chem. Phys.* **2001**, 115 (20), 9169.
- (39) Hénin, J.; Chipot, C. *J. Chem. Phys.* **2004**, 121 (7), 2904.
- (40) Blumberger, J.; Klein, M. L. *J. Am. Chem. Soc.* **2006**, 128, 13854.

- (41) Laio, A.; VandeVondele, J.; Röthlisberger, U. *J. Chem. Phys.* **2002**, *116*, 6941.
- (42) Martyna, G. J.; Klein, M. L.; Tuckerman, M. *J. Chem. Phys.* **1992**, *97*, 2635.
- (43) Dal Peraro, M.; Llarrull, L. I.; Rothlisberger, U.; Vila, A. J.; Carloni, P. *J. Am. Chem. Soc.* **2004**, *126*, 12661.
- (44) Díaz, N.; Suárez, D.; Merz, K. M., Jr. *J. Am. Chem. Soc.* **2001**, *123*, 9867.
- (45) Morihara, K.; Tsuzuki, H. *Eur. J. Biochem.* **1970**, *15*, 374.
- (46) Bakowies, D.; Kollman, P. A. *J. Am. Chem. Soc.* **1999**, *121*, 5712.
- (47) Lopez, X.; Inäki Mujika, J.; Blackburn, G. M.; Karplus, M. *J. Phys. Chem. A* **2003**, *107*, 2304.
- (48) Claeysens, F.; Harvey, J. N.; Manby, F. R.; Mata, R. A.; Mulholland, A. J.; Ranaghan, K. E.; Schütz, M.; Thiel, S.; Thiel, W.; Werner, H.-J. *Angew. Chem., Int. Ed.* **2006**, *45*, 6856.
- (49) Blumberg, S.; Vallee, B. L. *Biochemistry* **1975**, *14*, 2410.
- (50) Nam, K.; Prat-Resina, X.; Garcia-Viloca, M.; Devi-Kesavan, L. S.; Gao, J. *J. Am. Chem. Soc.* **2004**, *126*, 1369.
- (51) Roca, M.; Moliner, V.; Tunon, I.; Hynes, J. T. *J. Am. Chem. Soc.* **2006**, *128*, 6186.
- (52) Garcia-Viloca, M.; Gao, J.; Karplus, M.; Truhlar, D. G. *Science* **2004**, *303*, 186.
- (53) Antonczak, S.; Monard, G.; Ruiz-Lopez, M. F.; Rivail, J.-L. *J. Am. Chem. Soc.* **1998**, *120*, 8825.

CT7000792