Catalytic Mechanism of the Enzyme Papain: Predictions with a Hybrid Quantum Mechanical/Molecular Mechanical Potential

Martin J. Harrison, Neil A. Burton, and Ian H. Hillier*

Contribution from the Department of Chemistry, University of Manchester, Oxford Road, Manchester, M13 9PL United Kingdom

Received April 10, 1997. Revised Manuscript Received September 17, 1997[⊗]

Abstract: A hybrid quantum mechanical/molecular mechanical approach is used to elucidate structural and energetic features of amide hydrolysis by the enzyme papain. The role of the enzyme in stabilizing the thiolate−imidazolium ion pair is examined and the potential energy pathway for the subsequent attack of the cysteine anion and proton transfer from the imidazolium cation is determined. The reaction is found to be concerted rather than stepwise, and the transition state for the reaction is located. The effect of residue mutations both on the ion pair stability and on the barrier to amide hydrolysis is explored and found to be in agreement with experiment. In this work both high-level electronic structure and semiempirical MO methods are used, with location and characterization of stationary structures. Rearrangement of the enzyme in response to the changing electronic structure of the active site is also considered.

Introduction

Cysteine proteases are a class of enzymes that have been widely studied over the years and have been shown to play important roles in degradation of muscular protein[1−3] and in immunopharmacological response.[4,5] Defects in the regulation of cysteine protease activity have also been reported in connection with several disease processes, including cancer[6] and muscular dystrophy.[7] The overall principles of substrate recognition, catalysis, and inhibition are now reasonably well documented. However, the mechanism by which cysteine proteases hydrolyze their substrates is still poorly defined at the atomic level so that the relationship between the molecular structure of the enzyme and of the substrate to the rate of the catalytic reaction is not clear. It is now widely accepted that the activity of cysteine proteases is dependent on the establishment of an active site ion pair[8−10] between cysteine and histidine residues (Figure 1). Many other residues both in and around the active site are considered to be of importance for the proper functioning of the enzyme based on comparison of amino acid sequences and available X-ray structures [see ref 11 for a review].

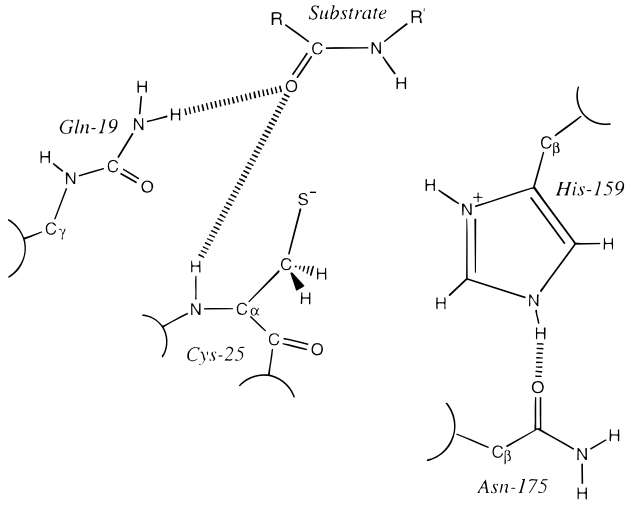

Figure 1. Cysteine protease active site geometry showing the cysteine−histidine ion pair. The C═O of the scissile peptide bond of the substrate is directed toward an oxyanion hole formed from the main chain NH of cysteine and the NH group of the amide side chain of glutamine. [The papain sequence numbering is used.]

The majority of the studies on cysteine proteases describe results obtained with the enzyme papain, since it is considered to be the archetype of this enzyme class. Papain is a globular protein consisting of a single polypeptide chain of 212 residues, folded to form two domains with a deep cleft between them,[12,13] The active site cysteine residue (Cys-25) is part of the L1 α-helix at the surface of the left domain, while the histidine (His-159) is in a β sheet at the surface of the right domain of the enzyme. The existence of these two ionizable groups as essential catalytic residues in papain is consistent with the pH dependency of activity in this enzyme. The acid limb, with a pK_a of 4, is usually attributed to the ionization of Cys-25, whereas the basic limb, with a pK_a of 8.5, is considered to reflect ionization of

[⊗] Abstract published in *Advance ACS Abstracts,* November 15, 1997.

(1) Komatsu, K.; Tsukuda, K.; Hosoya, J.; Satoh, S. *Exp. Neurol.* **1986**, *93*, 642−646.
(2) Rabbini, N.; Moses, L.; Anandaraj, M. P. J. S. *Biochem. Med. Metabol. Biol.* **1987**, *37*, 282−286.
(3) Gopalan, P.; Dufresne, M. J.; Warner, A. H. *Can. J. Physiol. Pharmacol.* **1984**, *65*, 124−129.
(4) Amamoto, T.; Okazaki, T.; Komurasaki, T.; Hanada, K.; Omura, S. *Microbiol. Immunol.* **1984**, *28*, 85−97.
(5) Amamoto, T.; Okazaki, T.; Komurasaki, T.; Oguma, K.; Tamai, M.; Hanada, K.; Omura, S. *Jpn. J. Pharmacol.* **1984**, *34*, 335−342.
(6) Lah, T. T.; Clifford, J. L.; Helmer, K. M.; Day, N. A.; Moin, K.; Honn, K. V.; Crissman, J. D.; Sloane, B. F. *Biochem. Biophys. Acta* **1989**, *993*, 63−73.
(7) Gopalan, P.; Dufresne, M. J.; Warner, A. H. *Biochem. Cell Biol.* **1986**, *64*, 1010−1019.
(8) Polgár, L. *FEBS Lett.* **1974**, *47*, 15−18.
(9) Lewis, S. D.; Johnson, F. A.; Shafer, J. A. *Biochemistry* **1976**, *82*, 5009−5017.
(10) Lewis, S. D.; Johnson, F. A.; Shafer, J. A. *Biochemistry* **1981**, *20*, 48−51.
(11) Storer, A. C.; Ménard, R. *Methods Enzymol.* **1994**, *244*, 486−500.
(12) Drenth, J.; Kalk, K. H.; Swen, H. M. *Biochemistry* **1976**, *15*, 3731−3738.
(13) Kamphuis, I. G.; Kalk, K. H.; Swarte, M. B. A.; Drenth, J. *J. Mol. Biol.* **1984**, *17*, 233−256.

S0002-7863(97)01147-5 CCC: $14.00

His-159. Generally, the pK_a values of the cysteine and histidine groups in proteins are about 8 to 9, and 6, respectively.

It might be expected that cysteine proteases have a similar mechanism to the serine proteases[14] since the nature and orientation of the catalytic groups are similar.[15] As the nucleophile is already present in cysteine proteases they may also appear to be more efficient. However, the same lower basicity that allows cysteine to be ionic in the resting state also makes it a poorer nucleophile compared to the neutral active site serine in the serine proteases. Howard and Kollman[16] have studied theoretically the nucleophilic attack of OH^- and SH^- on formamide to provide insight into the mechanism of hydrolysis by serine and cysteine proteases, respectively. They found that in the case of OH^-/formamide there is no barrier to the formation of a tetrahedral adduct, while no stable SH^-/formamide adduct is formed. Thus two key questions emerge concerning the amide hydrolysis step in the cysteine proteases: Does the reaction involve a tetrahedral intermediate and if so how is it stabilized? A possible mechanism involves histidine first protonating the substrate at which ionized cysteine attack is favorable, or alternatively[17] such attack is concerted with proton transfer, enabling a stable tetrahedral adduct to be formed.

The importance of particular residues in stabilizing the initial ion pair and in aiding the subsequent hydrolysis reaction has been studied experimentally by site-directed mutagenesis. In particular, the role of Asn-175 and Gln-19 in affecting ion-pair stability and amide hydrolysis has been investigated. The asparagine residue (Asn-175) that is adjacent to the catalytic His-159 residue is conserved in all cysteine protease sequences of the papain family. The amide oxygen of the Asn-175 side chain is hydrogen bonded to $N_{\epsilon 2}H$ of His-159, and this bond is approximately colinear with the His-159 $C_\beta-C_\gamma$ bond, allowing rotation of the imidazole ring about the $C_\beta-C_\gamma$ bond without disruption of the hydrogen bond. Since the side chains of Asn-175 and His-159 interact directly via hydrogen bonding, one of the obvious roles of Asn-175 is to stabilize the thiolate−imidazolium form of papain. It has been suggested that the proximity of the active site cysteine and histidine residues could be one of the most important factors contributing to the formation of an ion pair and that the proton affinities of Cys-25 and His-159 at the active site of papain are highly sensitive to the geometry of these residues.[18,19] Consequently, Asn-175 could stabilize the ion pair by keeping the imidazole ring of His-159 in a favorable orientation.

The side chain of Gln-19 and the backbone NH of Cys-25 have been proposed[12,20,21] to be of importance to catalysis due to formation of an oxyanion hole that is able to stabilize the transition state or intermediate. This oxyanion hole is similar to that found in serine proteases.[15] However, unlike serine proteases where there is clear support for the existence of an oxyanion hole, the situation for the cysteine proteases is not clear. Probably the most cited work against the existence of an operational oxyanion hole in cysteine proteases is that of Asbóth and Polgár, describing the hydrolysis of oxygen and thiono esters by papain and chymopapain.[22] The replacement of the carbonyl oxygen of the substrate by a sulfur atom significantly decreases the serine protease activity, yet cysteine proteases were shown to hydrolyze both types of substrates at comparable rates.[22−24] However, as pointed out by Storer and Carey,[24] cysteine protease could accommodate the sulfur atom of a thiono ester through an adjustment of both the substrate and the enzyme. The controversy over the putative role of an oxyanion hole in cysteine proteases has recently been clarified with the kinetic characterization of papain variants produced by site-directed mutagenesis.[20] The replacement of Gln-19 by Ala and Ser led to 60- to 600-fold decreases in k_{cat}/K_M, respectively, for hydrolysis of the substrate CBZ-Phe-Arg-MCA, which correspond to changes in free energy of catalysis of 2.4 and 3.8 kcal/mol for Gln19Ala and Gln19Ser, respectively. Since the Gln-19 residue in papain is conserved in all known cysteine proteases sequenced, it is likely that the oxyanion hole plays an important role in cysteine proteases in general.

The catalytic pathway involved in hydrolysis by cysteine proteases has been explored by Arad et al.,[17] who used semiempirical quantum mechanical calculations to study the energetics of active site/substrate structures suggested from molecular mechanical studies. Their calculations suggest that proton transfer from histidine to either oxygen or nitrogen of the substrate occurs prior to, or concerted with, the nucleophilic attack.

Subsequent to these studies, hybrid methods have been developed and used to model enzyme catalysis.[25−29] Here, quantum mechanics (QM) is used to treat the active site and substrate, and molecular mechanics (MM) is used to model the bulk of the enzyme not directly involved in the chemical reaction. Such methods are being increasingly used, employing both semiempirical[25−27] and *ab initio*[28,29] descriptions of the QM part of the system. In this paper we describe the use of such hybrid QM/MM methods to explore a number of aspects of catalysis by the enzyme papain.

Computational Details

The Hybrid QM/MM Method. In the hybrid QM/MM method the effective Hamiltonian is written as

$$\hat{H}_{\text{eff}} = \hat{H}_0 + \hat{H}_{\text{QM/MM}}$$

where $\hat{H}_0$ is the Hamiltonian of the QM fragment and $\hat{H}_{\text{QM/MM}}$ is the interaction between the QM and MM fragments.

$$\hat{H}_{\text{QM/MM}} = -\sum_{\text{iM}} \frac{q_{\text{M}}}{r_{\text{iM}}} + \sum_{\alpha\text{M}} \frac{z_\alpha q_{\text{M}}}{R_{\alpha\text{M}}} + \sum_{\alpha\text{M}} \left\{ \frac{A_{\alpha\text{M}}}{R_{\alpha\text{M}}^{12}} - \frac{B_{\alpha\text{M}}}{R_{\alpha\text{M}}^{6}} \right\} \quad (1)$$

where α and M label QM and MM centers, respectively. The first term on the right-hand side of eq 1 accounts for the effect of the formal charges of the MM fragment (q_M) on the QM part and may readily be incorporated into standard QM codes. The final two terms are the remaining Coulombic and Lennard-Jones interaction terms between the

(14) Daggett, V.; Schröder, S.; Kollman, P. A. *J. Am. Chem. Soc.* **1991**, *113*, 8926−8935.
(15) Garavito, R. M.; Rossmann, M. G.; Argos, P.; Eventoff, W. *Biochemistry* **1977**, *16*, 5065−5071.
(16) Howard, A. E.; Kollman, P. A. *J. Am. Chem. Soc.* **1988**, *110*, 7195−7200.
(17) Arad, D.; Langridge, R.; Kollman, P. A. *J. Am. Chem. Soc.* **1990**, *112*, 491−502.
(18) Dijkman, J. P.; Osman, R.; Weinstein, H. *Int. J. Quantum Chem.* **1989**, *35*, 241−252.
(19) Rullmann, J. A. C.; Bellido, M. N.; Duijnen, P. Th. van *J. Mol. Biol.* **1989**, *206*, 101−118.
(20) Ménard, R.; Carriere, J.; Laflamme, P.; Plouffe, C.; Khouri, H. E.; Vernet, T.; Tessier, D. C.; Thomas, D. Y.; Storer, A. C. *Biochemistry* **1991**, *30*, 8924−8928.
(21) Ménard, R.; Plouffe, C.; Laflamme, P.; Vernet, T.; Tessier, D. C.; Thomas, D. Y.; Storer, A. C. *Biochemistry* **1995**, *34*, 464−471.
(22) Asbóth, B.; Polgár, L. *Biochemistry* **1983**, *22*, 117−122.
(23) Asbóth, B.; Stokum, E.; Khan, I. U.; Polgár, L. *Biochemistry* **1985**, *24*, 606−609.
(24) Storer, A. C.; Carey, P. R. *Biochemistry* **1985**, *24*, 6808−6818.
(25) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227−249.
(26) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comp. Chem.* **1990**, *11*, 700−733.
(27) Aqvist, J.; Warshel, A. *Chem. Rev.* **1993**, *93*, 2523−2544.
(28) Singh, U. C.; Kollman, P. A. *J. Comp. Chem.* **1986**, *7*, 718−730.
(29) Waszkowycz, B.; Hillier, I. H.; Gensmantel, N.; Payling, D. W. *J. Chem. Soc., Perkin Trans. 2* **1991**, 225−231.

QM and MM fragments. The code AMBER[30] and associated force fields are used to carry out the MM calculations, and Gaussian94[31] is used to enable both semiempirical and *ab initio* QM calculations to be carried out. These codes are coupled *via* the Gaussian link structure, thus enabling all the standard facilities within both programs to be efficiently utilized. In particular our implementation of the hybrid QM/MM scheme allows the geometry of the QM fragment to be optimized in the field of the MM portion by using the redundant internal coordinate algorithm, and conversely the geometry of the MM portion can be optimized in the field of the QM fragment. Although such optimizations of the two portions are not directly coupled at present, an iterative optimization procedure does allow for the effective optimization of the combined QM/MM system. The characterization of transition geometries can now be routinely carried out by calculation of the second derivatives of the QM/MM energy. The van der Waal (vdw) and bond contributions to the second derivatives are calculated by finite difference of the vdw gradients whereas the electrostatic terms can be calculated by finite difference or analytically using standard methods. In the optimizations, the positions of several atoms were frozen to speed convergence by increasing the relative mass of these atoms in the Cartesian to internal coordinate conversions. A similar procedure is followed in the calculation of the vibrational frequencies enabling characterization of the constrained stationary state.

The treatment of the covalent linkages between the QM and MM portions is somewhat arbitrary. Hydrogen termination atoms are used to satisfy the valency of the QM portion, placed at a fixed distance from the appropriate QM atom. No interactions between these hydrogen termination atoms and the MM portion is included, and to avoid unrealistic interactions, the charges of the MM junction atoms attached to the QM system are set to zero, with the charges on the remaining MM atoms scaled to conserve total charge.

Construction of the Enzyme−Substrate Model. An initial substrate−enzyme structure was constructed following the approach of Arad et al.,[17] utilizing an experimentally determined inhibitor−enzyme structure together with a structure of the oxidized enzyme. The initial starting structure for the construction of our enzyme−substrate model was the papain−inhibitor complex (6PAD) studied by Drenth et al.[12] The substrate chosen for our work was Ace-Phe-Ser-Ile-Nme in view of experimental binding data for these residues. The initial geometry of the papain−substrate complex was constructed by replacing the inhibitor in structure 6PAD by our chosen substrate and the enzyme of 6PAD by the high-resolution structure 9PAP.[13] The inevitable bad contacts in this initial structure were relieved by energy minimization, and the structure was solvated with 95 TIP3P water molecules.[32] In these MM calculations, using AMBER 4.0, all atoms were explicitly represented by the force field and a nonbonded cutoff of 10 Å was used. For simulating the nonstandard residue Cys^-, the model CH_3S^- was optimized at the HF/4-31G level and the set of electrostatic potential charges obtained was fitted into the side chain of the cysteine anion. These charges were approximately −0.9 for the sulfur and 0.2 for the β carbon. (Charges for His^+ are available in the AMBER database.) In the final structure, the carbonyl to be cleaved points toward the "oxyanion hole" formed by the NH_2 group of Gln-19 and the NH group of the Cys-25 main chain, forming C═O−H distances of 2.01 and 1.87 Å, respectively. The S−C(O) distance is 3.40 Å. Other enzyme−substrate distances are in good agreement with X-ray crystallographic studies.[12] This structure, including the 95 water molecules, was used as the starting structure for the QM/MM calculations.

Hybrid QM/MM Calculations. In the calculations reported here we use the semiempirical AM1 Hamiltonian to describe the QM portion,

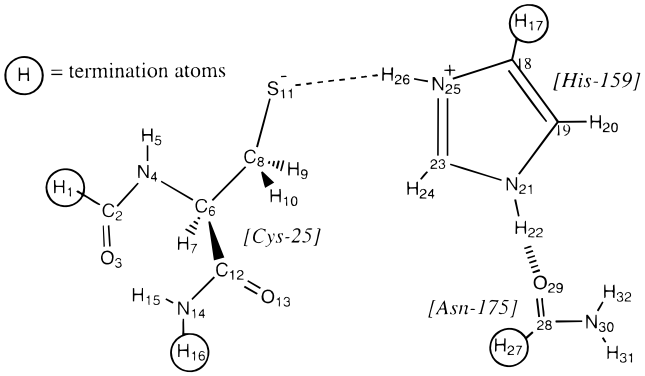

Figure 2. QM atoms involved in ion pair calculations. (The zwitterionic residue form is shown here.)

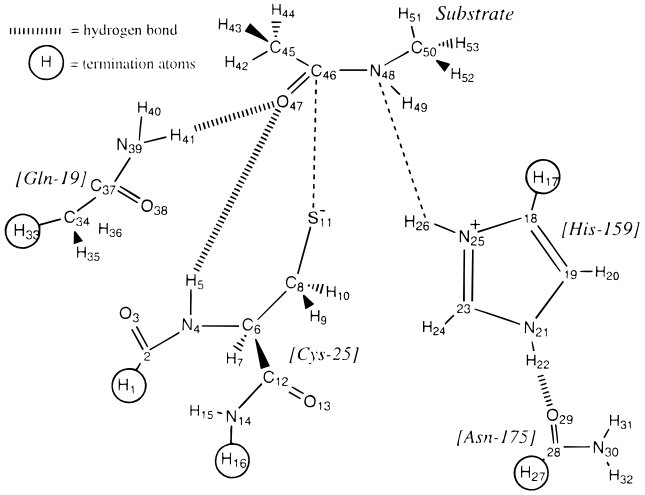

Figure 3. QM region of the substrate/active site.

allowing geometry optimization of substantial fragments to be carried out. (The sulfur AM1 parameters of Gaussian94 were replaced by the more recent values.[33]) To evaluate the performance of AM1 for the reaction to be studied here, we have calculated the proton affinities of the methane thiolate anion and imidazole, and the associated proton transfer reaction, at a number of levels of theory.

For the enzyme−catalyzed reaction, the potential energy surface associated with the thiolate−imidazolium ion pair and with the subsequent amide hydrolysis study was first studied at the QM(AM1)/MM level. The use of a higher level of theory involving a 3−21G* all−electron basis and a B3−LYP exchange correlation functional[34−36] (QM(B3LYP/3−21G*)/MM) was then explored. The QM fragments in the study of the thiolate−imidazolium ion pair and the amide hydrolysis are shown in Figures 2 and 3, respectively. In the former, the complete substrate from the MM optimization was considered. In the amide hydrolysis study, the substrate chosen was *N*-methylacetamide, which was modeled at the QM level. Such a small substrate was used to permit the local geometry changes associated with the chemical reaction to occur, which would be difficult with a larger substrate.

Computational Results

The Thiolate−Imidazolium Ion−Pair. In Table 1 we show the results of calculations involving methanethiol and imidazole. The AM1 result for the proton transfer energy from methanethiol to imidazole is in close agreement with the experimental value,[37] and so no major corrections need be made for this reaction.

(30) Pearlman, D. A.; Case, D. A.; Caldwell, J. C.; Seibel, G. L.; Singh, U. C.; Weiner, P.; Kollman, P. A. *AMBER 4.0*; University of California: San Francisco, 1992).

(31) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T. A.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomberts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez C.; Pople, J. A. *GAUSSIAN94*; Gaussian Inc.: Pittsburgh, PA, 1995.

(32) Jorgensen, W. L.; Chandrasekhar, J.; Madura, D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.

(33) Dewar, M. J. S.; Yuan, Y. C. *Inorg. Chem.* **1990**, *29*, 3881−3890.

(34) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

(35) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev.* **1988**, *B37*, 785−789.

(36) Becke, A. D. *Phys. Rev.* **1988**, *A38*, 3098−3100.

(37) Kollman, P. A.; Hayes, D. M. *J. Am. Chem. Soc.* **1981**, *103*, 2955−2961.

Table 1. Calculated Gas-Phase Proton Affinities of Methane Thiolate Anion and Imidazole, and the Relative Gas-Phase Energies for the Corresponding Proton-Transfer Reaction

	reaction[a]		
method	$CH_3SH \rightarrow CH_3S^- + H^+$	$ImH^+ \rightarrow Im^c + H^+$	$CH_3SH + Im \rightarrow CH_3S^- + ImH^+$
AM1	302.3	169.4	132.9
HF/STO-3G	496	283	213
HF/3-21G*	370.1	248.9	121.2
HF/4-31G	364.4	247.9	116.5
B3LYP/3-21G*	373.1	245.2	127.9
experiment[b]	355.0	220.0	135.0

[a] Energies in kcal/mol. [b] Reference 37. [c] ImH^+ is a protonated imidazolium ring; Im is a neutral imidazole molecule.

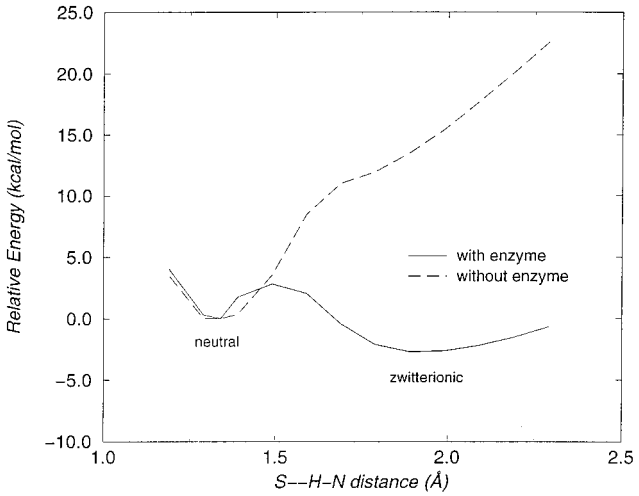

Figure 4. Potential energy curve for proton transfer. (Energy of the QM system at R(S−H(N)) = 1.34 Å was taken as zero.)

Indeed the AM1 result is close to the value at the B3LYP/3-21G* level, and is superior to those from lower level *ab initio* treatments.

We now turn to the results for the thiolate−imidazolium ion pair (Figure 2). The QM atoms of the active site system were defined as the whole of the Cys-25 residue (plus the backbone C_2═O_3 of Ser-24 and the backbone N_{14}−H_{15} of Trp-26), the imidazole ring of His-159, and the side chain of Asn-175 (not including the C_β atoms). This 32-atom QM system, of which 4 are termination atoms, was embedded in the rest of the enzyme modeled via MM, consisting of 3579 atoms. During optimization the positions of a number of peripheral atoms (1, 14, 15, 16, 17, 27, 30, 31, 32) were kept fixed to speed convergence of the geometry optimization. The stability of the thiolate−imidazolium ion pair was initially studied by optimizing the geometry of the QM region for a range of fixed S_{11}−H_{26} distances from 2.29 to 1.19 Å in 0.1-Å steps. Thus a series of 11 structures was generated with AM1, in which all geometric parameters of the QM system, except for the S−H(N) bond length, were fully optimized. The relative energetics are shown in Figure 4. To quantify the effect of the enzyme environment on the potential energy surface, the energies of the optimized QM structure were also calculated in the absence of the MM atoms. In the absence of the enzyme the pair of neutral residues are strongly favored. However, it can be seen that the protein environment has a large effect on proton transfer energetics. The potential energy surface for proton transfer with the protein environment present is extremely flat, the zwitterionic form being predicted to be stabilized over the structure having neutral residues by ≈3 kcal/mol. The barrier to proton transfer is seen to also be quite small at ≈3 kcal/mol. There has been much discussion concerning the position of the reactive sulfur atom of cysteine relative to that of the imidazole ring. X-ray diffraction studies[12] have suggested that the imidazole ring of His-159 is capable of rotating 30° to lie in the same plane as this reactive sulfur atom. Previous calculations[18,19] have suggested that for the neutral cysteine residue, the sulfur does indeed lie in the plane of the imidazole. Our calculations also come to the same conclusion, and indeed after proton transfer to imidazole has occurred, the sulfur atom remains approximately in the plane of the ring, the imidazole ring having rotated only slightly (by ≈2.5°) about the C_β−C_γ bond of His-159 during proton transfer.

Table 2. Energies of the Zwitterionic and Neutral Residue Structures at Different Levels of Theory

	energies of QM optimized structures	
Cys25, His159 residues	AM1[a] (kcal/mol)	B3LYP/3-21G* (au)
zwitterionic	−121.3	−1205.30170
neutral	−118.5	−1205.28864
zwitterionic − neutral[b]	−2.8	−8.2

[a] In all tables, the AM1 energies are heats of formation. [b] Energy difference in kcal/mol.

Table 3. Energies of the Zwitterionic and Neutral Residue Structures at Different Levels of Theory (Here the Enzyme Is Allowed To Relax After Each Cycle of QM Geometry Optimization)

	energies of QM optimized structures (enzyme relaxed)	
Cys25, His159 residues	AM1 (kcal/mol)	B3LYP/3-21G* (au)
zwitterionic	−123.9	−1205.30284
neutral	−122.0	−1205.29010
zwitterionic − neutral[a]	−1.9	−8.0

[a] Energy difference in kcal/mol.

The effect of using a higher level of theory was studied by performing geometry optimizations on the two extreme enzyme structures (zwitterionic and neutral) at the B3LYP/3-21G* level. The S−H(N) bond distance in the neutral residue structure was fixed at 1.34 Å whereas in the zwitterionic structure this bond distance was free to optimize. The energies of the AM1 and B3LYP/3-21G* calculations are given in Table 2. The greater relative stability of the zwitterionic form calculated at the B3LYP/3-21G* level compared to the AM1 level is consistent with the underestimation of the energy of proton transfer between methanethiol and imidazole at the B3LYP/3-21G* level (see Table 1).

Our study of the reactive center of the enzyme was taken further by examining the effect of allowing for the slight structural reorganization of the enzyme during proton transfer. This was accomplished by allowing for some MM relaxation (50 steps) at each cycle of the QM geometry optimization procedure. Such a calculation was carried out for the two extreme structures (neutral and zwitterionic) with the atomic charges appropriate to the two structures. The positions of the termination hydrogen atoms were maintained along the broken QM−MM bonds by keeping the junction atoms stationary during MM optimization. In view of the size of the QM region chosen, we believe that there will be little structural change in the region of the QM−MM junction over the reaction path studied. The results at both AM1 and B3LYP/3-21G* levels are shown in Table 3. These results show that the stabilization of the zwitterionic form relative to the neutral residue form is decreased by between ≈0.9 (AM1) and ≈0.2 kcal/mol (B3LYP/

Table 4. Effect of the Asn-175 Residue on the Stability of the Ion Pair

	energies of QM optimized structures (enzyme relaxed)	
Cys25, His159 residues	AM1 (kcal/mol)	B3LYP/3-21G* (au)
native enzyme		
zwitterionic	−123.9	−1205.30284
neutral	−122.0	−1205.29010
Asn-175Ala mutation		
zwitterionic	−65.0	−1036.31001
neutral	−66.5	−1036.30278
Asn-175 stabilization[a]	3.3	3.5

[a] Energy difference in kcal/mol.

3-21G*) compared to the previous calculations where the enzyme remains fixed. This result was not unexpected since the active site model used in the initial calculations to follow the proton transfer was constructed from the minimized zwitterionic enzyme−substrate complex, leading to a small bias toward the zwitterionic form.

Turning now to the effect of the enzyme structure on the proton transfer energetics, the AM1 energetics of the QM structures along the proton transfer pathway have been evaluated in the absence of interaction with the MM system (see Figure 3). The removal of the rest of the enzyme causes the zwitterionic form to be destabilized with respect to the neutral residue form by ≈16 kcal/mol, showing the essential role of the enzyme in achieving proton transfer. B3LYP/3-21G* single point calculations without the protein present also cause the zwitterionic form to be destabilized with respect to the neutral form by ≈16 kcal/mol.

The role of an individual residue, Asn-175, in stabilizing the zwitterionic form has been studied experimentally by using site directed mutagenesis,[38] replacing the Asn−175 by alanine (Asn−175Ala). If we assume that the mutation has no effect on the intrinsic activity of the catalytic reaction, the equilibrium constant involving the neutral and zwitterionic residue forms is smaller (by a factor of 735) than that of wild-type papain. This reduction in the equilibrium constant accompanying such a mutation leads to the conclusion that the zwitterionic form has been destabilized by 3.9 kcal/mol. This is an upper limit to the ion pair destabilization, the lower limit being zero if we assume the mutation has no effect on the ion pair. We have modeled this effect by repeating our calculation on the zwitterionic and neutral residue topology structures, allowing the enzyme to relax at each cycle of the QM geometry optimization. New enzyme structures were used in which replacement of Asn-175 by alanine has been carried out. The QM system now consists only of the Cys-25 and His-159 fragments, the Ala-175 residue being part of the MM region. The AM1 and B3LYP/3-21G* results described in Table 4 show that this mutation results in a relative destabilization of the zwitterionic form of 3−4 kcal/mol. This agrees closely with experiment and confirms that this mutation mainly effects the stability of the ion pair and not the intrinsic activity.

In both the AM1 and B3LYP/3-21G* results the reactive sulfur atom no longer lies in the plane of the imidazole ring when the thiol group is reduced to −SH, as is found by our calculations for the case of the natural enzyme. It is therefore likely that the stabilization due to the Asn-175 residue comes from keeping the imidazole ring of His-159 in a favorable orientation, with the imidazole ring and the reactive sulfur atom lying in the same plane.

(38) Vernet, T.; Tessier, D. C.; Chatellier, J.; Plouffe, C.; Lee, T. S.; Thomas, D. Y.; Storer, A. C.; Ménard, R. *J. Biol. Chem.*, **1995**, *270*, 16645−16652.

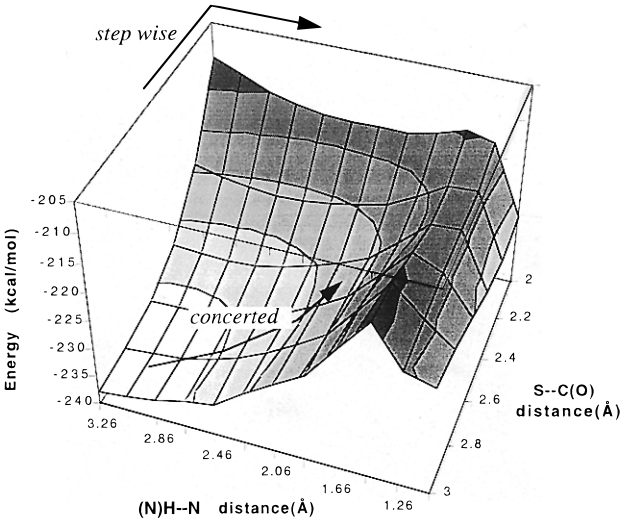

Figure 5. Potential energy surface for amide hydrolysis.

The Amide Hydrolysis Step. The next reaction step to be studied is hydrolysis of the substrate amide and the associated cleavage of the peptide bond, the QM region of our model for this being shown in Figure 3. To study the reaction, the substrate Ace-Phe-Ser-Ile-Nme was replaced by *N*-methylacetamide. The whole of this substrate was included as part of the QM region, with no termination atoms needed between the substrate and the MM environment. The QM region also included the QM atoms previously used in the proton transfer reaction (Figure 2) plus the side chain of Gln-19. In this enzyme−substrate structure there are 53 QM atoms of which 5 are termination atoms and 3509 MM atoms. When optimizing the geometry, 12 QM atoms were fixed (1, 12, 13, 14, 15, 16, 17, 27, 30, 31, 32, 33) to speed convergence.

To investigate the amide hydrolysis reaction pathway a potential energy surface was generated with a range of $S_{11}-C_{46}(O_{47})$ (3.0 to 1.9 Å) and $((N_{25})H_{26}-N_{48})$ (1.1 to 3.3 Å) distances. In total over 100 structures were thus generated, in which all geometric parameters of the QM system, except for these S−C(O) and (N)H−N lengths, were fully optimized. Initially the AM1 method was used for all optimizations. The structure of the MM environment was not altered during the series of optimizations because of the cavity produced on reducing the size of the substrate. For the minimum energy reactant structure we take R(S−C(O)) = 3.40 Å and R((N)H−N) = 3.66 Å, these being the values from the MM minimized enzyme−substrate complex where the full substrate was considered.

A broad scan of the potential energy surface was first performed, and these results are given in Figure 5 and Table 5. This broad scan gives an overall picture of the potential energy surface of amide hydrolysis, and from this it is clear that the reaction proceeds via S^- attack being concerted with histidine proton delivery. The maximum on the lowest energy path from reactants to products (i.e. the transition state) is close to R(S−C(O)) = 2.40 Å and R((N)H−N) = 1.46 Å. A finer scan was used to locate more precisely the transition geometry for the reaction. This structure was subsequently refined by using a transition state searching algorithm and confirmed to be a transition state by calculating the second derivatives of the energy. This structure is described in Figure 6, together with that of the reactants. It can be seen that the reaction proceeds via a transition state without the intervention of a tetrahedral intermediate. The (O)C−N bond of the substrate is beginning to break in the transition state structure to form the acyl enzyme complex with loss of H_2NCH_3.

Table 5. Energies of AM1 Optimized Structures for Amide Hydrolysis (Broad Scan)

(N)H--N distance (Å)	S--C(O) distance (Å)					
	3.00[a]	2.80	2.60	2.40	2.20	2.00
3.26	−**237.93**[b]	−237.53	−236.13	−232.53	−224.22	−211.05
3.06	−**237.33**	−237.22	−236.16	−233.08	−225.36	−212.90
2.86	−236.53	−**236.60**	−235.85	−233.15	−226.07	−214.99
2.66	−235.30	−**235.54**	−235.05	−232.81	−226.51	−216.29
2.46	−235.36	−**233.81**	−233.63	−231.90	−226.38	−216.98
2.26	−230.74	−231.40	−**231.55**	−230.42	−225.72	−217.16
2.06	−227.61	−228.47	−**229.06**	−228.52	−224.76	−217.11
1.86	−224.61	−225.29	−226.01	−**226.32**	−223.48	−216.97
1.66	−218.47	−219.56	−221.08	−**222.45**	−220.97	−216.00
1.46	−212.58	−213.89	−215.88	−**218.13**[c]	−**218.07**	−214.45
1.26	−219.46	−218.52	−218.01	−217.67	−**217.73**	−215.66
1.06	−220.76	−222.33	−224.50	−227.45	−**229.00**	−227.61

[a] AM1 heats of formation are in kcal/mol. [b] Bold font indicates lowest energy path. [c] Transition state.

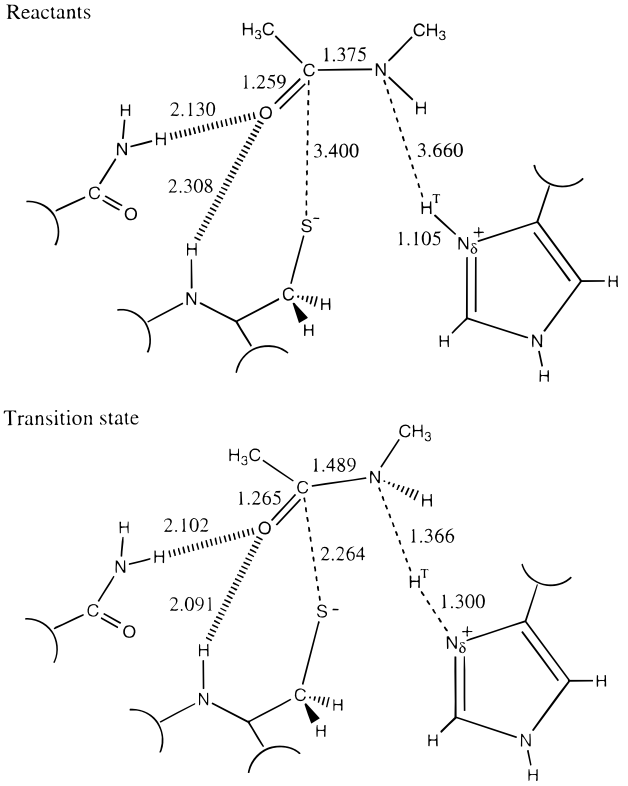

Figure 6. Active site/substrate geometry for reactants and transition state for amide hydrolysis. (Distances are in Å.)

As we have already seen, planarity of the reactive sulfur and the imidazole ring facilitates the formation of the thiolate−imidazolium ion pair. During amide hydrolysis, rotation of the imidazole ring by ≈20° about the $C_\beta-C_\gamma$ bond of His-159 puts the protonated $N_{\delta 1}$ in an optimal position to donate a proton to the substrate amide group. The energy of the transition state is 20.1 kcal/mol higher than that of the reactants *in situ*. *In vacuo* this transition state barrier increases to 23.1 kcal/mol. These *in vacuo* energies are evaluated by performing QM single point calculations on the optimized QM/MM reactant and transition state structures but without the enzyme environment. From these results the protein environment preferentially stabilizes the transition state by ≈3 kcal/mol.

However, this is a rather simplistic view of the *in vacuo* situation because from the results of previous calculations the reactant geometry may well be significantly different from that *in situ*. This is due in part to the Cys-25 and His-159 residues being more stable in their neutral form in the absence of

Table 6. AM1 Energies of Reactants and Transition State for Amide Hydrolysis

	energies of AM1 optimized structures (kcal/mol)	
structures	QM/MM (in situ)	QM (in vacuo)
reactants	−237.7	−157.3
transition state	−217.6	−118.1
transition state barrier	20.1	39.2

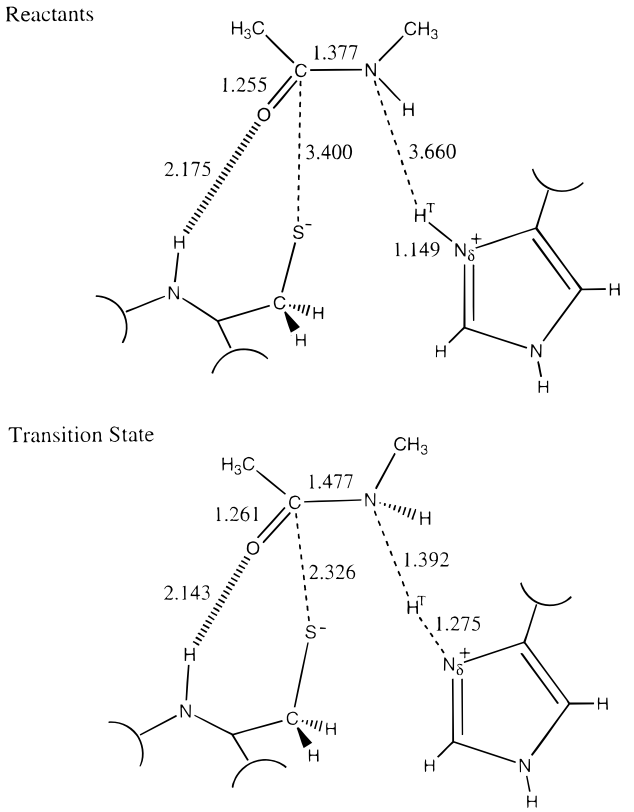

Figure 7. Active site/substrate geometry for reactants and transition state for amide hydrolysis with mutation of Gln-19 to Ala. (Distances are in Å.)

interaction with the protein environment (see Figure 4). The QM reactant system was therefore optimized in the absence of interaction with the MM atoms. A number of additional atoms were fixed (18, 34, 35, 36) to maintain the essential features of the enzyme active site. The reactant and transition state *in vacuo* energies are given in Table 6 and are compared against the *in situ* energies. The results show that there is a large increase in the transition state barrier without the enzyme present. The reason for this increase in barrier height is mainly due to the optimization of the QM reactant geometry. The geometries of the transition state and the reactants are similar to those calculated with the QM/MM model except that in the QM reactant structure proton transfer has occurred from the protonated histidine to the negatively charged cysteine. Thus, the *in vacuo* reaction must first form the thiolate−imidazolium ion pair before amide hydrolysis can proceed thereby resulting in an increased barrier for reaction. On the other hand, as this ion pair is already present in the enzyme, the *in situ* reaction can proceed more readily with a lower barrier.

The role of the Gln-19 residue has been investigated experimentally by mutating Gln-19 to Ala-19, which removes one of the hydrogen bond donors in the proposed oxyanion hole of papain. To model this mutation the QM Gln-19 atoms were replaced by MM Ala-19 atoms which are also unpolarizable, and the potential energy surface of the hydrolysis reaction was

calculated following the procedure adopted for the native enzyme. Such a mutation raises the calculated barrier to the reaction from 20.1 to 21.0 kcal/mol, close to the experimental increase of 2.4 kcal/mol.[20] The transition state structure also shows reduced interaction of the substrate with the active site thus reducing its stabilization by the enzyme (Figure 7) as reflected in the lengthening of the S−C(O) and (N)H−N bond distances.

We have not carried out higher level calculations of the reaction surface for amide hydrolysis in view of the success of the AM1 method both in predicting gas phase proton transfer energies (table 1) and in modeling the active site of the enzyme when compared to calculations at the B3LYP/3-21G* level, and bearing in mind the considerable computing resources needed to undertake such a study which must include electron correlation.

Conclusions

The results reported in this paper have shown the ability of the hybrid QM/MM scheme to credibly model the chemistry of large complex biomolecules. It has been clear for a long time that coupling of QM and MM approaches is useful and often necessary in modeling enzymatic reactions.[28,25,39−43] However, most of the attempts in the literature have involved single point calculations, or the prediction of minimum energy structures. We have shown here how such a hybrid approach allows efficient geometry optimization and transition state searching for complex molecular systems. In particular we allow the structural refinement of the bulk enzyme and solvent in response to the movements of the QM atoms.

By using this QM/MM approach several key features of the catalytic pathway of the enzyme papain have been identified. The calculations highlight the usefulness of the method in the determination of geometries and energies of structures that are dependent on the electrostatic and steric interactions within a macromolecular environment. When comparisons are possible, the results obtained with this QM/MM method on the enzyme papain are in good agreement with relevant experimental findings and have answered key questions concerning the catalytic mechanism which have proved difficult or impossible to answer through the use of experimental or other theoretical methods. The mutation of the active site residue Asn-175 to Ala-175 is seen to have a marked effect on ion pair stability. This was unclear from experimental observations as the role of the mutation could not be unequivocally assigned to ion pair stability or intrinsic activity. The potential energy surface for amide hydrolysis has been mapped out, and the transition state structure for the reaction *in situ* has been identified and properly characterized. Within the limitations of the AM1 method it is evident that the reaction is a concerted process without the involvement of a tetrahedral intermediate.

Acknowledgment. We thank EPSRC for support of this research.

JA9711472

(39) Alagona, G.; Desmeules, P.; Ghio, C.; Kollman, P. A. *J. Am. Chem. Soc.* **1984**, *106*, 3623−3632.

(40) Waszkowycz, B.; Hillier, I. H.; Gensmantel, N.; Payling, D. W. *J. Chem. Soc., Perkin Trans. 2* **1991**, 2025−2032.

(41) Bash, P. A.; Field, M. J.; Davenport, R. C.; Petsko, G. A.; Ringe, D.; Karplus, M. *Biochemistry* **1991**, *30*, 5826−5832.

(42) Karplus, M.; Evenseck, J. D.; Bash, P. A.; Field, M. J. *Faraday Discuss.* **1992**, *93*, 239−248.

(43) Vasilyev, V. V. *J. Mol. Struct. (THEOCHEM)*, **1994**, *110*, 129−141.