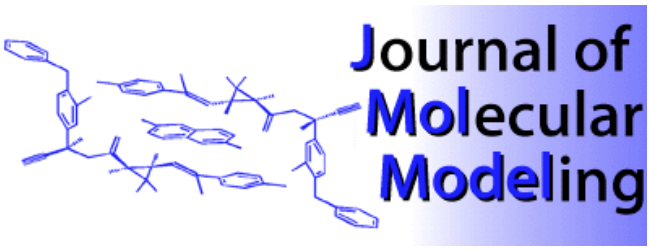
J. Mol. Model. **2000**, *6*, 527 – 538



FULL PAPER

Insights in the Peptide Hydrolysis Mechanism by Thermolysin: A Theoretical QM/MM study

Serge Antonczak‡, Gérald Monard, Manuel Ruiz-López, and Jean-Louis Rivail

Laboratoire de Chimie Théorique, Unité Mixte de Recherche n° 7565, Université Henri Poincaré - Nancy I., BP 239, F-54506 Vandœuvre-lès-Nancy Cedex, France. Tel.: +33-(0)3 83 91 20 50; Fax: +33-(0)3 83 91 25 30.
E-mail:rivail@lctn.u-nancy.fr

Received: 9 February 2000/ Accepted: 28 March 2000/ Published: 16 August 2000

Abstract The hydrolysis by thermolysin of a Gly-Phe-Leu peptide, considered as a model substrate of the enkephalin family, has been studied with a mixed QM/MM method with the AM1/AMBER parameterization. This study is based on the mechanism proposed by Matthews in which the Glu-143 residue plays the role of a proton shuttle in the course of the reaction. The study focused on the description of every step of the process, reaction intermediates and transition states, and on the influence, both energetical and structural, of the whole protein on these stationary points. The overall mechanism appears to be quite realistic, but the study shows that some reaction steps that were assumed to be concerted should occur in two phases. Analysis of the role of the amino-acids surrounding the active site has shown their important participation in the fluctuations of the energy. In particular, the major role of His-231 on the overall mechanism has been confirmed. This study shows that modeling reaction mechanisms for enzymes is quite feasible and opens the way for computer experiments that may be helpful in devising and interpreting detailed experimental investigations.

Keywords Hybrid methods, QM/MM, LSCF, Zinc metalloproteases, Thermolysin, Peptide hydrolysis, Enzymatic mechanism

Introduction

Thermolysin (TLN), a zinc metallopeptidase, is the best-studied member of the clan of enzymes containing the HExxH+E zinc-binding motif.[1] The first X-Ray structure of thermolysin was solved in 1972 by Matthews *et al.*[2,3] Since, several structures of TLN have been characterized with different inhibitors with different ways of binding. In this HExxH+E sub-family, the structures of pseudolysin [4] and bacillolysin [5,6] have also been determined but many human enzymes of the same group, such as neprilysin (NEP), peptidase O (pepO), endothelin-converting enzyme (ECE) have not yet been crystallized. Indeed, thermolysin being a bacterial enzyme, the interest of studying such a metalloprotesase would have been limited if the characteristics (hydrolysis mechanism and specificity) of the above mentioned human enzymes and TLN were not so close. Therefore, thermolysin is often considered as a model for the other metalloproteases of its clan.

Correspondence to: J.-L. Rivail

‡*Present adress:* GRECFO-LARTIC, Université de Nice-Sophia Antipolis, F-06108 Nice Cedex 2, France.

Thermolysin is a thermostable enzyme produced by *Baccilus thermoproteolyticus*. This thermal stability [2,3,7] arises from the presence of four calcium ions in its structure. The three amino acid ligands of the zinc cation are His-142, His-146 and Glu-166 and therefore, the second conserved glutamate residue is Glu-143. These amino acids are positioned on two helices (residues 137-150 and 160-180), among the seven present in thermolysin, which form a cross centered at the active site. The 137-150 helix plays the particular role of link between the two lobes of thermolysin, namely residues 1-136 and 151-316. The fourth ligand position is filled either by a water molecule in the free wild type enzyme or by a substrate or an inhibitor. The successive inhibitor/TLN co-crystallized structures also brought important informations on the role of several amino acids close to the zinc atom. In the active site, a S1' hydrophobic subsite is involved in the recognition of the inhibitors and substrates. It is constituted by the sidechains of the following residues: Phe-130, Leu-133, Val-139, Ile-188, Gly-189, Val-192 and Leu-202. Other residues, such as Asn-112, Ala-113, Arg-203 and His-231, are involved in the stabilization of the inhibitors or substrates through hydrogen bonds. Generally speaking, the hydrogen bond network between the amino acids of the first two shells of residues surrounding the metal is important for the activity of the enzyme.[8] Mutating to Ala the Asp-170 residue, which is involved in hydrogen bonds with both Arg-203 and His-142 to Ala, leads to drastic decrease of the specific activity of thermolysin.

Thermolysin hydrolyses peptidic bonds on the amino side of a hydrophobic residue. Among the wide variety of substrates, one can put forward the benzocarbonyl derivatives,[9] Leucine Enkephalin,[10] the glutaryl derivatives [11,12] but also the atrial natriuretic factor [13] which is a physiological substrate of Neprylisin. A first hydrolysis peptide mechanism by thermolysin has been proposed on the basis of the X-Ray structures which have permitted to establish structural models for the different reaction steps,[14] In this process, the Glu-143 residue plays an important role, as a proton shuttle between the reactant water molecule and the peptidic bond to be hydrolyzed. More recently, an alternative mechanism has been proposed,[15,16] in which the amino acid involved in the successive proton transfers is His-231. In this scheme, contrarily to the former one, His-231 is supposed to be neutral in the enzyme and a proton transfer occurs from the reacting water molecule on the histidine residue. Then, the resulting OH^- ion is involved in a nucleophilic attack of the carbon atom of the peptidic bond, leading to the breaking of the substrate. Even if some results based on kinetics [16] and crystallographic experiments [17] give arguments in favour of this latter mechanism, the former remains favored. Indeed, on mutation of Glu-143, a drastic loss of activity of thermolysin is noted and at the moment, this fact cannot be accommodated by the second mechanism.

Taking all these factors into account is not simple [18] when one considers theoretical investigations of the catalytic mechanism of such a metalloprotease. Indeed, studying the breaking and the formation of chemical bonds is not simple by means of molecular mechanics methods since the electrons that describe these bonds are not taken into account. On the other hand, in full quantum calculations, the size of the system is limited by constraints regarding the computer time. An alternative approach is the use of hybrid methods,[19] often denoted QM/MM methods, which combine the advantages of the both formalisms. The region of the system in which the chemical reaction takes place is described in a quantum way while the rest of the macromolecule is described by means of a classical force field. Many such methods have been developed [20–24] this last decade and used to investigate enzymatic activity. Such methods have been applied to investigate the importance of the coordination of the metal [25,26] and to point out the role of specific residues in alcohol dehydrogenase [27] but to the best of our knowledge none have been employed for the study of a entire hydrolysis mechanism by a zinc metalloprotease.

Most such methods are limited to the semiempirical [19] level but more recently hybrid methods based on Density Functional [28-30] or *ab initio* [31] formalisms have been developed as well. However, the mixed methods using semi-empirical description of the reactive part remain the most widely used in the case of enzymatic mechanism investigations due to the number of atoms involved in the chemical process. Contrarily to most of the available methods, which use an artificial link atom to treat the split between the two subsystems, the Local Self Consistent Field method [31-33] (LSCF) used in this work explicitly treats the bond separating the quantum part from the classical one. In a previous study,[34,35] this method proved to be efficient in predicting the catalytic strength of thermolysin by comparison with full quantum studies on model systems and allowed us to discuss the influences, both energetic and structural, of the whole enzyme [36] on the reaction mechanisms.

Thermolysin is known to hydrolyze peptides of the enkephalin family and the Tyr-Gly-Gly-Phe-Leu (YGGFL) peptide is considered as one of its best natural substrates. Simulating the reactivity of such an enzyme using theoretical chemistry tools should at best involve a whole quantum description of the system composed of the enzyme, its YGGFL substrate and the surrounding solvent water molecules. Extensive use of molecular dynamics should then allow us to access reaction free energies which could be directly compared to experiment results. Unfortunately this ideal way of simulating enzymatic reactions is still far from being achieved due to the actual limited power of computers. Therefore, one must make several approximations in the description of the system to get a good qualitative and even semi-quantitative insight into the reactivity of thermolysin.

The first approximation to make is to reduce the part of the system described at a quantum level. Use of QM/MM methods allows us to describe the active site of thermolysin directly involved in the hydrolysis mechanism by quantum mechanics, whereas the rest of the enzyme, which does not participate directly in the reaction, is described by classical molecular mechanics. In the present work, in order to reduce the size of the substrate, and since it seems that the Tyr-Gly N-terminal part of the substrate does not take an active part in the hydrolysis process, we focused our attention on the

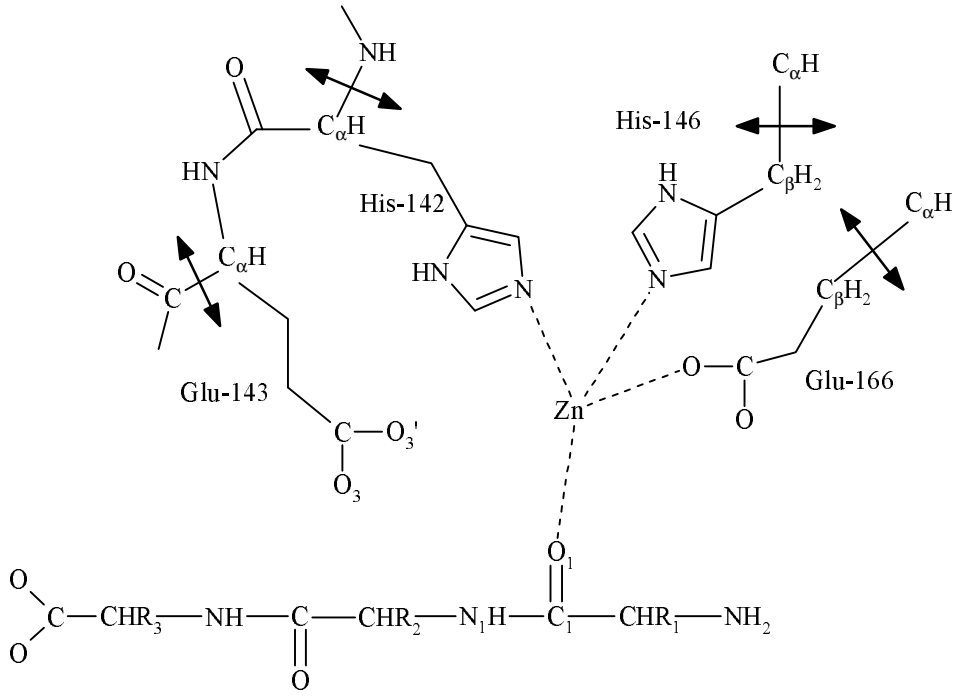


Figure 1 *Schematic description of the active site of thermolysin complexed with the GFL tripeptide. Bold arrows locate the border between the QM and MM regions*

hydrolysis mechanism by thermolysin of the Gly-Phe-Leu peptide considered as a model peptide of the enkephalin family. We based our investigations on the reaction scheme proposed by Matthews since, as mentioned above, it seems to be the most widely accepted mechanism.

Methodology

The main problem arising with QM/MM methods is the description of the QM/MM interactions, especially when simulating enzyme reactivity, where covalent bonds must be shared between atoms located at the border between the two QM and MM regions (see Figure 1). This so-called frontier bond must be described at the quantum level in order to reproduce its behavior as if the whole system were included in the quantum computation. One way to carry out such a computation is to use the Local Self Consistent Field approach which has been extensively described elsewhere.[31-33] Its basic principles are the following. The frontier bonds are described by strictly localized bond orbitals (SLBOs) whose electronic properties are considered constant during the whole reaction process (this assumption is valid provided that the reaction is located far enough away from the frontier bonds). These SLBOs are frozen (i.e. they do not change during the computation) and a local SCF procedure is carried out using a basis set of orbitals orthogonal to these frozen SLBOs in order to generate the molecular orbitals describing the quantum subsystem.

The LSCF formalism has been developed at the *ab initio* [31] and semiempirical levels [32,33] and provides a very powerful tool for studying reactivity in enzymes.[37] In the present case, the necessity of considering in the quantum subsystem at least the chemically important groups of the active site, a polypeptidic chain to be hydrolyzed and a water molecule, the number of atoms rises to more than 100. The use of an *ab initio* or a DFT method is still not feasible. Therefore, we limited our study to the semiempirical level. Previous studies on model peptide hydrolysis at semiempirical, DFT and *ab initio* levels [34,35] have shown a good agreement between the structures obtained for this different kind of computations, although the energy barriers are systematically overestimated at the semiempirical level compared to *ab initio* and DFT, but these data can still be used to support a qualitative reasoning. Furthermore, there is strong evidence that the substrate is strongly tied to the thermolysin active site, so that we made the assumption that the geometry of the MM part of the system (*i.e.* the rest of the enzyme) does not change significantly during the reaction and has been then considered as fixed along the reaction path.

LSCF semiempirical computations have been performed using the GEOMOP program [38] at the NDDO level with AM1 parameterization.[39,40] The electrostatic interactions between the quantum and the classical subsystems (point charges) are included in the Hamiltonian. The classical force field used is AMBER [41] with special potentials for the zinc atoms developed by Jacob.[42]

Preparation of the system has been made using a complete MM description of the system. Minimization procedures have been performed on the whole system with the DISCOVER program [43] developed by MSI using either steepest descent or conjugate gradients methods.

Intermediate and transition states along the reaction path have been then located by varying the coordinates of the atoms entering the quantum subsystem. Transition states have been characterized using Schlegel's algorithm,[44] verifying that the Hessian has only one negative eigenvalue, corresponding to an imaginary frequency. Reaction intermediates were directly related to the TS by following the transition vector on both sides, using the IRC formalism,[45,46] and then fully optimized.

Preparation of the system

The initial atomic coordinates of the enzyme are taken from the crystallographic structures available in the Protein Data Bank.[47] The hydrogen atoms are added using the usual procedures. The proton of the acidic and basic sidechains are distributed according to the standard pK_a of these groups for an hypothetical pH of 7. The consequence of this is that the total charge of the protein is -2e. In our computations we did not add the counterions which are often considered as having a constant contribution to the energy of the system and we only discussed the energy variations along the reaction path.

As mentioned previously, the Tyr-Gly-Gly-Phe-Leu (YGGFL) peptide [11] is often used by experimentalists as a basis of comparison to measure the activity of thermolysin or related metalloproteases. The hydrolysis of such a peptide takes place between the Gly and the Phe amino-acids. The second glycine amino-acid is linked to the enzyme through a bond between its carbonyl oxygen and the zinc cation (see Figure 1). No other covalent or coordination bonds are proposed to exist but several H-bonds between the substrate and the enzyme are assumed to stabilize the complex. Furthermore, the phenyl group of the Phe amino-acid of the substrate peptide is probably inserted in the S_1' hydrophobic subsite of the thermolysin protease. All these conclusions arise from analyses of co-crystallized structures of TLN with several inhibitors.

Because of the above considerations, and in order to reduce the size of the system, we chose to limit our study to a Gly-Phe-Leu tripeptide but, in order to study a system that can be considered as a model of a larger peptide, the Gly N-terminal residue was not protonated. On the other hand, the Leu C-terminal is assumed to be in its charged form as in the pentapeptide.

This tripeptide has been docked in the enzyme with the help of the cocrystallized structures of the TLN protease with some inhibitor and following the above propositions. This structure was then minimized fully at the MM level using both steepest descent and conjugate gradient algorithms. The resulting Gly-Phe-Leu/thermolysin complex shows the same structural trends as those of the crystallized structures. Under these assumptions the net electronic charge of the substrate is -1, so that the enzyme-substrate interaction has a positive Coulomb contribution, which would vanish on adding counterions. This unphysical situation will be corrected by considering the energy variations along the reaction path.

The above complex was used as a starting point for the QM/MM study and was divided into two parts (see Figure 1). In the quantum part, we have included the metal, the His-142, His-146 and Glu-166 ligands, the GFL substrate as well as the Glu-143 residue since, following the hypothesis of Matthews, this residue plays an important role in the hydrolysis mechanism. For the two His-146 and Glu-166 ligands, the split was made between the C_α and the C_β atoms. This point is acceptable since it has been demonstrated previously [36] that along a reaction path the most important changes in these ligands occur at their π-system, *i.e.* the imidazole ring for His and the carboxylic group for Glu. The fragment including His-142 and Glu-143 was considered as a whole *i.e.* the sidechains and the peptidic unit between them. The splits were made between the C_α and the N atoms of the His-142 residue on one side and between the C_α and the C atoms of the Glu-143 residue on the other side. The rest of the enzyme was described by means of the classical AMBER force field.

Results

Hydrolysis mechanism

The breaking of the peptidic bond of the substrate involves a series of stationary points whose main characteristics can be found in Table 1, while a schematic representation of the reaction path is described in Figure 2. The atom numbering is defined on Figure 3.

Initial Complex I1 Starting from the enzyme-substrate complex I0, a water molecule is introduced into the active site of the enzyme between the Gly-Phe peptidic bond and the carboxylic group of the Glu-143 residue. The quantum part is then optimized. In the resulting structure (see Figure 3), this water molecule can be considered as a fifth ligand for the zinc atom if one considers the zinc-oxygen interatomic distance (Zn-O_2: 2.275 Å). At the AM1 level of approximation, the two hydrogen atoms of the incoming water molecule are involved in two hydrogen bonds with the oxygen atoms of the Glu-143 carboxylic group (O_3-H_2: 1.941 Å and O_3'-H_2': 2.009 Å). Computations performed on a model system including a water molecule and a formate ion with other semi-empirical methods or at the *ab initio* level show that this interaction may involve only one hydrogen atom. Nevertheless, in the protein, the constraints imposed by the limited space are also in favor of this double interaction. At this stage, no strong interactions arise between the water molecule and the substrate (C_1-O_2: 2.501 Å and H_2-N_1: 2.790 Å). Conversely, the introduction of this new ligand leads to an increase of the Zn-O_1 bond-length (+0.061 Å) and to an important decrease of the C_1-N_1 bond-length (-0.084 Å) while C_1-O_1 does not vary.

TS1 transition State According to our results, the next step is a proton transfer from the water molecule to the gluta-

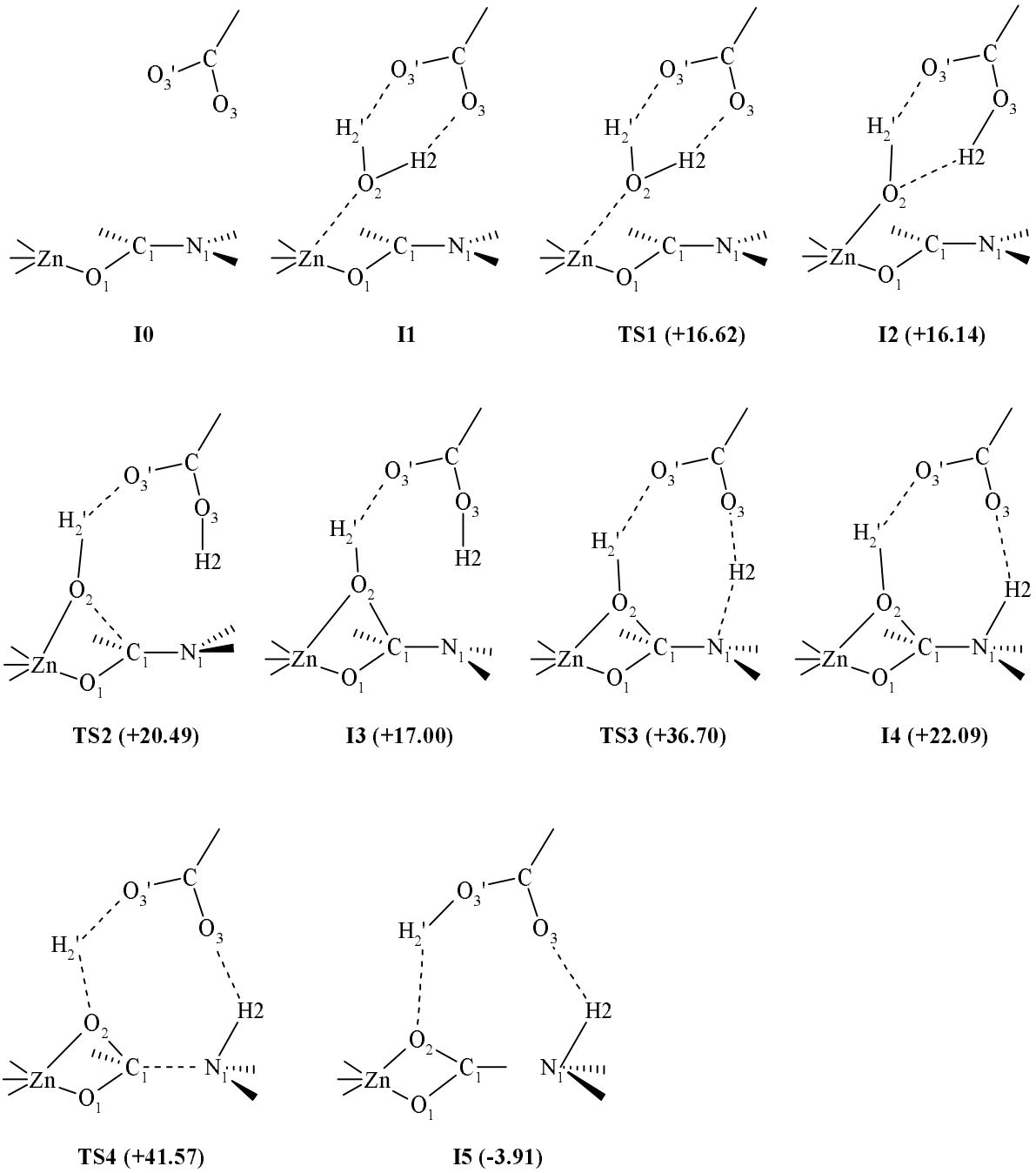


Figure 2 *Schematic description of the reactive part for the stationary points along the reaction path. Energies (in kcal·mol^{-1}) with respect to intermediate I1 are given in parenthesis*

mate143, which occurs through a transition state denoted TS1 on Figure 2 and located 16.6 kcal·mol^{-1} above the previous complex. The two hydrogen atoms of the water molecule exhibit a very different behavior and a careful analysis of the various possibilities shows that the only possible process is the transfer of the hydrogen atom denoted H_2 in Figure 2. Transfer of the H_2' proton appears impossible as well as a concerted attack of the carbon of the peptidic bond by the oxygen atom of the water molecule. In the characterized transition state, the hydrogen transfer is almost complete: the O_3-H_2 bond is short (1.052 Å) while O_2-H_2 is elongated to a large extent (1.470 Å). Therefore, an OH^- group, involving the H_2' and O_2 atoms of the water molecule is formed. The electronic population of the oxygen atom increases drasti-

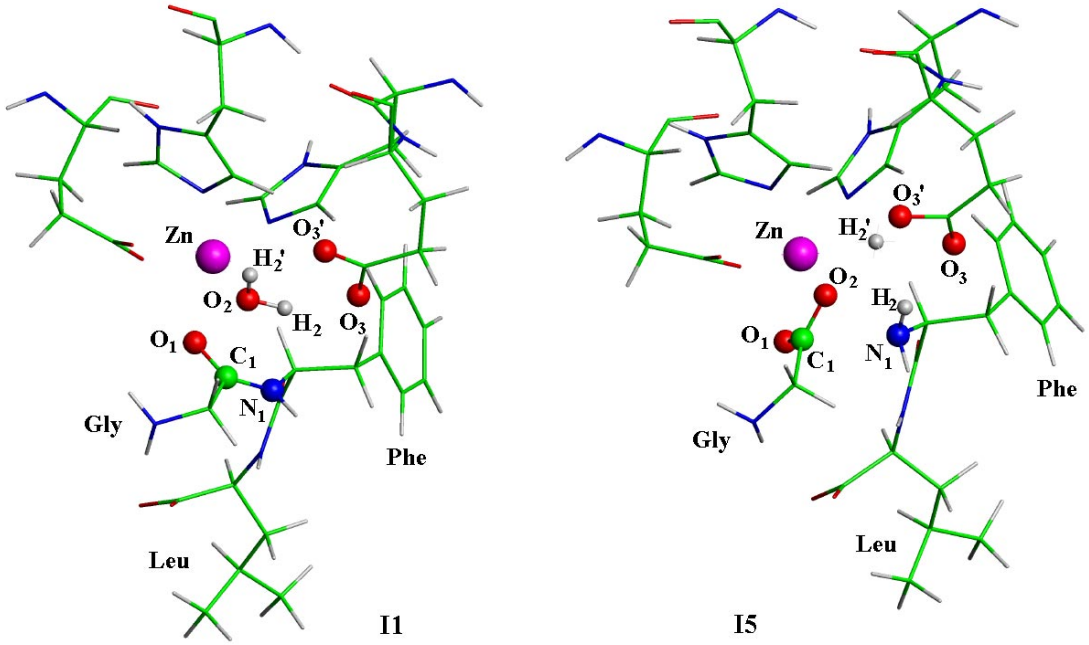


Figure 3 *Structures of the initial and final intermediates, I1 and I5 respectively, for the hydrolysis mechanism. Only the quantum part is presented*

cally and leads to an important decrease of the Zn-O_2 bond-length. A consequence of this change is an increase of the Zn-O_1 bond-length associated to a slight elongation of the peptidic bond and a slight shortening of the carbonyl bond. These changes in the structure are reflected by the evolutions of the net atomic charges. The peptidic bond to be broken is only slightly perturbed in a way which prepares its breaking by a slight departure from planarity (the $O_1C_1N_1C_1'$ dihedral angle is 172°) in conjunction with a shortening of the C_1-O_2 bond (-0.132 Å).

I2 Intermediate By following the eigenvector of the TS1 transition state, an intermediate in which the H_2 hydrogen atom is completely transferred to the carboxylic group of Glu-143 is characterized and denoted I2 (see Figure 2). In this complex, one can consider that the fifth ligand of the metal is a OH^- group. Except for a slight shortening of the O_2-C_1 bond with respect to the previous transition state, only small evolutions of the geometry can be noted. Thus, I2 is only weakly stabilized with respect to TS2 (-0.48 kcal·mol^{-1}) and the reality of these two stationary points cannot be considered as certain.

TS2 Transition State The following step appears to be the nucleophilic attack of the peptidic carbon atom by the OH^- group formed during the previous step. Here again several possible reaction paths have been considered and the computations have shown that the only feasible one is the nucleophilic attack of the O_2 atom on the C_1 carbon atom of the peptidic bond. This approach leads to a transition state which is only 4.35 kcal·mol^{-1} above the I2 intermediate. The approach of the OH^- plays the major role in the reaction coordinate. However, in this transition state, the C_1-O_2 bond-length remains long (1.836 Å) and indicates that the interaction of the carbonyl group with the zinc atom does not activate the peptidic bond strongly. This approach leads to an important perturbation of the π-system of the peptidic bond as illustrated by the evolutions of the O_1-C_1 and C_1-N_1 bond-lengths, by the charges of the corresponding atoms and by the $O_1C_1N_1C_1'$ dihedral angle which decreases to 149°. This is accompanied by a slight pyramidalization at the nitrogen atom. Note also that Zn-O_1 and Zn-O_2 bonds have almost the same length but cannot be considered as equivalent since, for example, the charges of the two oxygen atoms are different (Table 1).

I3 Intermediate The nucleophilic attack leads to an intermediate denoted I3, in which the main features observed in TS2 are enhanced except for the metal oxygen bond-lengths. The Zn-O_2 distance is now shorter than Zn-O_1 but this is easily understandable since the O_2 oxygen atom of the reacting water molecule is involved in two covalent bonds, with H_2' and C_1. This point is reflected by the drastic decrease of the negative charge of this oxygen atom. Pyramidalization of the nitrogen atom and tetrahedralization of the carbon atom of the peptidic bond are now visible. Furthermore, taking into account the relatively long H_2'-O_3' interatomic distance and the relatively weak negative charge of the O_3' atom, a hydro-

Table 1 *Bond-lengths (Å) and Mulliken net Atomic Charges Corresponding to the Stationary Points. See Figure 2 for Notations*

Bond-lengths										
	I0[a]	**I1**	**TS1**	**I2**	**TS2**	**I3**	**TS3**	**I4**	**TS4**	**I5**
ZnO_1	2.16	2.22	2.26	2.27	2.15	2.07	2.13	2.17	2.14	2.27
O_1C_1	1.28 (1.26)	1.27	1.27	1.27	1.0	1.34	1.33	1.31	1.32	1.27
C_1N_1	1.44 (1.37)	1.35	1.37	1.37	1.40	1.45	1.52	1.58	1.65	3.04
ZnO_2		2.28	2.06	2.03	2.14	2.37	2.27	2.28	2.18	2.25
O_2C_1		2.50	2.37	2.37	1.84	1.51	1.46	1.44	1.38	1.30
H_2N_1		2.79	2.87	2.82	2.81	2.87	1.36	1.04	1.03	1.01
H_2O_3		1.95	1.05	1.00	0.99	0.99	1.31	1.93	2.08	2.20
$H_2'O_3'$		2.02	2.15	2.18	2.47	2.64	2.05	2.07	1.16	0.98
$H_2'O_3$		-	-	-	-	-	-	2.40	-	-
$H_2'N_1$		-	-	-	-	-	-	-	-	-
O_2H_2		0.98	1.47	1.73	1.82	-	-	-	-	-
O_2H_2'		-	-	-	-	-	-	-	1.31	1.95
Mulliken Net Atomic Charges										
	I0[a]	**I1**	**TS1**	**I2**	**TS2**	**I3**	**TS3**	**I4**	**TS4**	**I5**
Zn	0.74	0.72	0.69	0.68	0.68	0.68	0.68	0.69	0.68	0.72
O_1	-0.41 (-0.44)	-0.42	-0.41	-0.42	-0.50	-0.55	-0.53	-0.51	-0.55	-0.45
C_1	0.35 (0.26)	0.30	0.39	0.34	0.38	0.31	0.26	0.27	0.32	0.31
N_1	-0.36 (-0.37)	-0.33	-0.35	-0.35	-0.38	-0.39	-0.28	-0.14	-0.15	-0.34
O_2		-0.41	-0.69	-0.72	-0.62	-0.46	-0.39	-0.39	-0.60	-0.56
H_2		0.26	0.34	0.33	0.32	0.31	0.37	0.28	0.23	0.17
H_2'		0.28	0.19	0.18	0.19	0.22	0.25	0.29	0.36	0.31
O_3		-0.57	-0.38	-0.34	-0.33	-0.32	-0.53	-0.62	-0.48	-0.35
O_3'		-0.64	-0.46	-0.41	-0.41	-0.40	-0.52	-0.59	-0.46	-0.36

[a] The values in parenthesis refer to the free Gly-Phe-Leu peptide.

gen transfer from O_2 to O_3' is difficult to imagine. Furthermore, no transition state has been located corresponding to the concerted hydrogen transfers H_2' on O_3' and H_2 on N_1. On the other hand, the H_2-N_1 is also long (2.87 Å) but, since the lone pair of the nitrogen points in the opposite direction of the H_2 hydrogen atom, a inversion of this group at the nitrogen atom is conceivable, leading to a shorter H_2-N_1 distance and to a possible transfer (see Figure 4).

TS3 Transition State This transition state, 19.70 kcal·mol^{-1} higher in energy than I3, corresponds to an hydrogen transfer from Glu143 on the nitrogen atom of the peptidic bond to be hydrolyzed. This hydrogen is the one which has been previously transferred from the water molecule on Glu-143 in the I1 → I2 step of the reaction. This transfer is more symmetrical than the transfer which occurs in the TS1 transition state: the O_3-H_2 and H_2-N_1 bond-lengths are comparable. Some important changes can be reported in the bond-lengths between the C_1 carbon atom and the atoms to which it is bonded. While C_1-N_1 bond-length increases, C_1-O_1 and C_1-O_2 bond-lengths both decrease. These evolutions are followed by a drastic decrease of the H_2'-O_3' interatomic distance. The path going from I3 to TS3 involves an inversion at the N_1 atom in order to make the H_2-N_1 hydrogen bond possible. A systematic analysis of this path shows that no stable intermediate with the proper configuration at the N_1 atom can be found, so that the I3 → I4 step is a one step mechanism.

I4 Intermediate In this structure, the C_1-N_1 is elongated but cannot be considered to be broken. Furthermore, two hydrogen bonds between the two oxygen atoms of the carboxylic group of the Glu-143 residue and the two H_2 and H_2' hydrogen atoms can be reported.

TS4 transition state The hydrogen transfer between the O_2 oxygen atom and the O_3' atom of the Glu-143 carboxylic group is coupled with the complete breaking of the C_1-N_1 peptidic bond. Here again, during the transfer, the O_2-H_2' and O_3'-H_2' bond-lengths are nearly the same. The two O_1 and O_2 oxygen atoms tend to be equivalent with respect to the C_1 and Zn atoms and the difference between O_1 and O_2 can be explained by the O_2-H_2' interaction which makes the Zn-O_2 and C_1-O_2 slightly longer. The carboxylic group $O_1C_1O_2$ is now almost formed.

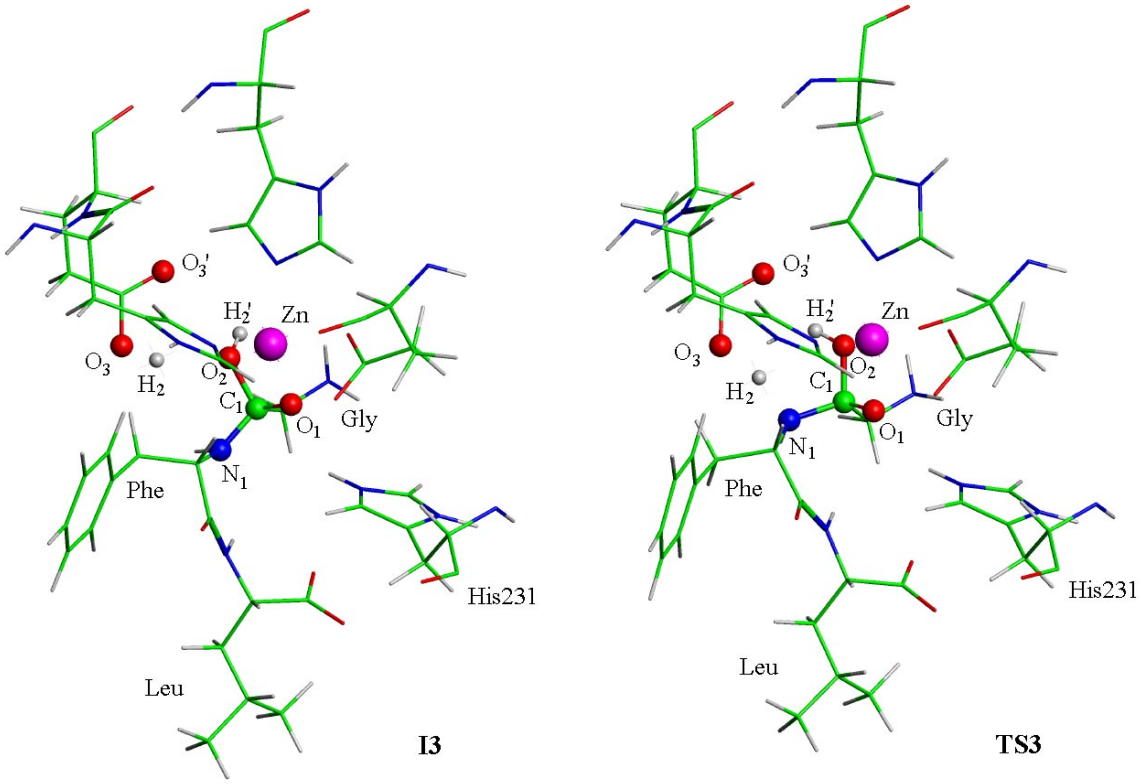


Figure 4 *Structures of the most stabilized intermediate I3 and of the following transition state TS3. The quantum part and the His-231 residue, described classically, are presented*

I5 Intermediate If one limits oneself to the hydrolysis of the Gly-Phe peptidic bond, this intermediate corresponds to the products (see Figure 3) but if the whole catalytic cycle involving the release of the separated Gly and Phe-Leu peptides, this stationary point is only a reaction intermediate. Nevertheless, the Gly amino acid and Phe-Leu peptide are well characterized. Indeed, the C_1-N_1 bond is now broken: 3.036 Å. Additionally, by inspecting the OC and OZn bond-lengths, the O_2 and O_1 oxygen atoms can now be considered as equivalent. The slight differences may be explained by the weak interaction which remains between the H_2' transferred atom and the O_2 oxygen atom of the carboxylic group of the forming Gly amino-acid. Five atoms are coordinated to the metal center since the Gly peptide is linked to the metal in a bidentate way. Furthermore, the C-terminal carbon atom of the forming Gly peptide presents a sp2 configuration. Finally, the drastic energy decrease between the TS4 and the I5 stationary points, -45.48 kcal·mol^{-1}, has to be noted. This diminution allows us to conclude that the reverse reaction is not feasible.

Termination of the reaction and regeneration of the enzyme

As mentioned above, the I5 intermediate cannot be considered as the final product of the complete biological process. The possibilities that the release of the products (Gly and Phe-Leu peptides in our case) occurs through a process involving a motion of the whole enzyme has been discussed.[48] The authors of this study mentioned the possible hinge-bending motion between the two lobes of thermolysin around the 137-150 and 160-180 helices since they form a cross centered at the active site. Simulation of such a dynamic process remains too expensive when using a QM/MM method and will not be envisaged here. However, our computations have shown that the intermediate I5 may undergo a variety of transformations, basically through proton transfer processes, for which the products could also be obtained. Below, such structures are described (see Figure 5).

I6 Intermediate This intermediate is expected to be the result of a transformation of I5 through an hypothetical transition state TS5, which is not considered here. In this structure, the H_2' hydrogen atom which was present on the O_3' oxygen atom of the Glu-143 carboxylic group is now bonded to the O_3 oxygen atom, on the side of the N-terminal nitrogen atom of the Phe-Leu peptide. Another important difference in the structure by comparison with I5 lies on the orientation of the lone pair of the N_1 nitrogen atom. In I6, this lone pair points towards the Glu-143 residue and therefore a hydrogen bond exists while, in the I5 intermediate, it points towards the opposite direction permitting then a hydrogen

Table2 *Energetic contributions (in kcal·mol^{-1}) along the reaction path to the total energy of several subsystems described classically: single residues, whole thermolysin, water molecules of crystalisation and total contributions. See Figure 2 and text for notations*

	I0	I1	TS1	I2	TS2	I3	TS3	I4	TS4	I5
Residues participating to H-bonds[a]										
Asn-112	-5.4	-5.3	-5.2	-5.5	-5.5	-5.5	-7.3	-7.3	-7.8	-7.6
Ala-113	-2.6	-2.6	-0.8	-0.3	0.1	0.6	1.6	-2.8	0.8	0.9
Tyr-157	0.9	0.4	-0.3	-0.3	0.1	0.5	1.7	1.3	1.2	0.3
Arg-203	-52.0	-52.2	-55.2	-55.8	-57.3	-58.0	-55.9	-51.7	-55.4	-57.8
His-231	-92.4	-84.8	-90.7	-91.5	-98.1	-102.4	-96.0	-90.3	-97.9	-99.8
Residues participating to the S1' hydrophobic subsite[b]										
Phe-130	-1.7	-1.7	-1.9	-1.8	-1.8	-1.8	-1.3	-1.5	-1.4	-1.5
Val-139	-5.1	-5.1	-5.2	-6.1	-6.6	-6.9	-5.5	-4.9	-5.8	-6.4
Ile-188	0.4	0.5	1.1	0.9	1.2	1.1	2.4	2.2	2.6	2.0
Leu-202	-1.2	-1.2	-1.7	-1.7	-1.7	-1.7	-1.7	-1.7	-1.8	-1.6
Contributions of Subsystems										
TLN[c]	3.1	6.5	1.5	0.6	-7.7	-13.5	-1.5	3.0	-4.7	-12.0
H_2O[d]	5.5	5.2	6.4	6.6	7.5	8.1	7.4	5.9	7.5	7.8
Total	8.6	11.7	8.0	7.2	-0.2	-5.4	5.9	8.9	2.8	-4.1

[a] Amino acids interacting through H-bond with the substrate

[b] Amino acids interacting through hydrophobic interaction with the substrate at the S1' subsite

[c] All atoms of thermolysin excepted those included in the quantum part

[d] Water molecules of crystalisation

bond between N_1 and H_2, the hydrogen transferred in the previous step. No transition state has been characterized for the process leading from I5 to I6. Nevertheless, the hydrogen transfer between the two oxygen atoms of a carboxylic group as well as the inversion of the amino group are known to be energetically easy. Therefore, reaching this structure from I5 is undoubtly feasible if one considers the high energy activation of the previous processes. This structure is energetically and structurally close to I5 if one excludes the two most important changes, mentioned above.

TS6 Transition State The proton transfer from the Glu-143 carboxylic group to the lone pair of the terminal nitrogen atom of the Phe-Leu peptide constitutes the most important

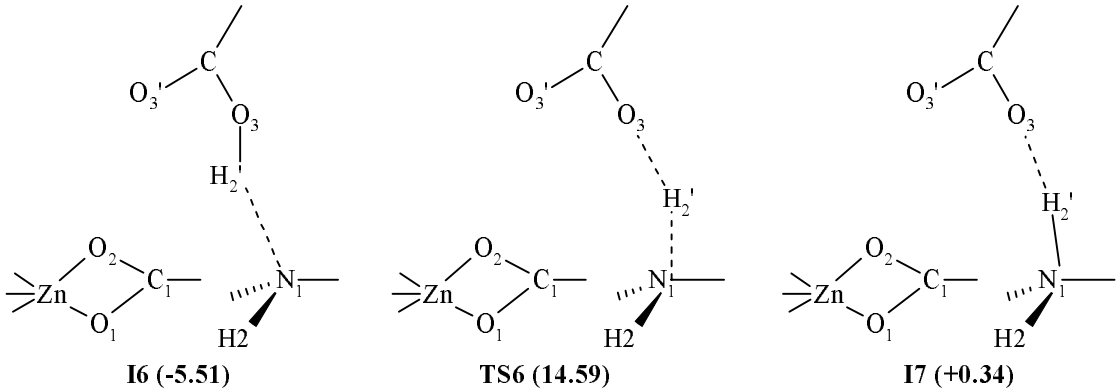


Figure 5 *Schematic description of the reactive part for the stationary points of the termination of the reaction. Energies (in kcal·mol^{-1}) with respect to intermediate I1 are given in parenthesis*

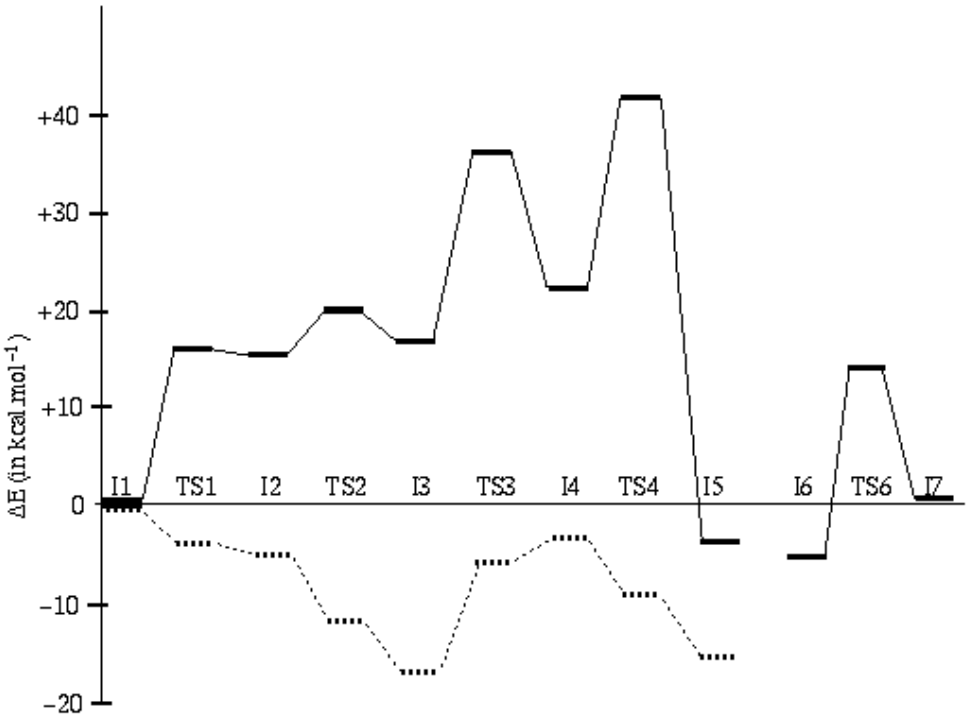


Figure 6 *Energy profile of the hydrolysis mechanism (in kcal·mol^{-1}). Total energy (plain lines) and energy contributions of the surroundings (hatched lines) are given with respect to the initial complex I1*

contribution to the reaction coordinate. Here again, this transfer is rather symmetrical and the energy barrier, 20.10 kcal·mol^{-1} is almost the same that the previous one (TS4).

I7 Intermediate At this point, the hydrogen atom is completely transferred and the Phe-Leu product of the hydrolysis is a peptide charged at its C- and N-terminal parts, which is the common way to consider a peptide in a solution or in an enzyme. The Glu-143 carboxylic group is regenerated but an H-bond still exists between the transferred hydrogen atom and the O_3 atom of the Glu-143 carboxylic group.

Products As shown in the intermediates I5, I6 and I7, the Gly amino acid (considered as a product of the hydrolysis) is doubly ligated to the metal, whereas the Phe-Leu peptide interacts with the enzyme through electrostatic interactions only. The final product could involve the separated Gly-enzyme complex and the Phe-Leu peptide. The latter may represent two forms, a non-protonated one, as found in I5 and a protonated one as found in I7. In both cases, the process appears to be energetically unfavorable in vacuo and, solvation effects, which are not taken into account here, are expected to play a major role. Therefore, this step will not be discussed further.

Discussion

The peptide hydrolysis mechanism (steps I1 → I5) presented here shows several differences with the scheme proposed by Matthews and exhibits, for example, more steps, including transition states and local minima. In order to analyze the importance of the surroundings, for each of these stationary points, the contribution of the enzyme to the total energy has been computed (see Table 2 and Figure 6). This contribution corresponds to the pure $E_{QM/MM}$ term, which can be extracted easily from the total energy. Furthermore, the contribution of individual residues may be extracted by $\langle\Psi|H_i^{QM/MM}|\Psi\rangle$ where Ψ is the polarized wavefunction and $H_i^{QM/MM}$ represent the perturbation due to the residue *i* (note that this term includes in part the non additive contributions to the energy). Considering the total contribution, one should keep in mind that the rather surprising positive values in the interactions energies come from the fact that both the quantum and the classical subsystems are negatively charged. Similarly, the introduction of the water molecule (I0 to I1) induces an increase of energy, which may come from the fact that the system has been fully optimized in the absence of this molecule. However, the chemically interesting energy changes occur in the rest of the process.

The first interesting feature concerns the nucleophilic attack of the carbonyl group by the oxygen atom of the water. This reaction, which is usually considered as a concerted mechanism, appears as a two step reaction in which a proton transfer to the glutamate group (I2) occurs previously to the nucleophilic attack by the resulting OH^- group. Nevertheless, the well defined energy minimum corresponding to I2 is only 0.5 kcal·mol^{-1} below the transition state and this quantity is smaller than the error bars of the method. If one analyzes the energy of interaction between the quantum and the classical parts of the system, one notices that this quantity de-

creases by 0.7 kcal·mol^{-1} from TS1 to I2. This means that the hypothetical intermediate I2 would appear more stable if one optimizes the classical structure at this step. Finally, from a structural point of view one notices that the O_2-C_1 distance does not decreases much while the distance O_2-Zn and H_2-O_3 decrease strongly when passing from I1 to I2. This remark together with the fact that the proton transfer occurs on O_3 only support the idea that the intermediate I3 is formed from I1 by two successive steps.

From a kinetic point of view, one must point out that the rather low energy barrier that separates I2 from I3 at TS2 is mainly due to the presence of the enzyme since it appears (Table 2) that the classical-quantum interaction energy decreases by 7.4 kcal·mol^{-1} from I2 to TS2. The intermediate state I3 again is strongly stabilized by the presence of the surroundings. By considering the individual contributions of the most important amino acids, as shown in Table 2, it appears that His-231 plays a major role in this stabilization. This finding confirms some experimental results indicating that His-231 is strongly involved in the lowering of the energy barriers.[14-16] Finally, TS2, in which the Zn atom is pentaligated appears to be closer to the structure found in the TLN-hydroxamic inhibitor complex [49] than in the TLN-phosphoramidon one,[50] in which the Zn is tetraligated.

The subsequent breaking of the peptidic bond again appears as a multistep reaction. The protonation of the peptidic nitrogen requires a nitrogen inversion process mentioned previously (TS3). The resulting intermediate I4 still shows a rather short C_1-N_1 bond (1.58 Å). The C_1-N_1 bond breaking occurs in the next step and is accompanied by the transfer of proton H_2' from O_2 to O_3'. The final structure I5 is characterized by a long C_1-N_1 distance (3.0 Å) and a planar $C\alpha C_1O_1O_2$ carboxy group of the Gly residue. The same general features have been found in the case of the catalyzed hydrolysis of formamide [34-36] in which a second water molecule plays the role of a proton relay.

From an energetic point of view, the most striking feature is the strong positive variation of the classical quantum interaction energy between I3 and TS3. It is clear from Table 2 that, again, the role of His-231 is crucial in this interaction. By analyzing the interatomic distances it appears that in I3, the O_1 oxygen atom of the substrate is hydrogen bonded with Hε of His-231 (2.3 Å) and that N_1 is strongly interacting with this hydrogen (3.0 Å) especially if one takes into account the fact that the lone pair of this nitrogen atom is pointing towards Hε. In TS3, the O_1-Hε distance does not vary significantly but the N_1-Hε distance increases (4.0 Å) and the nitrogen atom has undergone an inversion. It is highly probable that a full optimization of the whole enzyme-substrate structure would lower the energy of TS3 and, subsequently of the following stationary points.

The strong energy decrease between TS4 and I5 (-45 kcal·mol^{-1}) is in favor of a non-reversible reaction and the important processes in the overall reaction kinetics occur during the sequence I1 → I5.

During this discussion we focused our attention on the contribution of the whole classical subsystem or on His-231 only. A more detailed analysis may consider the contributions of other residues that are often mentioned as being important. Contributions of these residues are detailed in Table 2 and distinction has been made between the residues able to establish a H-bond with the substrate and the hydrophobic ones. Similarly, the contribution of the crystallization water molecules has been separated from the rest of the classical part. One notices that the interaction energy of these water molecules varies in a range of 2.9 kcal·mol^{-1} although the variations of the interactions with the rest of the enzyme reach 20 kcal·mol^{-1}. The interaction with the hydrophobic residues do not vary much. The strongest variations, 2.2 kcal·mol^{-1} are found with Ile-188 which has a positive interaction energy with the quantum subsystem. Apart from His-231, the strongest interactions are observed with Arg-203 and Asn-112. Ala-113 and Tyr-157, which have been proposed as transition state stabilizers, do not appear as important partners in the process of interest.

Conclusion

Hybrid QM/MM computations appears as a useful tool to analyze enzymatic reactions processes. The necessity of using semiempirical quantum chemical methods, which generally exaggerate the energies of the transition states, prevents us from deriving quantitative conclusions. Nevertheless, the comparison of the various energies, and the determination of the stationary points along the reaction paths allows us to derive some useful information. In the present case, our findings are in general good agreement with the mechanism proposed by Matthews, but they show that some postulated global steps require a further analysis which is difficult to perform on the basis of experimental results only, when they come out quite clearly from the calculations. In addition, the structures found for the transition states are very important data. For instance, they can be directly used for the design of specific inhibitors that mimic these transition states. This study shows that they can be used to propose a realistic mechanism when there are not enough experimental data to conclude or to perform a series of fast, cheap computer experiments in order to select the best experimental strategy in studying reaction mechanisms. Therefore, modeling reactions by means of QM/MM methods is becoming a very important tool for biochemists.

Acknowledgements The authors thank Pr. B.P. Roques (U266 INSERM, URA D1500 CNRS) and the members of his metalloproteases research group as well as Dr. B. Maigret for fruitful discussions. We acknowledge the support of the Centre Charles Hermite in Nancy for the provision of computer time on the SG Power Challenge machines and for the kind help of their technical staff.

References

1. Lipscomb, W. N.; Sträter, N. *Chem. Rev.* **1996**, *96*, 2375-2433.
2. Matthews, B. W.; Colman, P. M.; Jansonius, J. N.; Titani, K.; Walsh, K. A.; Neurath, H. *Nature, New Biology* **1972**, *238*, 41-43.
3. Matthews, B. W.; Jansonius, J. N.; Colman, P. M.; Schoenborn, B. P.; Dupourque, D. *Nature, New Biology.* **1972**, *238*, 37-41.
4. Thayer, M. M.; Flaherty, K. M.; McKay, D. B. *J. Biol. Chem.* **1991**, *266*, 2864-2871.
5. Pauptit, R. A.; Karlsson, R.; Picot, D.; Jenkins, J. A.; Niklaus-Reimer, A. S.; Jansonius, J. N. *J. Mol. Biol.* **1988**, *199*, 525-537.
6. Stark, W.; Pauptit, R. A.; Wilson K. S.; Jansonius, J. N. *Eur. J. Biochem.* **1992**, *207*, 781-791.
7. Feder, J.; Garret L. R.; Wildi, B. S. *Biochemistry* **1971**, *10*, 4552-4556.
8. Marie-Claire, C.; Ruffet, E.; Antonczak, S.; Beaumont, A.; O'Donohue, M.; Roques, B. P.; Fournié-Zaluski, M. C. *Biochemistry* **1997**, *36*, 13938-13945.
9. Hersh, L. B.; Morihara, K. *J. Biol. Chem.* **1986**, *261*, 6433-6437.
10. Benchetrit, T.; Fournié -Zaluski, M. C.; Roques, B. P. *Biochem. Biophys. Res. Commun.* **1987**, *147*, 1034-1040.
11. O'Donohue, M. J.; Roques, B. P.; Beaumont, A. *Biochem. J.* **1994**, *300*, 599-603.
12. Pozsgay, M.; Michaud, C.; Liebman, M.; Orlowski, M. *Biochemistry* **1986**, *25*, 1292-1289.
13. Roques, B. P.; Noble, F.; Daugé, V.; Fournié-Zaluski, M. C.; Beaumont, A. *Pharmacol. Rev.* **1993**, *45*, 87-146.
14. Matthews, B. W. *Acc. Chem. Res.* **1988**, *21*, 333.
15. Mock, W. L.; Aksamawti, M. *Biochem. J.* **1994**, *302*, 57-68.
16. Mock, W. L.; Standford, D. *Biochemistry* **1996**, *35*, 7369-7377.
17. Holland, D. R.; Hausrath, A. C.; Juers, D.; Matthews, B. W. *Protein Sci.* **1995**, *4*, 1955-1965.
18. Warshel, A.; Lewitt, M. *J. Mol. Biol.* **1976**, *103*, 227-249.
19. Åqvist, J.; Warshel, A. *Chem. Rev.* **1993**, *93*, 2523-2555.
20. Cunningham, M. A.; Ho, L. L.; Nguyen, D. T.; Gillilian, R. E.; Bash, P.A. *Biochemistry* **1997**, *36*, 4800-4816.
21. Ranganathan, S.; Gready, J. E. *J. Phys. Chem. B.* **1997**, *101*, 5614-5618.
22. Liu, H.; Müller-Plathe, F.; van Gunsteren, W. F. *J. Mol. Biol.* **1996**, *261*, 454-469.
23. Lyne, P. D.; Mulholland, A. D.; Richards, W. G. *J. Am. Chem. Soc.* **1995**, *117*, 11345-11350.
24. Arad, D.; Langridge, R.; Kollman, P. A. *J. Am. Chem. Soc.* **1990**, *112*, 491-502.
25. Ryde, U. *Eur. Biophys. J.* **1996**, *24*, 213-221.
26. Ryde, U. *J. Comput-Aided Molec. Design.* **1996**, *10*, 153-164.
27. Ryde, U. *Prot. Sci.* **1995**, *4*, 1124-1132.
28. Tuñón, I.; Martins-Costa, M. T. C.; Millot, C.; Ruiz-López, M. F. *J. Mol. Mod.* **1995**, *1*, 196-201.
29. Tuñon, I.; Martins-Costa, M. T. C.; Millot, C.; Ruiz-López, M. F. *J. Chem. Phys.* **1997**, *106*, 3633-3642.
30. Strnad, M.; Martins-Costa, M. T. C.; Millot, C.; Tuñón, I.; Ruiz-López, M. F.; Rivail, J.-L. *J. Chem. Phys.* **1997**, *106*, 3643-3647.
31. Assfeld, X.; Rivail, J.-L. *Chem. Phys. Lett.* **1996**, *263*, 100-106.
32. Théry, V.; Rinaldi, D.; Rivail, J.-L.; Maigret, B.; Ferenczy, G. G. *J. Comp. Chem.* **1994**, *15*, 269-282.
33. Monard, G.; Loos, M.; Théry, V.; Baka, K.; Rivail, J.-L. *Int. J. Quantum Chem.* **1996**, *58*, 153-159.
34. Antonczak, S.; Ruiz-López, M. F.; Rivail, J.-L. *J. Am. Chem. Soc.* **1994**, *116*, 3912-3921.
35. Antonczak, S.; Ruiz-López, M. F.; Rivail, J.-L. *J. Mol. Model.* **1997**, *3*, 434-442.
36. Antonczak, S.; Monard, G.; Ruiz-López, M. F.; Rivail, J.-L. *J. Am. Chem. Soc.* **1998**, *120*, 8825-8833.
37. Reuter, N.; Loos, M.; Monard, G.; Cartier, A.; Maigret, B.; Rivail, J.-L. *Mol. Eng.* **1997**, *7*, 349-365.
38. Rinaldi, D.; Hoggan, P.E.; Cartier, A.; Baka, K.; Monard, G.; Loos, M.; Mokrane, A.; Dillet, V.; Théry, V. *GEOMOP* to be published.
39. Stewart, J. J. P. *J. Comp. Chem.* **1989**, *10*, 209-221.
40. Merz, Jr., K.M.; Dewar, M. J. S. *Organometallics* **1990**, *7*, 522-524.
41. Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Comput. Chem.* **1986**, *7*, 230-252.
42. Jacob, O.; Cardenas, R.; Tapia, O. *J. Am. Chem. Soc.* **1990**, *112*, 8692-8705.
43. Discover, Molecular Simulation Inc., 9685 Scranton Road, San Diego, CA 92121-3752 USA.
44. Schlegel, H. B. *J. Comput. Chem.* **1982**, *3*, 214-218.
45. Gonzales, C.; Schlegel, H.B. *J. Phys. Chem.* **1990**, *94*, 5523-.
46. Beapark, M. J.; Robb, M.A.; Schlegel, H.B. *Chem. Phys. Lett.* **1994**, *223*, 269-272.
47. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. *Nucleic Acids Research* **2000**, *28*, 235-242.
48. Holland, D. R.; Tronrud, D. E.; Pley, H. W.; Flaherty, K. M.; Stark, W.; Jansonius, J. N.; McKay, D. B.; Matthews, B. W. *Biochemistry* **1992**, *31*, 11310-11316.
49. Holmes, M. A.; Matthews, B. W. *Biochemistry* **1981**, *20*, 6912-6920.
50. Weaver, L. H.; Kester, W.R.; Matthews, B. W. *J. Mol. Biol.* **1977**, *114*, 119-132.