geometries corresponding to critical points on the surface have been fully optimized by gradient methods and characterized by determining the index of the relevant Hessian matrices.

The possibility of three kind of attacks to produce dioxetane has been considered: (1) a concerted [2 + 2] attack of C_{2v} (supra-supra) or C_2 (supra-antara) symmetry; (2) a diradicaloid attack of C_1 (gauche) or C_s (syn or anti) symmetry, which could be concerted (but asynchronous) or not, depending on the existence of diradical minima; (3) an attack leading directly to peroxirane, which could be an intermediate for the reaction leading to dioxetane as final product.

A synchronous concerted path does not exist: both the planar supra-supra approach and the supra-antara approach, obtained from the supra-supra by rotation of O_2 along the C_2 axis, correspond to saddle points of order higher than 1, not related to a transition structure. These points are located ca. 50 kcal mol^{-1} above the dissociation limit with both basis sets.

For a C_s syn approach, having some diradical character, the lowest energy surface corresponds to a singlet state that could be labeled $\Pi 3$ from the occupancy of the π molecular orbitals orthogonal to the molecular symmetry plane. This surface puts the lowest singlet state of the syn-diradical minimum in direct relation to one of the two degenerate singlet states of O_2 (reactants at infinite separation), but it cannot connect the syn $\Pi 3$ saddle point directly with the product dioxetane, whose ground state is $\Pi 4$. Due to the symmetry of the system, these two species are related by a real crossing of the $\Pi 3$ and $\Pi 4$ surfaces. For the same reason a real crossing occurs between the supra-supra critical point and the two syn saddle points that have been found on the $\Pi 3$ and $\Pi 4$ surface. These syn saddle points are both second-order, and the lowest one is located ca. 30 kcal mol^{-1} above the dissociation limit.

Corresponding to a C_1 gauche attack, a first-order saddle point having some diradical character is found. The transition structure is located ca. 29 kcal mol^{-1} above the dissociation limit. Corresponding to this transition structure, a gauche minimum of the peroxy diradical is found. In the gauche attack, the symmetry of the system is lowered and the two electronic states considered above can mix: the system can therefore proceed from the reactants through a C_1-diradical path to the dioxetane minimum in two steps.

A second first-order saddle point is found for a C_s anti approach, leading to another conformational minimum of the peroxy diradical. The barrier height is the same as for the gauche approach. The conformational minima of the peroxy diradical are very close in energy and present a barrier for redissociation of ca. 12 kcal mol^{-1}.

A path leading directly to peroxirane does not appear to exist because the C_s peroxirane-like critical point that has been found is a second-order saddle point. Consequently, the peroxirane minimum (as the dioxetane minimum) seems to be reachable only passing through the gauche-diradical minimum.

Acknowledgment. This work was supported by grants from NATO (RG 096-81) and the NSF (CHE 83-12505 and CHE 87-11901). We thank CSI-Piemonte and Wayne State University for generous allocations of computer time.

Registry No. O_2, 7782-44-7; ethene, 74-85-1.

A Simulation of the Sulfur Attack in the Catalytic Pathway of Papain Using Molecular Mechanics and Semiempirical Quantum Mechanics

Dorit Arad, Robert Langridge, and Peter A. Kollman*

Contribution from the Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143. *Received August 29, 1988*

Abstract: Recent theoretical studies revealed a unique behavior of the sulfur nucleophile in its attack on formamide, relative to the analogous oxygen reaction. These studies together with experimental evidence suggest that an exact correspondence between the serine and sulfhydryl enzymatic hydrolysis mechanism cannot be made. This prompted us to simulate alternative pathways on the sulfhydryl peptide hydrolysis reaction surface by molecular mechanics and semiempirical quantum mechanical methods in order to attempt to find those that might be in accord with theoretical as well as the experimental results. The molecular mechanics AMBER minimizations of two different enzyme-substrate conformations in papain led to a new conformation which is 29.2 kcal/mol lower in energy then the classic Michaelis complex in which the substrate carbonyl is pointed toward an oxyanion hole. This new conformation is a noncovalent complex between the charged sulfur and a NH bond of the substrate, and is stabilized by electrostatic interactions between sulfur and the substrate. Both Michaelis complexes were used as a basis to construct two tetrahedral covalent structures. The one based on the more stable Michaelis complex was calculated to be more stable by 14.5 kcal/mol. Finally, semiempirical AM1 reaction paths simulations were performed on each of the conformations, simulating the approach of the substrate reactants to the sulfur nucleophile, and the proton-transfer reaction from histidine-159 to the substrate. Our results suggest a new mechanism for papain-catalyzed hydrolysis of peptides, in which a proton transfer from histidine either to oxygen or nitrogen occurs prior to or concerted with the nucleophilic attack. The results suggest reinterpretation of some experimental data and encourage new experiments to test these predictions.

The cysteine proteases constitute a group of endopeptidases whose members owe their catalytic activity to the presence of cysteine and histidine residues in their active site.[1] The three-dimensional structure of the cysteine protease papain has been determined at high resolution using X-ray diffraction by Kamphius et al.[2] who have also proposed a mechanism for the action of papain on the basis of the observed binding of chloromethyl ketone substrate analogues to the enzyme.

The mechanism proposed by these authors is similar to the mechanism of catalysis postulated for the serine proteases.[3] Since

(1) (a) Polgar, L.; Halasz, P. *Biochem. J.* **1982**, *207*, 1-10. (b) Willenbrock, F.; Kowlessur, D.; O'Driscoll, M.; Patel, G.; Quenby, S.; Templeton, W.; Thomas, E. W.; Willenbrock, F. *Biochem. J.* **1987**, *244*, 173-181 and references cited therein.

(2) Kamphuis, G.; Kalk, K. H.; Swarte, B. A.; Drenth, J. *J. Mol. Biol.* **1984**, *179*, 233-256.

0002-7863/90/1512-491$02.50/0

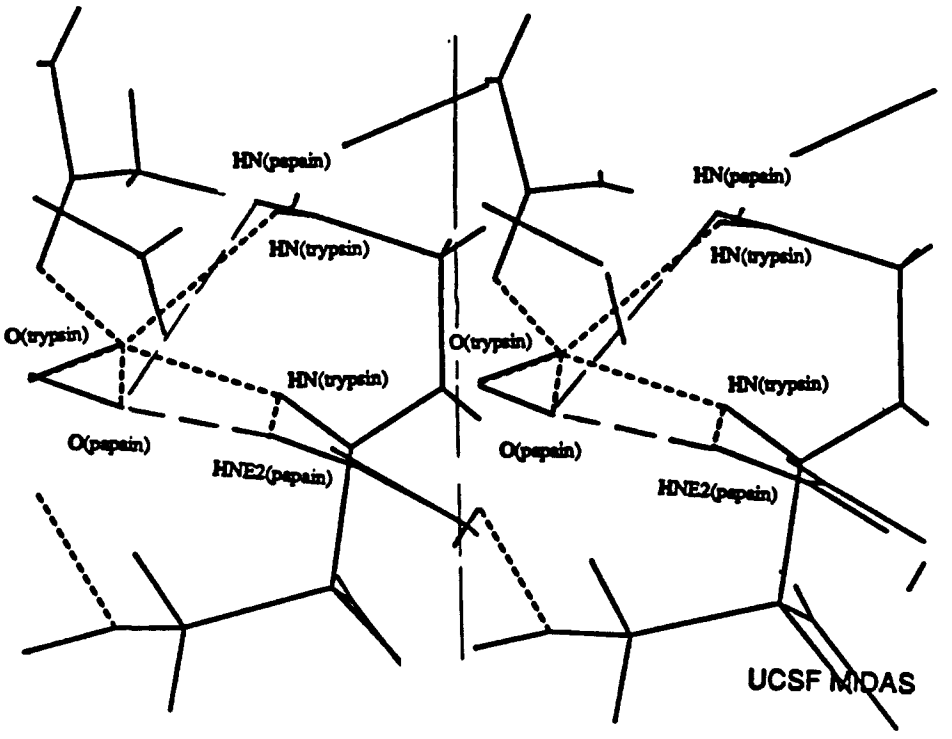

Figure 1. A stereoview of the oxyanion holes in papain versus trypsin. The models displayed are the enzyme–substrate tetrahedral complexes of papain and trypsin. The papain model building is described here (Figure 8), and the trypsin enzyme–substrate model was constructed by Seibel et al.[27] Superposition of the two models show that the substrate oxygens are displaced by 0.66 Å and the hydrogens that form the oxyanion holes (e.g., HN and HNE2(pap) and HN (tryp)) by 0.41 and 0.23 Å, respectively.

the nature and orientations of the catalytic groups in serine and sulfhydryl proteases are similar,[4] the assumption of analogous mechanisms is quite reasonable. The proposed mechanism may be summarized as follows. The side chains of residues Cys-25 and His-159 from an ion-pair $ImH^+\cdots S^-$. This ion-pair conformation has been suggested to be the "resting state" of the enzyme.[5] The orientation of the sulfur atom in the resting state is approximately coplanar with the imidazolium ring of His-159. The substrate binds to the enzyme in an "oxyanion hole" conformation formed by the main chain NH of Cys-25 and the side-chain NH_2 of Gln-19, thereby facilitating the nucleophilic attack of the sulfur on the carbonyl of the substrate. The hydrogen bond in the oxyanion hole polarizes the oxygen of the carbonyl in the bound substrate and thus facilitates the nucleophilic attack by sulfur on the carbonyl by lowering the activation energy for this process.[6]

The next step is the formation of the charged tetrahedral intermediate resulting from the attack of the charged sulfur atom of Cys-25 on the carbonyl of the substrate. The tetrahedral intermediate is protonated on the amide nitrogen by a proton transfer from the imidazolium ion. Loss of the NH_2R group from this charged tetrahedral intermediate forms the acyl enzyme complex EA. The acyl-enzyme complex EA is then hydrolyzed to give products. (In the serine proteases, the rate-determining step for amides is acylation, and for esters it is the deacylation step.)

A survey of the literature, however, suggests some difficulties in accepting the complete analogy between the serine and sulfhydryl mechanisms.

a. The rate for the hydrolysis reaction catalyzed by papain is slower by a factor of 10^2–10^3 than the rate for the analogous serine protease catalyzed reaction.[7]

b. The S1 (P1) selectivity index for papain is very wide in papain compared to the serine proteases. Papain cleaves a large variety of amino acids residues in its P1 site.[8]

c. The "oxyanion hole" which plays an essential role in the catalytic mechanism of the serine proteases may not be essential in the hydrolytic mechanism of papain.[9] This was shown by Polgar et al.,[9] who replaced the carbonyl of the substrate with a thiocarbonyl bond and assumed that since the C=S bond is much longer than the C=O bond, the thiocarbonyl may not fit the same oxyanion hole site. In the serine proteases thioesters are not cleaved at all, yet the same replacement does not affect the rate of hydrolysis of papain the acyl-enzyme step. The rate is reduced by a factor of 2 in the deacylation step. One might argue that these differences in activity may result from a different structure of an oxyanion hole in sulfhydryl and serine proteases, and that papain should have a larger oxyanion hole than trypsin, for example, so that a larger C=S bond can be accommodated into this hole. We tested this argument by graphically superimposing the oxyanion holes of papain and trypsin and found that the structure of the oxyanion holes are very similar (Figure 1).

d. Recent C-13 NMR studies of an inhibition complex of papain and substrate aldehyde (**1**[10]) showed that two enantiomers, thiohemiacetals **2a** and **2b** (the result of S^- attack on the aldehyde carbonyl), are observed. This observation is different from that made in analogous NMR experiments in the serine proteases in which only one enantiomer could be detected.[11] As the binding features of substrate **1** are specific to papain (e.g., the P2 binding site is a phenylalanine residue which has a specific hydrophobic pocket in the enzyme), this result suggests that the substrate is not locked in an "oxyanion conformation" but can adopt other

(3) Drenth, J.; Kalk, K. H.; Swen, H. M. *Biochemistry* **1976**, *15*, 3731–3738.
(4) Garavito, R. M.; Rossmann, M. G.; Argos, P.; Eventoff, P. *Biochemistry* **1977**, *16*, 5065.
(5) van Duijnen, P. T.; Thole, B. T.; Broer, R.; Nieuwpoort, W. C. *Int. J. Quantum Chem.* **1980**, *17*, 651–661.
(6) (a) Kraut, J. *Annu. Rev. Biochem.* **1977**, *46*, 331–358. (b) Willenbrock, F.; Brocklehurst, K. *Biochem. J.* **1985**, *227*, 521–528. (c) Asboth, B.; Polgar, L. *Biochemistry* **1983**, *22*, 117–122.
(7) (a) Polgar, L.; Bender, M. L. *Biochemistry* **1969**, *8*, 136. (b) Hill, R. L.; Scmidt, W. R. *J. Biol. Chem.* **1962**, *237*, 389.
(8) (a) Smith, E. L.; Kimmel, J. R. In *The Enzymes*; Boyer, P. D., Lardy, H., Myrbach, K., Eds.; Academic Press: New York, 1960; Vol. 4, p 133. (b) Lowe, G. *Tetrahedron* **1976**, *32*, 291–302.
(9) Asboth, B.; Stokum, E.; Khan, I. U.; Polgar, L. *Biochemistry* **1985**, *24*, 606–609.
(10) Mackenzie, N. E.; Grant, S. K.; Scott, A. I.; Malthouse, J. G. G. *Biochemistry* **1986**, *25*, 2293–2298.
(11) Shah, D. O.; Kofen, L.; Gorenstein, D. G. *J. Am. Chem. Soc.* **1984**, *106*, 4272–4273.

conformations in which the attack of the sulfur moiety would lead to the other enantiomer (Figure 2). These new data still agree that the basic features of the serine and sulfhydryl mechanisms are similar, inasmuch as both involve an acyl-enzyme intermediate that is formed by nucleophilic attack on the carbonyl.

Theoretical investigations of the gas-phase reaction coordinate for nucleophilic attack of hydrosulfide ion on formamide by Howard and Kollman[12] revealed that this reaction is different from the analogous hydroxide–formamide reaction. The striking difference is the absence of a stabilizing potential for a tetrahedral intermediate complex of SH^-/formamide. High-level ab initio calculations show that the only minima found on the reaction potential surface is an ion–dipole complex which is 28 kcal/mol lower in energy then the reactants.

Howard and Kollman suggest that the gas-phase results may apply to similar solution reactions, and alternative pathways in the hydrolysis mechanism of sulfhydryl proteases should be considered. The experimental discrepancies regarding the mechanism suggested by Kamphius et al. together with the results found by the ab initio model calculations on the behavior of the sulfur nucleophile prompted us to model and simulate a papain–substrate structure, and to investigate computationally a number of possible alternatives pathways in the hydrolysis of peptides by papain. Our strategy was to build several enzyme–substrate models with molecular mechanics and follow the reaction paths of each using semiempirical quantum mechanical methods. Our results suggest alternative pathways in which protonation of the oxygen or nitrogen on the substrate occurs prior to the nucleophilic attack by the sulfur or in a concerted manner with such an attack.

Methods

The molecular mechanics studies were accomplished using the molecular mechanics program AMBER.[13] The force field equation for AMBER is given by eq 1:

$$E_{total} = \sum_{bonds} K_r(r - r_{eq})^2 + \sum_{angles} K_\theta(\theta - \theta_{eq})^2 + \sum_{dihedrals} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] + \sum_{i<j}\left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\epsilon R_{ij}}\right] + \sum_{H\ bonds}\left[\frac{C_{ij}}{R_{ij}^{12}} - \frac{D_{ij}}{R_{ij}^{10}}\right] \quad (1)$$

The calculations were carried out on VAX8650 and FPS-264 computers. An all-atom force field was used; i.e., all atoms were explicitly represented by the force field, using the force-field parameters that have been previously published.[14] The geometries of the complexes were optimized until the root-mean-square (rms) energy gradient was less than 0.1 kcal mol.$^{-1}$ Throughout all calculations a nonbonded van der Waals cutoof of 8 Å was used, and a hydrogen bond cutoff of 4 Å. The minimizations were performed on the protein, "in vacuo", using a distance-dependent dielectric constant. Sets of partial charges were assigned to the nonstandard residues and the binding sites of the ES complex by performing single-point calculations on a model system, using the program GAUSSIAN 80 UCSF[15] with a 4-31G set, and fitting the calculated electrostatic potential to partial charges.[16] For example, for simulating the ion-pair $His^+\cdots Cys^-$ state in papain, a set of partial charges for Cys^- had to be derived. (Charges for His^+ are available in the AMBER data base.[13]) The model CH_3–S^- was optimized at the HF//4-31G level and the set of charges obtained was fitted into the side chain of cysteine$^-$. These charges were approximately −0.9 for the sulfur and 0.2 for the β carbon. Visualization, graphical manipulations, and minor structure refinements were done using the program MIDAS[17] on the PS2 Evans and Sutherland display system. Solvent accessible surfaces for the enzyme's active site region were created using the program MS.[18]

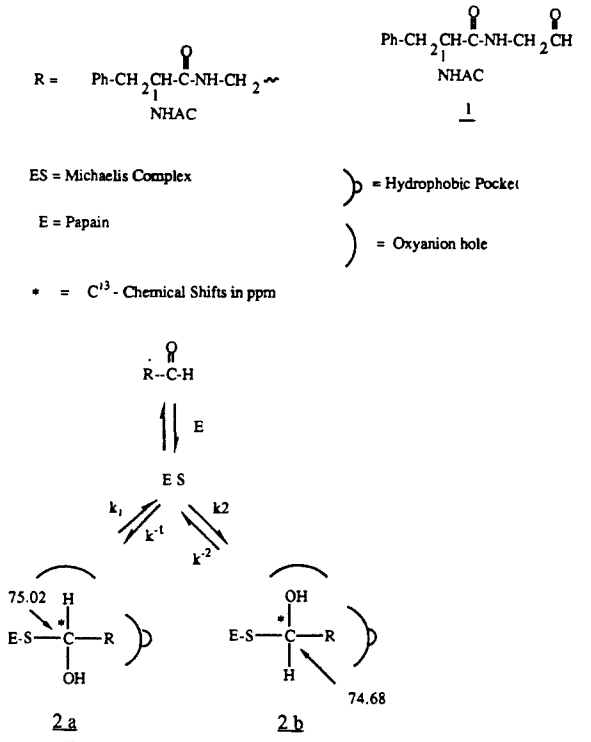

Figure 2. ^{13}C chemical shifts in ppm of the aldehyde **1** bound to papain. Two signals are observed in the NMR spectrum of the thiohemiacetal complexes, which correspond to the formation of two bound enantiomers **2a** and **2b**.

Table I. Calculated and Experimental Gas-Phase Proton Affinities of Methanol, Methanethiol, and Imidazole and the Relative Gas-Phase Energies for the Corresponding Proton-Transfer Reactions (kcal/mol)

reaction[a]	AM1	4-31G[b]	exp[b]
$CH_3OH \rightarrow CH_3O^- + H^+$	410.7	407.0	377.0
$CH_3SH \rightarrow CH_3S^- + H^+$	320.6	370.0	355.0
$ImH^+ \rightarrow Im^d + H^+$	170.0	248.0	220.0
$CH_3OH + Im \rightarrow CH_3O^- + ImH^+$	163.0		157.0
$CH_3SH + Im \rightarrow CH_3S^- + ImH^+$	146.6		135.0
$H\text{-}C{=}O(NH_2) + ImH^+ \rightarrow H\text{—}C{=}OH^+(NH_2) + Im$	20.0		26.0[c]

[a] Energies in kcal/mol. [b] Reference 21. [c] Reference 22. [d] ImH^+ is a protonated imidazolium ring; Im is a neutral imidazole molecule.

The reaction paths were calculated using the AM1 program which is available from the semiempirical package AMPAC.[19] AM1 is an improved version of MNDO which partially overcomes the major weakness of MNDO, which is the failure to reproduce hydrogen bonds correctly.[20] To evaluate the performance of AM1 for the enzyme system it is necessary to make a comparison between the calculated and experimental values of some relevant molecules. In our case the appropriate molecular systems will be methanol, methanethiol, and imidazole. We compared the experimental and calculated proton affinities of these compounds, and the results are given in Table I, along with energies for proton transfer reactions.

AM1 underestimates absolute values for proton affinities of thiomethane by 30 kcal/mol and for HS^- by 10 kcal/mol; however, it slightly overestimates the values for photon-transfer reactions from methanol and methanethiol to imidazole. The corrections that have to be made for these reactions are relatively small compared with other semiempirical methods, only 6–11 kcal/mol compared to 60 kcal/mol in CNDO, for example.[21] On comparing the proton-transfer energies, we see that AM1 overestimates the energies for a proton transfer for methanol and methanethiol to imidazole by similar amounts. However, the value for proton

(12) Howard, A.; Kollman, P. A. *J. Am. Chem. Soc.* **1988**, *110*, 7195–7200.
(13) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. J. *J. Am. Chem. Soc.* **1984**, *106*, 765. (b) Singh, U. C.; Weiner, P. K.; Caldwell, J.; Kollman, P. A. AMBER 3.0, University of California—San Francisco, 1986.
(14) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Comput. Chem.* **1986**, *7*, 230–252.
(15) Singh, U. C.; Kollman, P. A. GAUSSIAN 80 UCSF, *QCPE Bull.* **1982**, *2*, 17.
(16) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1984**, *5*, 129–145.
(17) Ferrin, T. E.; Huang, C. C.; Jarvis, L. E.; Langridge, R. *J. Mol. Graphics* **1988**, *6*, 2.
(18) Connolly, M. MS, Quantum Chemistry Program Exchange (QCPE) Program.
(19) Dewar, M. J. S.; Zoelisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902–3209.
(20) Dewar, M. J. S.; Thiel, W. J. *J. Am. Chem. Soc.* **1977**, *99*, 4899–4907.
(21) Kollman, P. A.; Hayes, D. M. *J. Am. Chem. Soc.* **1981**, *103*, 2955.
(22) Kollman, P. A.; Rothenberg, S. *J. Am. Chem. Soc.* **1977**, *99*, 1333.

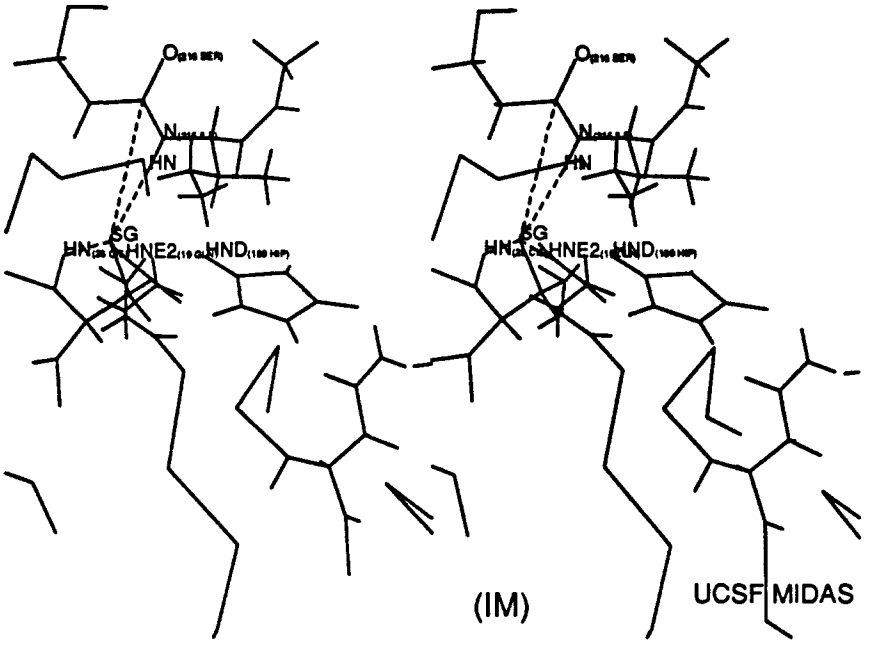

Figure 3. Stereoview of a noncovalent structure of "ion-molecule conformation" (IM). The ion-molecule interaction is between the sulfur (Cys-25) at the enzyme's surface and the amide hydrogen (Ile-216). The carbon-sulfur bond distance is 4.0 Å. Note that the angle of approach of the substrate to the enzyme's surface is almost perpendicular to the enzyme's surface, e.g., ∠SG-C-O = 160.0°.

transfer from imidazole to formamide is underestimated by 6 kcal/mol, which implies that a proton is transferred more easily to the oxygen on an AM1 reaction surface than would be observed experimentally.

We also used AM1 to study the SH^-···formamide model system that we had earlier examined with ab initio calculations. We calculated the energy of the system as a function of S–C distance. In contrast to the ab initio calculations, the AM1 calculations led to a local minimum for a tetrahedral complex with R(C–S) of 2.04 Å and a stabilization relative to the isolated molecules of 18 kcal/mol. However, the ion–dipole complex with R(S–C) = 3.8 Å and the S^- hydrogen bonding to the NH_2 was the global minimum, 25 kcal/mol more stable than the separate species.

Results

Model Building. The initial starting conformation for our model was the papain–substrate complex studied by Drenth and coworkers.[2] We were interested in building a model which contains a peptide substrate bound to papain, for illustrating a hydrolysis reaction pathway which is as close as possible to the real one. The peptide substrate was chosen from the binding specificity of experimental data. Berger and Schechter,[24] showed that the active site for papain contains sites for at least six residues along the chain of the peptide or protein substrate. Such a chain may be represented as follows:

P4–P3–P2–P1–P1′–P2′

Our model contains the P2–P1–P1′ residues. P2 is hydrophobic. Phe is one of the residues that has the best binding values to the P2 cleft and was chosen in the model for this reason. For P1 we chose Ser. Ala and Ser have better cleaving rates then Gly, but charged residues such as arginine are also cleaved at high rates. As the binding specificity at this site is low, either of the above residues could have been chosen in our model. For P1′ the common residues that cleave from this side are highly hydrophobic. It was found[25] that papain can accommodate a very large hydrophobic cleft at this site. ILE is commonly observed, and we thus chose it for our model substrate.

Our substrate contains the following residues

Ace–Phe–Ser–Ile–NMe

where Ace is an acetyl group and NMe is an N-methyl group. The initial geometry for the papain–substrate Michaelis complex was created by superimposing its coordinates on the coordinates of the enzyme–inhibitor covalent complex of Drenth that was available from the Brookhaven data bank[3] using the model 6PAD for the substrate. For the protein, however, we used a newer high-resolution structure of papain (1.65 resolution) available in the data bank as a model called 9PAP.[26] For the superposition of our model on the crystal data model we used the program EDIT of AMBER. The initial structure was manipulated. A solvent accessible surface of the active site was generated, and the bad contacts and steric effects on the substrate active site surface was reduced using the program MIDAS. This initial structure was then subjected to a minimization in which residues in the region of the active site (18–29, 64–69, 131–137, 157–160, 175–177, 204–207) and the substrate were allowed to move and the rest of the protein was kept fixed. The only constraint was to keep the S···C=O distance fixed at a distance of 3.0 Å, with a force constant of 150. The carbonyl to be cleaved pointed toward the "oxyanion hole" formed by the NH_2 group of Gln-19 and NH group of the Cys-25 main chain. At this point only the substrate was minimized keeping the protein fixed until an rms energy gradient of 0.01 kcal/(mol Å) was achieved.

The new superimposed structure was subjected to further minimization, releasing all the degrees of freedom and minimizing the entire protein substrate structure. The result is structure IM (Figure 3). This structure can be described as an "ion–molecule complex" since its conformation in the active site region resembles the optimized 4-31G structure of the ion–molecule complex minima of SH^- and formamide (Figure 4).[12] The bond distance S···C=O in IM is 4.0 Å and the interesting feature is that the carbonyl is no longer pointing toward the oxyanion hole but lies in an almost perpendicular plane to the enzyme surface forming an angle of 160° with the S^- nucleophile. The NH bond of the substrate direction forms a hydrogen bond with the sulfur atom of Cys-25 with a S···H distance of 2.2 Å.

We then decided to compare the energy of the "ion–molecule structure" (IM) with a conventional "oxyanion" conformation. We converted the IM structure to an oxyanion structure by reversing the chirality of the carbonyl in the substrate and leaving all the rest of the molecule untouched. This structure was then energy minimized with no constraints. In the resulting structure,

(23) For example: Tapia, O.; Stamato, F. M. L. G.; Smeyers, Y. G. *Theochem* **1985**, *123*, 67–84.

(24) Berger, A.; Schechter, I. *Phil. Trans. R. Soc. London*, **1970**, *B257*, 249–264.

(25) Carotti, A.; Smith, R. N.; Wong, S.; Hansch, C.; Blaney, J. M.; Langridge, R. *Arch. Biochem. Biophys.* **19848** *229*, 112–125.

(26) Kamphius, I. G.; Kalk, K. H.; Swarte, M. B. A.; Drenth, J. *J. Mol. Biol.* **1984**, *179*, 233.

(27) Seibel, G. L.; Kollman, P. A., private communication.

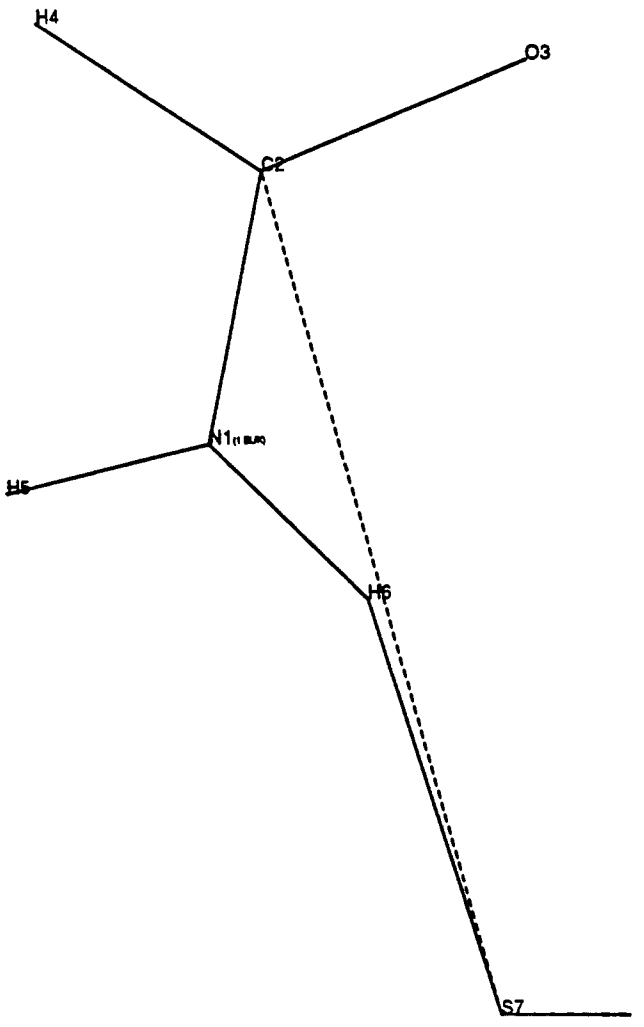

Figure 4. A gas-phase calculated structure of an ion–molecule complex between SH^- and formamide. Geometry was optimized at the RHF/4-31G ab initio method.

OI (Figure 5), the carbonyl of the substrate points toward the oxyanion hole that is formed by the side chain of HN_2 Gln-19 and NH of Cys-25 (main chain). This structure (OI) is higher in energy than the ion–molecule (IM) structure by 29.2 kcal/mol.

For simulating the reaction paths we needed to have reasonable structures of tetrahedral intermediate complexes. The noncovalent complexes IM and OI were constrained using geometrical constraints from the 4-31G optimized structure of the tetrahedral intermediate **6a**. This structure results from the nucleophilic attack of hydrosulfide ion on doubly hydrated formamide. In **6a** the tetrahedral geometry of the intermediate is maintained throughout the optimization at the 4-31G basis set level, with the S–C distance fixed at a distance of 2.1 Å. The details of the distance (C–S = 2.1 Å) and angle constraints applied to IM and OI in order to convert them to tetrahedral structures are given in Figure 6. All constraints were applied with force constants of 150 kcal/(mol Å^2). The partial charges for the tetrahedral structures were assigned to the complexes by fitting the appropriate point charges obtained from a single-point calculation at 4-31G using GAUSSIAN 80 UCSF.[16] The structure used for obtaining the electrostatic potential fit charges is the tetrahedral structure **6b**, but without the two hydrating water molecules. The values of the calculated point charges obtained directly from the model and the actual charges assigned to the protein are given in Figure 6, b and c, respectively.

The structures of the modified ES complexes were then subjected to another set of minimizations, obtaining two tetrahedral complex structures. The first, which is derived from IM, is IM-Tet (Figure 7), and the second, derived from OI, is OI-Tet (Figure 8). In both new tetrahedral structures, the C–S bond distance lengthened from the initial 2.1 Å constrained value to 2.5 Å.[12] The total force-field energy of IM-Tet came out 16.2 kcal/mol lower than the energy for OI-Tet, suggesting that the former may be a more likely candidate for the hydrolysis reaction pathway.

To derive a better picture about the origin of the energy difference between the OI and IM of the four structures (Figure 3, 5, 7 and 8), we analyzed the total energy of each protein as follows using the ANAL facility of AMBER. The residues of each protein were initially divided into three groups: the substrate residues (group A), the active site residues (group B), and the residues of the rest of the protein (group C). The total energy is broken into various components, including bond, angle, dihedral, nonbonded energy, electrostatic energy, and hydrogen bond (HB) energy. Recall that, in the Weiner et al. force field, most of the hydrogen bond energies are included in the electrostatic energy.[13] The energy of each group is divided into internal and group-interaction energy components. The results of this analysis are given in Table II. We shall focus in the text only on the interactions between the active site and the substrate, because much of the differences can be rationalized by differences in these energies.

IM is more stable then OI by 29.2 kcal/mol. This energy difference results mainly from better electrostatic interactions in the ion–molecule conformation. The total nonbonded electrostatic energy is 37.7 kcal/mol lower in IM compared to OI. A small contribution to the energy lowering of IM comes from the HB energy which is 1.7 kcal/mol lower for IM. The other energy components (VDW, bond, angle, dihedral, and 1–4 electrostatics) are higher for IM (0.1–5.0 kcal/mol) and reduce the total energy difference. The interaction part of the electrostatic energy stabilization of IM relative to OI is 17.2 kcal/mol, of which a major part, 15.7 kcal/mol, is due to the electrostatic interaction between the substrate (group A) and the active site region (group B). The HB interaction energy is also slightly lower for IM.

The total energy difference between the IM and the OI conformations is reduced from 29.2 to 16.2 kcal/mol in the tetrahedral pair IM-Tet and OI-Tet. Whereas the internal energy difference remained almost the same as for the complex pair (13.4 kcal/mol, favoring IM-Tet), the interaction energy stabilization is substantially reduced (only 2.5 kcal/mol).

The striking reduction in stabilization of the IM conformation in the tetrahedral form comes from the substrate–active site (A–B) interaction energies. The electrostatic energy difference between IM-Tet and OI-Tet is only 5.9 kcal/mol compared to 15.7 kcal/mol in the complex pair IM/OI. This reduced stabilization can either originate from a better electrostatic interaction in the OI tetrahedral conformation or loss of stabilizing electrostatic interaction in the tetrahedral IM conformation. In order to localize the factors and groups which are responsible for energetic differences, we further broke the interacting groups A and B into fragments. The substrate was divided into three fragments: P2(residues Ace–Phe)P1, which is the carbonyl part in the reaction center (residue Ser), and P1′, which contains the nitrogen part of the reaction center and the rest of the substrate (residues Ile–NMe). Recall that the bond that is being broken in the hydrolysis is the amide bond between Ser and Ile. The active site was divided into eight groups: residues 18–24, 25, 26–27, 64–69, 131–137, 157–160, 175–177, and 204–207. The active site groups that are important to our energy difference discussion are Gln-19 in group 4, the catalytic Cys-25 in group 5, and His-159 in group 9. As can be seen from the energies shown in Table III, there are almost no differences in the energy values between the OI and IM conformations for the rest of the active site groups. Cysteine exhibits a 3.3-kcal/mol better electrostatic interaction with the serine in the IM conformation rather than in the OI conformation. At the distance of 4.0 a for the noncovalent complexes, the OI and IM conformations show about the same electrostatic repulsion between group 3 and the charged sulfur of cysteine.

Although the OI arrangement is slightly preferred electrostatically (4.2 kcal/mol) by the protonated His-159 of group 9 interacting with the carbonyl of Ser, the major contribution for the electrostatic preference of the IM conformation (21.0 kcal/mol) originates from the interaction between His^+ and group

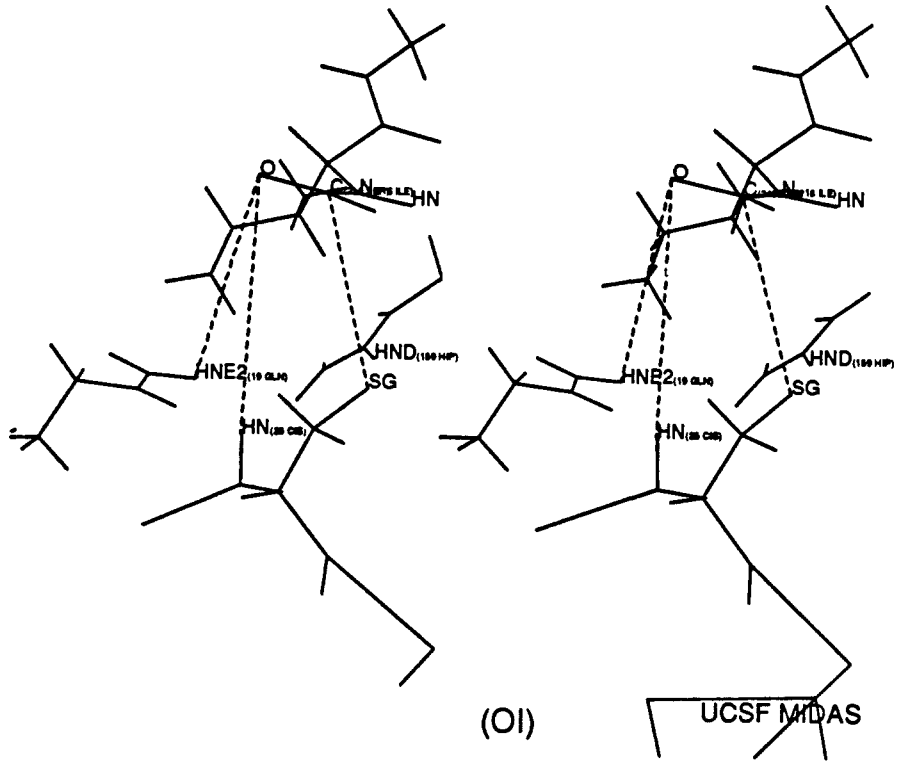

Figure 5. Stereoview of an "oxyanion conformation" (OI) noncovalent complex of papain and substrate. The carbon–sulfur distance is 4.0 Å, the carbonyl oxygen points toward the oxyanion hole formed by HN (Cys-25) and HNE2 (Gln-19), forming C=O···H distances of 2.4 and 2.5 Å, respectively.

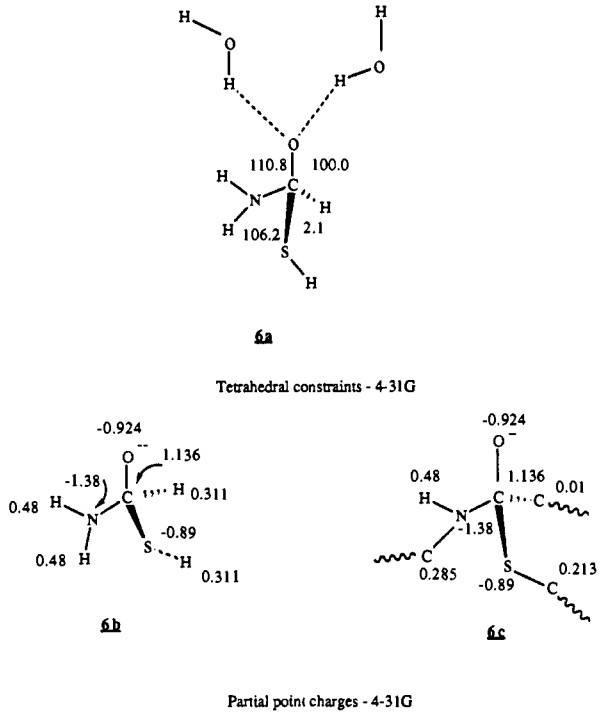

Figure 6. The tetrahedral constraints and the partial charges of the gas-phase structures used to construct a tetrahedral enzyme–substrate complex: **6a**, RHF/4-31G//4-31G optimized structure of a tetrahedral intermediate that results from a nucleophilic attack of hydrosulfide ion on hydrated formamide (the C–S distance was fixed at 2.1 Å, and all the other parameters were optimized); **6b**, point charges derived from GAUSSIAN 80 UCSF of the tetrahedral adduct **6a** without the water molecules; **6c**, actual point charges assigned to the enzyme–substrate structures IM-Tet and OI-Tet.

3. This energy difference may be attributed to the interaction of the carbonyl group in Ile in group 3 with the histidine.

When the reactants get closer in the tetrahedral structures IM-Tet and OI-Tet, new factors dominate the energy differences. At a carbon sulfur distance of 2.95 Å the electrostatic interaction between sulfur and group 3 is ca. 33 kcal/mol in both conformations, but the electrostatic repulsion between the group 3 (ILE-NME) of the substrate and the negative sulfur is much worse in the OI-Tet conformation than it is in IM-Tet. The gain in electrostatic energy as a result of the hydrogen bonding in both conformations compensate each other; e.g., in OI-Tet the hydrogen bond between GLN-19 and the carbonyl of serine is 11.9 kcal/mol better than in IM-Tet, whereas in IM-Tet the electrostatic interaction of His-159 with the carbonyl of serine is 11.3 kcal better than in the OI conformation. Given that the tetrahedral structures have been artificially constrained, one should not overinterpret the energy component differences.

Our conclusions from analyzing the energy components are that the Michaelis complex for papain is energetically favored in a conformation that does not require stabilization by an oxyanion hole, e.g., the ion–molecule conformation. It is an attractive idea to consider the ion–molecule conformation as a possible intermediate on the reaction pathway. Yet some questions to be answered are: Why is the formation of the ion molecule conformation specific to sulfhydryl proteases? Will an ion–molecule conformation in the serine proteases exhibit similar stability relative to the "classic" oxyanion conformation? If we assume that IM is a preferred conformation, how does the hydrolysis reaction proceed from this step?

We suggest that the IM conformation is specific to sulfur enzymes. In the resting state of serine proteases the His–Ser pair does not exist as a charged ion pair but as a neutral H-bonded pair, whereas in the resting state of papain the His–Cys pair exists as an ion pair of negatively charged sulfur and protonated histidine. Thus, an interaction similar to the IM observed in the papain complex will not be as favorable in the serine proteases since it involves only a neutral molecule interaction.

In papain the charged sulfur attracts the substrate to form an ion–molecule conformation. Starting from this conformation there are possibly two or more pathways which can "activate" the enzyme (Figure 9). One is the bending of the carbonyl toward an oxyanion conformation (OI-Tet, route 1) in which the substrate nitrogen is pyramidized and is headed toward the charged histidine. The other (route 2) is the bending of the oxygen toward an IM conformation in which the carbonyl oxygen is headed toward the protonated histidine.

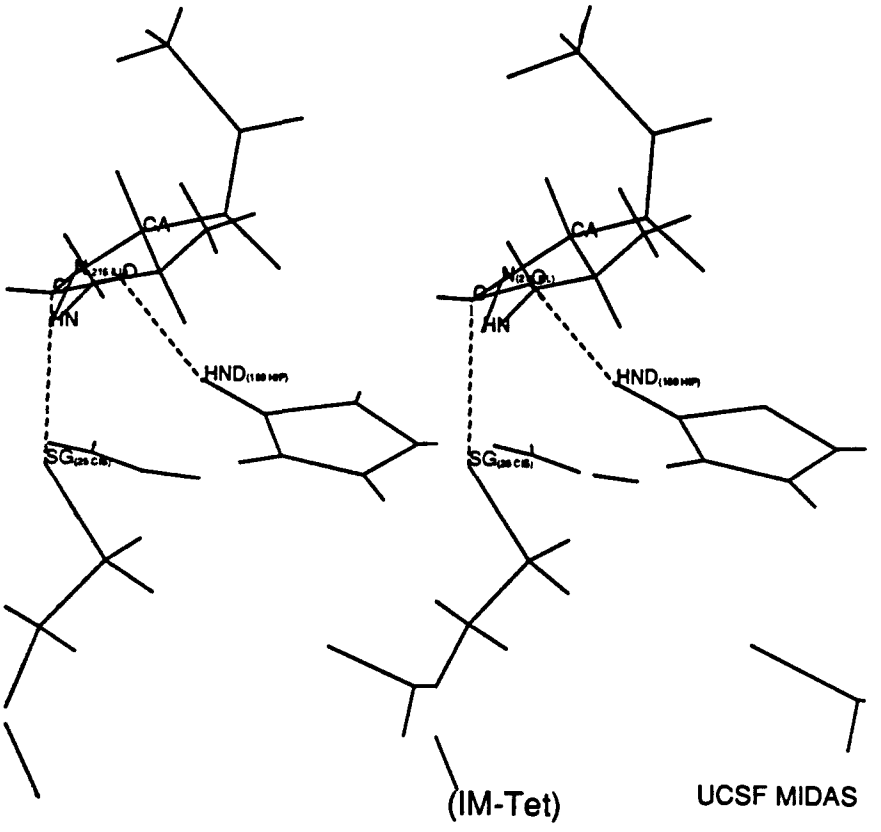

Figure 7. Stereoview of IM-Tet, the covalent tetrahedral ES complex derived from IM (Figure 3). The C–S distance is 2.5 Å. The carbonyl points toward the imidazolium hydrogen of His-159 forming a C=O···H(His) distance of 1.81 Å.

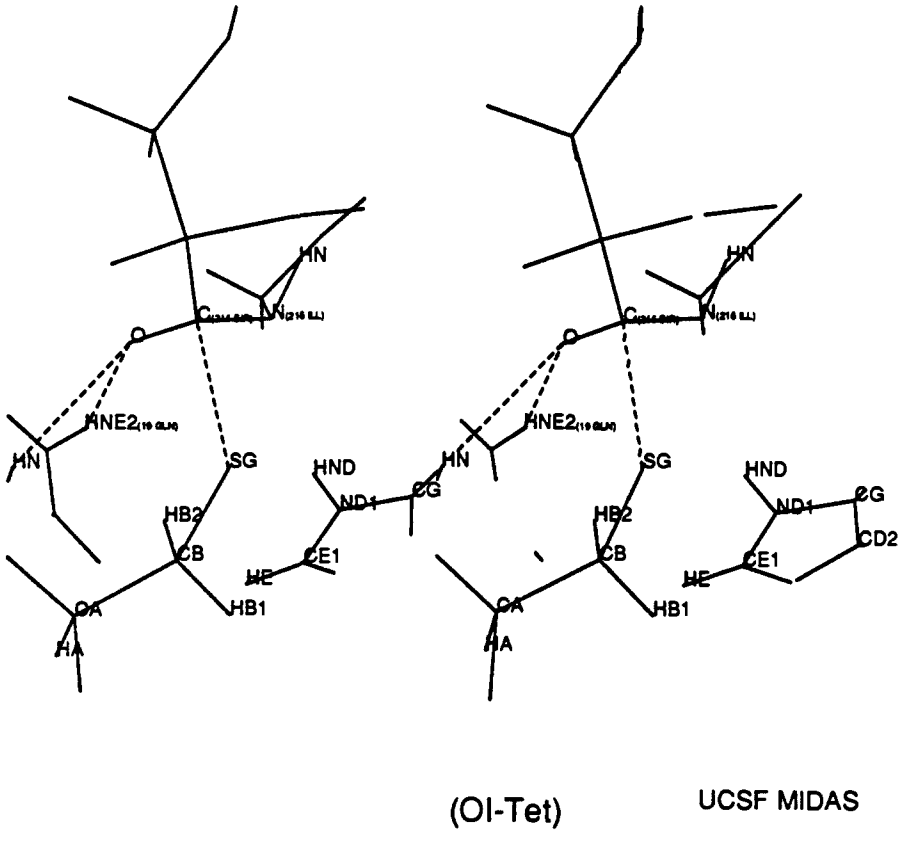

Figure 8.

In both cases the distance of either the oxygen (in IM-Tet) or the nitrogen (in OI-Tet) from the proton of histidine-159 is that of a strong hydrogen bond. (1.9 and 1.85 Å, respectively), and it is possible that proton transfer from this protonated histidine to the substrate may initiate the catalytic process.

The existence of both IM-Tet and OI-Tet conformations is compatible with the experimental results[10] which detect both structures **2a** and **2b** as active species on the reaction surface (Figure 2). Our molecular mechanics calculations showed that, in a S···C=O distance of 2.5 Å, the IM conformation is more

Table II. Total, Internal, and Interaction Energies (kcal/mol) of OI, IM, IM-Tet, and OI-Tet

	IM-Tet[e]	OI-Tet[f]	ΔE	IM[g]	OI[h]	ΔE
			Total Energies[a]			
nonbonded VDW	-1337.7	-1336.9	-0.8	-1362.2	-1364.3	+2.0
nonbonded EEL	-16356.2	16337.9	21.7	-16204.2	-16166.5	-37.7
H bond	-111.1	-110.5	-0.6	-111.9	-110.2	-1.7
bond	63.0	61.9	+1.1	36.9	36.8	+0.4
angle	278.3	283.4	-5.0	265.5	266.7	1.2
dihedral	295.8	301.2	-5.6	285.4	283.4	+2.0
nonbonded 14 VDW	461.8	467.0	-5.2	463.9	464.85	-0.9
nonbonded 14 EEL	7891.4	7912.4	-21.0	7808.1	7803.2	+4.9
bond constraint	24.6	24.4	+0.2	0.0	0.0	0.0
angle constraint	2.6	1.4	1.2	0.0	0.0	
dihedral constraint	0.4	0.23	+0.3	0.0	0.0	
total	-8814.8	-8798.6	-16.2	8817.2	8787.3	-29.2
			Internal Energies of Groups[a]			
bond, angle dihedrals[i]						
A[b]	36.6	45.7	-9.1	10.0	10.2	-0.2
B[c]	88.8	89.5	-0.7	90.3	91.6	-1.3
C[d]	481.0	482.0	-1.0	484.3	479.0	+5.0
nonbonded (VDW)[j]						
A[b]	6.9	7.9	-1.0	2.8	5.1	-2.3
B[c]	-68.2	-63.7	-4.5	-63.0	-63.3	+0.3
C[d]	-633.2	-631.0	-2.2	-631.7	-631.5	-0.2
electrostatic[k]						
A[b]	-142.2	-155.65	+13.4	-83.4	-90.6	+7.2
B[c]	-966.3	-961.25	-5.1	-960.5	-961.6	-1.1
C[d]	-6742.2	-6739.6	-2.6	-6729.3	-6706.8	-22.5
HB[l]						
A[b]	-0.4	-1.1	+0.7	-0.6	-0.9	+0.35
B[c]	-10.0	-9.3	-0.7	-10.4	-9.4	-1.0
C[d]	-87.1	-86.6	-0.5	-87.8	-86.9	-0.9
total						
A	-99.1	-103.1	-4.0	-71.3	-76.1	+4.8
B	-955.8	-944.8	-11.0	-943.7	-942.8	-0.9
C	-6981.4	-6975.4	-6.04	-6964.8	-6946.2	-18.6
bond, angles, dihedrals[i]						
A-B	25.9	24.4	+1.5	0.0	0.0	0.0
B-C	4.5	5.0	-0.5	4.9	4.8	0.0
A-C	0.0	0.0	0.0	0.0	0.0	0.0
nonbonded[j]						
A-B	-10.0	-11.7	+1.7	-33.7	-36.4	+2.7
B-C	-168.9	-168.8	-0.1	-169.9	-170.5	+0.6
A-C	-1.85	-2.1	0.2	-2.0	-2.2	0.2
electrostatic[k]						
A-B	-40.6	-34.7	-5.9	-28.5	-12.8	-15.7
B-C	-574.1	-575.2	+1.7	-595.8	-593.5	-2.3
A-C	0.5	1.7	-1.1	1.3	2.1	0.8
HB[l]						
A-B	-1.9	-1.8	0.1	-1.5	-1.5	0.0
B-C	-11.7	-11.6	0	-11.7	-11.5	-0.2
A-C	0.0	0.0	0.0	0.0	0.0	0.0
			-2.5			-13.9

[a] Energies given in kcal/mol. [b] A = substrate, residues Ace-Phe-Ser-Ile-NMe. [c] B = active site residues: 18-27, 64-69, 131-137, 157-160, 175-177, 204-207. [d] C = rest of the protein. [e] Ion-molecule conformation, tetrahedral constraints, C-S = 2.5 Å. [f] Oxyanion conformation, tetrahedral constraints, C-S = 2.5 Å. [g] Ion-molecule conformation, full minimization, no constraint. [h] Oxyanion conformation, full minimization, no constraint. [i] Sum of bond length, bond angle, and dihedral angle strain. [j] Van der Waals energy. [k] Electrostatic energy. [l] Hydrogen bond energy.

favorable, but since most of the energy difference is due to the internal energies of the reactants and not from better interaction between the substrate and the active site, we suggest that both conformations IM-Tet and OI-Tet may compete in the reaction pathway.

AM1 Calculations. Further study of the energetics for the potential reaction surface was performed using the semiempirical program AM1.[23] The minimized molecular mechanics structures IM-Tet and OI-Tet (Figures 7 and 8) were our initial models for the AM1 calculations. The coordinates for the bond-breaking residues of the substrate Ser and Ile, and the catalytically important active site residues (19, 25, 159) were extracted from the protein structures in both IM and OI conformations. The end residues Ace and NMe were added to each residue using the results from the molecular mechanics model. The resulting structure had 111 atoms.

To get a rough estimate if the AM1 energy differences between the OI and IM orientations were reasonable, the molecular mechanics geometry in which the carbon–sulfur bond distance in both conformations is 2.5 Å was run through a single SCF calculation. The energy difference between the structures is 20.5 kcal/mol, favoring the IM orientation. This energy difference is similar to the number obtained in the molecular mechanics calculations (16.2 kcal/mol). Both structures were then optimized. The parameters that were optimized were the bond lengths, angles, and dihedrals of the heavy atoms of Cys (active site residues), Gln (active site residue), Ser, and NMe (of the substrate at the reaction center). The rest of the parameters, including the heavy atoms of the His ring, end residues, and the hydrogens for each fragment were kept fixed.

The optimized structures are IM-AM1 (Figure 10a) and OI-AM1 (Figure 11). The energy difference between them is 3.0 kcal/mol favoring the structure in Figure 10a (Table IV). The important geometrical features in the two optimized structures are as follows.

a. Ion–Molecule Orientation (Figure 10). The S···C=O bond

Table III. Analysis Table of the Substrate and Active Site Interactions (kcal/mol) Divided into Groups

	noncovalent complexes 1M, OI					
	1M group 1[d]	OI group 1[d]	1M group 2[e]	OI group 2[e]	1M group 3[f]	OI group 3[f]
group 4 (18–24)						
NB[a]	−0.4	−0.4	−1.2	−1.5	−3.4	−1.5
EL[b]	0.4	0.4	1.5	−0.0	−0.1	−0.5
group 5 (25)[g]						
NB	−1.1	−0.1	−0.7	−0.8	−1.3	−1.0
EL	−7.3	−7.6	−5.7	−2.4	5.7	6.3
group 6 (26–27)						
NB	−1.8	−1.7	−0.3	−0.3	−0.5	−0.1
EL	−1.0	−1.4	0.5	0.1	−0.1	−0.3
group 7 (64–69)						
NB	−7.1	−7.6	−0.7	−1.0	−0.1	−0.1
EL	−6.0	−6.2	0.7	0.1	−0.2	−0.3
HB[c]	−0.8	−1.0	0.0	0.0	0.0	0.0
group 8 (131–137)						
NB	−2.0	−2.1	−0.1	−0.1	0.0	−1.9
EL	0.4	0.2	−0.1	−0.2	1.7	0.5
group 9 (157–160)[h]						
NB	−5.8	−5.2	−1.3	−1.3	−0.8	−6.4
EL	−1.3	−1.12	2.6	−1.6	−21.2	0.5
HB	0.0	0.0	0.2	−0.4	−0.7	−0.1
group 10 (175–177)						
NB	0.0	0.0	−0.0	−0.0	−2.2	−1.8
EL	0.0	0.0	0.2	−0.1	0.5	0.3
HB	0.0	0.0	0.0	−0.1	0.0	0.0
group 11 (204–207)						
NB	−0.3	−0.4	0.0	0.0	1.6	0.0
EL	0.1	0.1	0.0	0.0	0.0	0.0

Table IV. Total Energies (kcal/mol) of Active Site–Substrate Clusters, Calculated by the AM1 Semiempirical Quantum Mechanical Method

	1M-AM1[a]	OI-AM1[b]	1M-AM1-2[c]	OI-AM1-2[d]
E	−274.2	−271.2	−279.5	−285.0
relative energy	10.8	13.8	5.5	0.0
C–S bond length[e]	2.9	2.04	1.94	1.94

[a] Figure 10a. [b] Figure 11. [c] Figure 12. [d] Figure 13. [e] Bond length in ångströms.

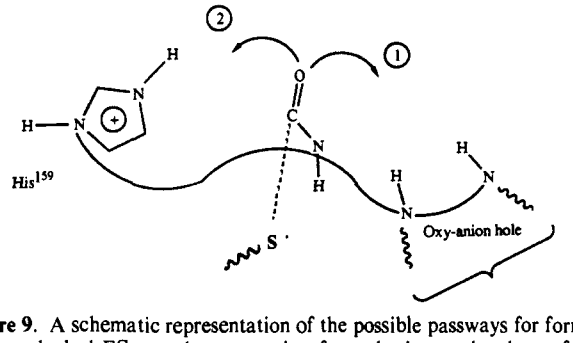

Figure 9. A schematic representation of the possible passways for forming tetrahedral ES complexes, starting from the ion–molecule conformation 1M. Route 1 would yield OT-Tet, and route 2 would yield 1M-Tet, corresponding to the observed ^{13}C NMR structures **2b** and **2a**, respectively, which are shown in Figure 2.

length is 2.9 Å. The central carbon becomes only slightly tetrahedral with an O–C–N angle of 119.1°. The amide nitrogen adjacent to the reaction center is planar (H–N–C = 120.1°). The distance between the imidazolium ion hydrogen and the oxygen of the substrate is 1.9 Å.

b. Oxyanion Orientation (Figure 11). The S···C=O bond length is 2.04 Å. The central carbon becomes tetrahedral with an O–C–N angle of 113.0°, and the hydrogen bonds that form the oxyanion hole, e.g., NH of the cysteine main chain and HN_2 of the Gln side chain, become very symmetrical, with a distance of 2.04 Å to the oxygen. An interesting geometrical change is a sharp pyramidization of the amide nitrogen together with a shortening of the distance between the hydrogen of the imidazolium ion and the nitrogen to 1.65 Å. To follow the potential surface of the reaction, three sets of reaction paths were performed. The histidine ring in IM-AM1 (Figure 10a) was allowed to optimize, resulting in a similar structure which is shown in Figure 10b, and which has a lower energy (−283.8 kcal/mol). Starting from the latter structure the C···S distance was gradually decreased from 2.9 to 2.0 Å. The energy of the system constantly increased.

Table V. Various Reaction Coordinates for the Hydrolysis Reaction of Papain (kcal/mol)

A[a]

	S···C=O bond distance[d]	energy	relative energy
1.	2.9	−283.8	0.0
2.	2.6	−279.0	4.8
3.	2.4	−274.2	9.6
4.	2.2	−266.1	17.7
5.	2.0	−255.0	28.8

B[b]

	C=O—H⁺···N(HIS) bond distance[d]	energy	relative energy	C—S bond length[d]
1.	0.99	−279.5	4.0	1.94
2.	1.3	−259.6	23.9	2.12
3.	1.4	−267.0	16.5	2.60
4.	1.5	−269.5	14.0	2.64
5.	1.7	−275.0	8.5	2.66
6.	1.94	−283.5	0.0	2.66

C[c]

	N···H···N bond distance[d]	energy	relative energy	C—S bond length[d]
1.	0.99	−271.5	13.5	2.1
2.	1.3	−275.0	10.0	1.95
3.	1.5	−279.3	5.7	1.94
4.	1.8	−285.0	0.0	1.9

[a] The reaction coordinate is the C–S bond that is being formed in the hydrolysis reaction (e.g., enzyme-S⁻—C=O bond). [b] The reaction coordinate is the axis for the proton-transfer reaction from the oxygen of the substrate 1M-AM1-2 (Figure 12) to the nitrogen of the imidazole ring of His-159 (e.g., C=OH⁺···N(His)). [c] The reaction coordinate is the axis for the proton-transfer reaction from protonated histidine in OI-AM1 (Figure 11) to the nitrogen of the substrate Ile (e.g., O=C—N···H···N(His⁺)). [d] Bond length in angströms.

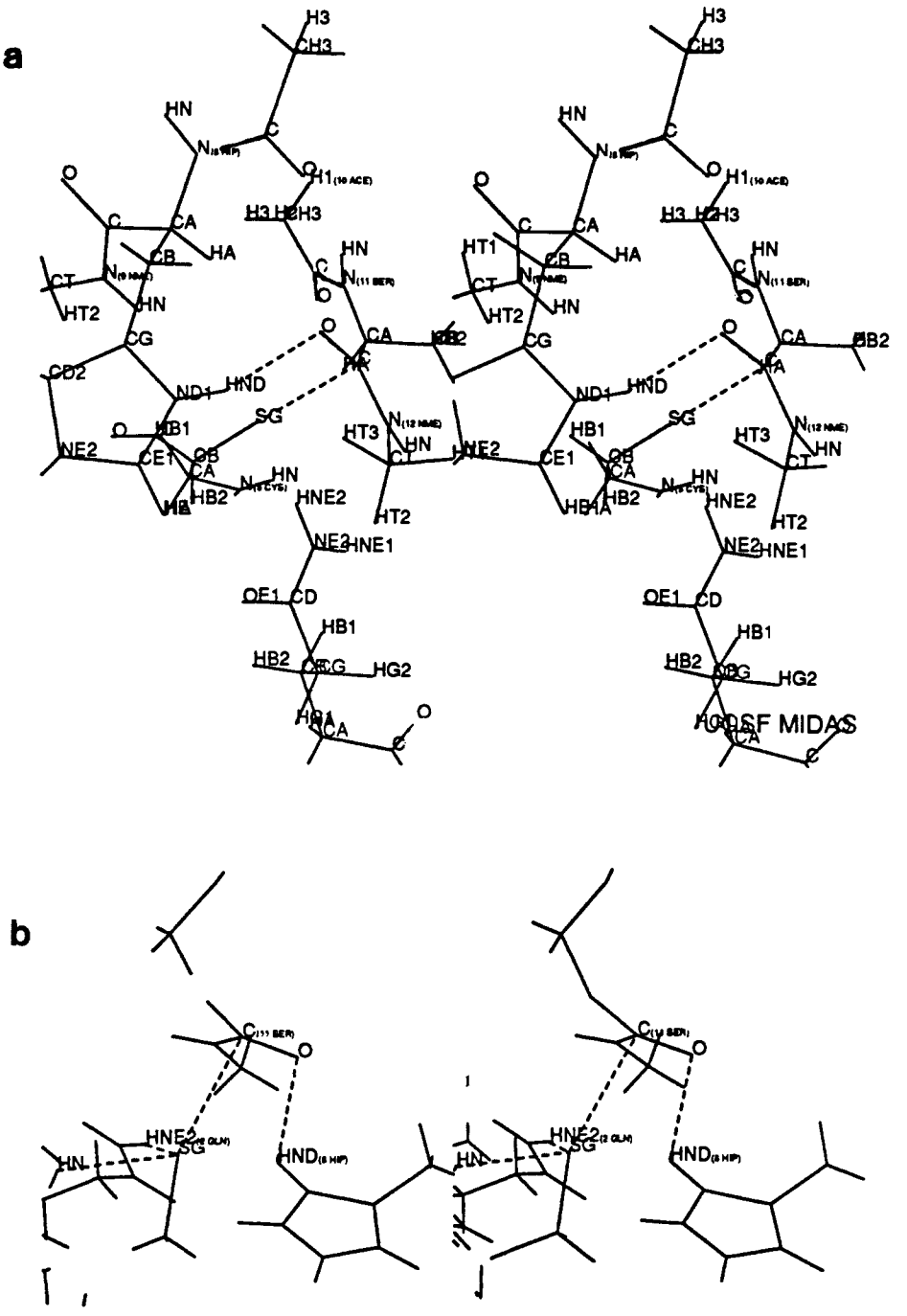

Figure 10. Stereoviews of IM-AM1, the active site + substrate cluster of the IM-Tet conformation, optimized by AM1. The cluster was constructed from the catalytic residues Gln-19, Cys-25, Hsp-159, and substrate residues Ser-215 and Ile-216. To form a neutral cluster, a beginning ACE (acetyl) and ending NME (*N*-methyl) groups were added to each of the residues. The initial coordinates for the cluster were extracted from the coordinate file of the IM-Tet conformation. In (a) the imidazolium ring parameters were kept fixed and the heavy atoms in the rest of the residues parameters were optimized. In (b) the imidazolium parameters were released as well, which reduced the energy of the system, but the geometry did not changed significantly. The C–S distance in both a and b is 2.9 Å and C=O···H (His) is 2.2 in a and 2.3 in b. b, however, is not a true minima on the potential surface; i.e., when His parameters are released, the C–S bond distance constantly increases the energy of the system drops.

At a sulfur carbon distance of 2.0 Å the energy of the system is 28.8 kcal/mol higher than the optimized structure at 2.9 Å (Table V, entry a). This result is consistent with the ab initio study results of Howard and Kollman,[12] which showed that in the reaction, $SH^- + H_2NCHO \rightarrow [H_2N(HS)CHO^-]$, at a S···C distance of 2.0 Å the energy of the system is 29.0 kcal/mol higher than the energy of the reactants.

Starting with the optimized structure that was obtained at point 4 on the reaction path a (with C···S = 2.2) the ϵ N···H distance in HIS was increased from a standard bond length of 0.99 to 1.3 Å. The bond, angle, and dihedral parameters of the heavy atoms of His were then released as well, and the geometry was optimized. The optimized structure, IM-AM1-2 (Figure 12), is 5.3 kcal/mol lower in energy than IM-AM1 (Figure 10a, Table IV). The important geometrical change is a complete proton transfer from the imidazolium ring to the oxygen yielding a neutral tetrahedral thioketal structure. The C···S distance in the optimized structure is 1.94 Å and the carbon is completely tetrahedral.

To evaluate the energy barrier for the proton-transfer reaction from the charged His ring to the oxygen of the substrate, the C=O···H^+ distance in Figure 12 was varied from a regular O···H bond (0.99) to 1.94 Å. Thus, the reaction coordinate is the C=O—H^+···N(HIS) bond forming and breaking (entry b in Table V). We find that the barrier for the proton transfer from His to the substrate oxygen is 23.9 kcal/mol. Thus, our assumption that there is a proton transfer from His^+ to the oxygen of the substrate prior to the attack of the sulfur does not seem likely at the first glance. Nevertheless, the coordinate that we have chosen

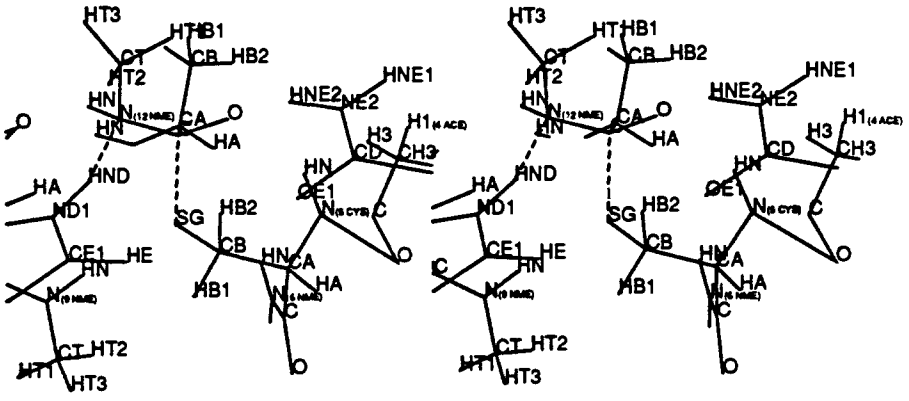

Figure 11. Stereoview of OI-AM1, the AM1 cluster constructed from OI-Tet (see Figugure 10). All heavy atoms in catalytic residues are being optimized except those of the begining and ending residues. The C–S distance in the optimized structure is 2.1 Å and CO—N···H(His) is 1.65 Å.

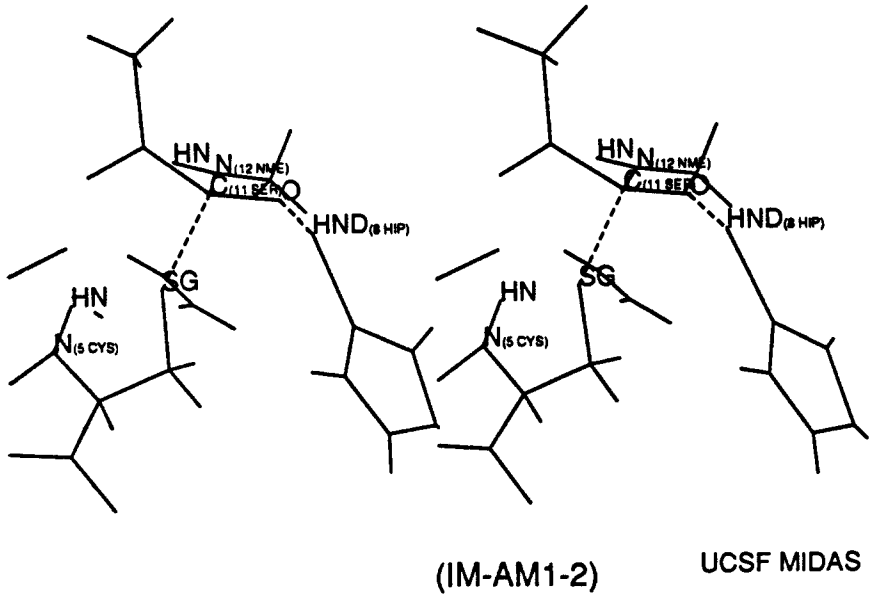

Figure 12. Stereoview of IM-AM1-2, which is a final point on the reaction coordinate for a proton transfer from histidine to the carbonyl oxygen of the substrate (Ser). The C–S distance is 1.98 Å, and the C=O···H^+ distance is 0.97 Å. Note that the carbon at the reaction center is tetrahedral and C–S bond shortened in comparison to IM-AM1.

forces the hydrogen to lie all the way in a straight line going from nitrogen to oxygen, and this route may not be the energetically preferred one. Considering the small energy difference (4 kcal/mol) between the starting and ending structures (e.g., IM-AM1-2 and point 6 on the reaction path b in Table V) and assuming that there are, as in other proton-transfer reactions, tunneling effects that lower the potential energy barrier suggest that a protonation of the substrate oxygen via a proton transfer from HIS^+ is a possible pathway.

We observed that in OI-AM1 the His—H^+···NH(substrate) distance is only 1.65 Å. When the parameters of the imidazolium ring in OI-AM1 are allowed to optimize, the energy of the system drops and it optimizes without a barrier to OI-AM1-2, in which a complete proton transfer from the His to the nitrogen has taken place. Entry c in Table V shows some points on this reaction surface. OI-AM1-2 (Figure 13) is the lowest energy point on the reaction surface that we calculated. The energy difference between 13 and the oxygen protonated species 11 is 5.5 kcal/mol. Our results show that of the three reaction paths we calculated, the "simple" nucleophilic attack of sulfur on carbonyl, without the participation of the imidazolium hydrogen, is not likely to occur. The preferred pathway is a nucleophilic attack on a protonated substrate. Of the two possible pathways the reaction proceeds more easily when the sulfur approaches the substrate in its OI conformation, followed by a proton transfer from histidine to the pyramidized nitrogen. We do not exclude, however, the second pathway (entry b in Table V) in which the proton is transferred

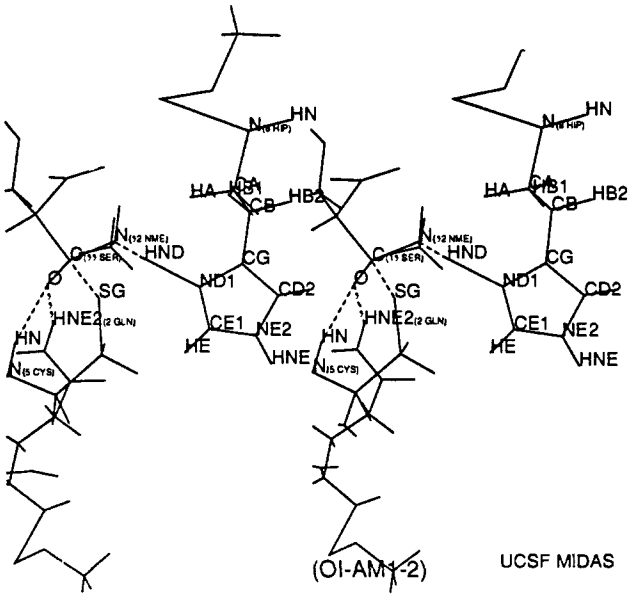

Figure 13. Stereoview of OI-AM1-2, which is a final point on the reaction coordinate for a proton transfer from histidine to the nitrogen of the substrate. The C–S distance is 1.88 Å, and the CO—N···H^+ distance is 1.02 Å. The hydrogen bonding distances of the oxyanion holes are C=O···HN(Cys) = 1.97 Å and C=O···HNE2(Gln) = 2.07 Å.

to the oxygen from a more remote position, when the C···S distance is 2.66 Å.

Discussion and Conclusions

We have presented molecular mechanical and semiempirical quantum mechanical calculations on models of noncovalent and covalent substrate complexes with papain. In the molecular mechanical calculations, we have shown that a structure (IM) which involves a substrate amide N–H···Cys-25/S^- interaction, similar to that found in the ion–molecule complex of SH^-··· formamide, is 29 kcal/mol more stable than a structure which has the substrate carbonyl group pointing in the direction of the oxyanion hole (OI). Of course, given our simple representation of solvent and electrostatic energies, this energy value should not be interpreted quantitatively. Nonetheless, it is likely that, because of the ion pair in its active site, substrates of papain might bind rather differently than those of the serine proteases.

We have then used molecular mechanics with constraints to force an attack of the Cys-25–S^- bond on the carbonyl carbon of the substrate to construct tetrahedral structures from the above two noncovalent complexes. The total energy difference still favors the IM-like structure over OI, but by significantly less than in the Michaelis complex, in part due to the oxyanion hole stabilization of the substrate C–O in the OI structure. Again, one should not interpret these energy differences too quantitatively and realize that both complexes are candidates to be the true "observed" structure.

We have then used the above two tetrahedral structures as a basis for semiempirical quantum mechanical calculations on a fragment of the protein and substrate. The results show, consistent with an earlier ab initio model study, that attack of S^- on a carbonyl carbon does not involve a stable anionic tetrahedral structure, rather that proton transfer to the substrate carbonyl oxygen or the substrate amide nitrogen should proceed or be concerted with S^- attack to form a stable tetrahedral adduct. At this time, the calculations are not accurate enough to decide which occurs. Protonation of isolated amides is known to occur on the oxygen, but the pK of this protonation is near 1.[28] If this occurs in the enzyme, one would have a stable tetrahedral adduct, but the mechanism for breakup to form the acyl enzyme is not clear. On the other hand, protonation of the nitrogen would occur only if the amide nitrogen became tetrahedral, presumably due to the presence of the S^- near the carbonyl carbon. (The calculations show that, even though the energy rises for S^- attack on the carbonyl carbon, the amide nitrogen becomes tetrahedral during this process.) Nitrogen protonation would lead to the acyl enzyme, since it makes N–C bond breakage facile. AM1 favors the nitrogen protonated structure, and the error it makes in the relative proton affinities of amide and imidazole nitrogen would suggest that it would tend to calculate oxygen protonation more favorable compared to nitrogen than that observed. Thus, on balance we favor a mechanism by which the sulfur can continually approach the carbonyl carbon, but the potential is repulsive until accompanied by His···amide N proton transfer.

In summary, our calculations demonstrate reasonable alternative binding modes for substrates interacting with papain, as well as suggesting two variations in the accepted mechanism for papain-catalyzed hydrolysis. These calculations lead us to speculate that papain has an Asn H bonded to its His, rather than an Asp as the serine proteases because an Asn does not hold onto His^+ as tightly and allows the His to be in a position to deliver the proton effectively once the S^- begins to approach the substrate carbonyl carbon. In the serine proteases, Asp is H bonded to His so it can facilitate the key step of proton abstraction of His from Ser. Thus, the key role of His in the serine proteases is proton abstraction, but in the cysteine proteases, it is proton delivery. The fact that a Cys–His–Asn triad in trypsin is quite inactive[29] does not rule out this explanation, given the H bonding network that stabilizes the Asp-102 in trypsin and which leads as Asn-102 to point its NH_2 toward the His.[30] Such a configuration would likely prevent His^+–Cys^- ion pairing. The key experiments would be the substitution of Asn for Asp in papain by site specific mutation and a study of its structure and catalytic activity. The above speculations would suggest that, even if the structure of the active site is preserved, the catalytic activity might be greatly reduced.

Acknowledgment. We are pleased to acknowledge research support from the NIH (GM-29072 to P.A.K. and RR-1081 to R.L.). The facilities of the UCSF Computer Graphics Laboratory, supported by RR-1081, are gratefully acknowledged as is the assistance of Allison Howard, George Seibel, and Eric Petterson. The coordinates of the various structures presented here can be obtained upon request to the authors.

Registry No. Ace-Phe-Ser-Ile-NMe, 124043-71-6; S, 7704-34-9; papain, 9001-73-4; cysteine proteinase, 37353-41-6; $[Fe_2O\{O_2P(OC_6H_5)_2\}_2(HBpz_3)_2]$, 96502-34-0; $[Fe_2O\{O_2P(C_6H_5)_2\}_2(HBpz_3)_2]$, 124042-43-9; $HO_2P(OC_6H_5)_2$, 838-85-7; $[Fe_2O(O_2CCH_3)_2(HBpz_3)_2]$, 86177-70-0; Fe, 7439-89-6; O_2, 7782-44-7; diphenylphosphinic acid, 1707-03-5.

(28) Perrin, D. D.; Dempsey, B.; Sergeant, E. P. In *pK_a Prediction for Organic Acids and Bases*; Chapman and Hall: London, 1981.

(29) Higaki, J. N.; Gibson, B. W.; Craik, C. S. *Cold Spring Harbor Symp. Quant. Biol.* **1987**, *32*, 615–621.

(30) Sprang, S.; Standing, T.; Fletterick, R. J.; Stroud, R. M.; Finer-Moore, J.; Xuong, N.-H.; Rutter, W. J.; Craik, C. S. *Science* **1987**, *237*, 905–909.