Protein Science (1996), *5*:1355–1365. Cambridge University Press. Printed in the USA.

A theoretical study of the active sites of papain and S195C rat trypsin: Implications for the low reactivity of mutant serine proteinases

ALLAN J. BEVERIDGE

Daresbury Laboratory, Daresbury, Warrington, Cheshire, United Kingdom

(RECEIVED August 9, 1995; ACCEPTED May 1, 1996)

Abstract

The serine and cysteine proteinases represent two important classes of enzymes that use a catalytic triad to hydrolyze peptides and esters. The active site of the serine proteinases consists of three key residues, Asp···His···Ser. The hydroxyl group of serine functions as a nucleophile and the imidazole ring of histidine functions as a general acid/general base during catalysis. Similarly, the active site of the cysteine proteinases also involves three key residues: Asn, His, and Cys. The active site of the cysteine proteinases is generally believed to exist as a zwitterion ($Asn\cdots His^{+}\cdots Cys^{-}$) with the thiolate anion of the cysteine functioning as a nucleophile during the initial stages of catalysis. Curiously, the mutant serine proteinases, thiol subtilisin and thiol trypsin, which have the hybrid Asp···His···Cys triad, are almost catalytically inert. In this study, ab initio Hartree–Fock calculations have been performed on the active sites of papain and the mutant serine proteinase S195C rat trypsin. These calculations predict that the active site of papain exists predominately as a zwitterion ($Cys^{-}\cdots His^{+}\cdots Asn$). However, similar calculations on S195C rat trypsin demonstrate that the thiol mutant is unable to form a reactive thiolate anion prior to catalysis. Furthermore, structural comparisons between native papain and S195C rat trypsin have demonstrated that the spatial juxtapositions of the triad residues have been inverted in the serine and cysteine proteinases and, on this basis, I argue that it is impossible to convert a serine proteinase to a cysteine proteinase by site-directed mutagenesis.

Keywords: catalytic triad; cysteine proteinase; Hartree–Fock; serine proteinase

The ubiquitous occurrence of catalytic triads in many hydrolytic enzymes emphasizes the versatility and efficiency of this structural motif in catalysis. There are no less than three distinct groups of hydrolytic enzymes (the chymotrypsin family of proteases, the cysteine proteases, and the α/β hydrolase fold enzymes) that use a catalytic triad at the active site. The overall three-dimensional structure of each class is very different, indicating that each has evolved as a result of convergent evolution. The serine and cysteine proteinases have been studied extensively because, although these two enzymes catalyze similar reactions (hydrolysis of peptides and esters) they exhibit some significant structural differences. The catalytic triad of the serine proteinases consists of Ser-195, His-57, and Asp-102; in the native enzyme, the imidazole ring of H57 is neutral and the hydroxyl group of S195 functions as the attacking nucleophile during the initial stages of catalysis. In the cysteine proteinases, the catalytic triad consists of Cys-25, His-159, and Asn-175. However, in contrast to the serine proteinases, experimental studies (Lewis et al., 1976, 1981; Creighton et al., 1980; Johnson et al., 1981) have shown that the active site exists as a zwitterion; the nucleophilic thiol group of Cys-25 is ionized and the catalytic histidine is protonated.

The catalytic mechanisms of both enzymes are very similar. In the serine proteinases, the first step in catalysis involves nucleophilic attack by the hydroxyl group of Ser-195 on the carbonyl carbon of the substrate, resulting in the formation of a tetrahedral covalent enzyme/substrate intermediate. His-57 is believed to function as a base that enhances the reactivity of Ser-195 by abstracting the hydroxyl proton in a concerted manner during the initial stages of nucleophilic attack (Fig. 1). The second step involves the protonated imidazole ring of His-57 transferring a proton (which was initially abstracted from Ser-195) to the scissile nitrogen of the substrate, thus initiating cleavage of the scissile bond and resulting in the formation of a covalent acyl-enzyme intermediate. The amine portion of the substrate then diffuses away from the active site and is replaced by a water molecule that reacts with the covalent acyl-enzyme intermediate to form a second tetrahedral intermediate before breaking

Reprint requests to: Allan J. Beveridge, Daresbury Laboratory, Daresbury, Warrington, Cheshire, United Kingdom; e-mail: a.j.beveridge@dl.ac.uk.

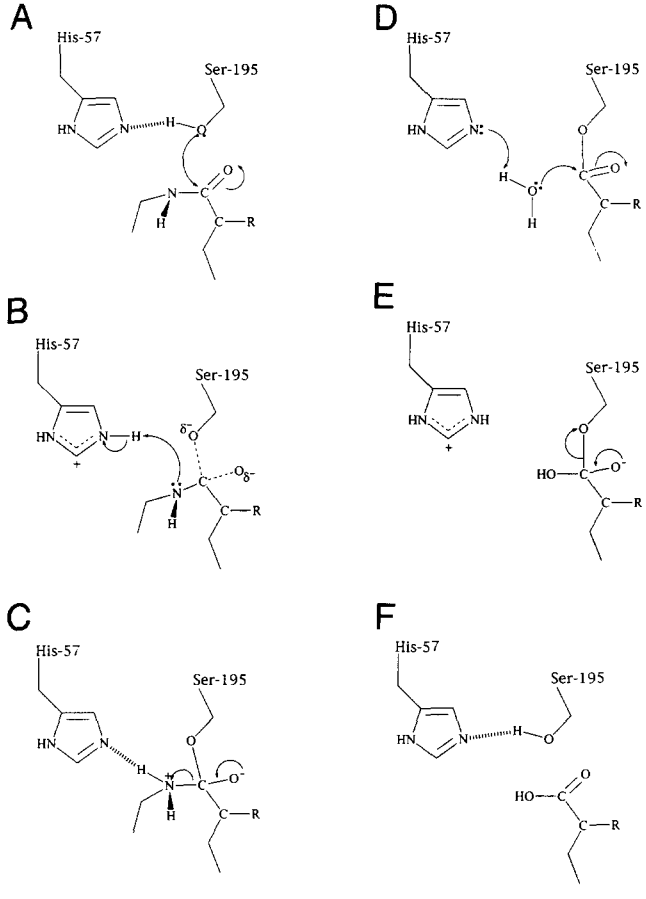

Fig. 1. Schematic representation of the catalytic mechanism of the serine proteinases (numbering of residues with respect to trypsin). Nucleophilic attack by the side chain of Ser-195 results in the formation of a covalent enzyme–substrate intermediate (**A,B,C**). Cleavage of the scissile bond (C) leads to the formation of an acyl-enzyme intermediate, which is subsequently hydrolyzed by a water molecule (**D,E,F**).

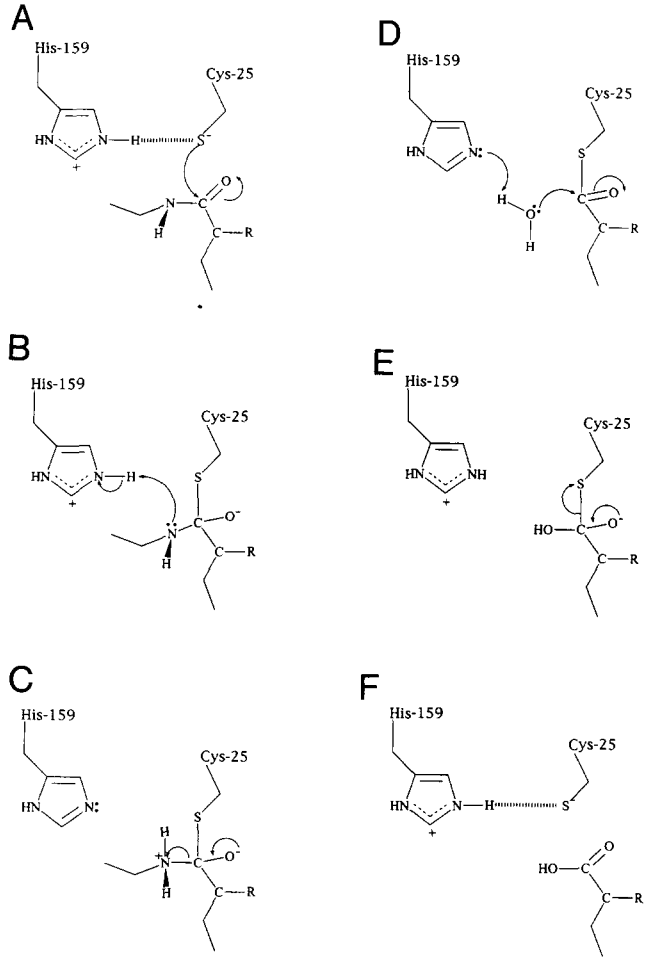

Fig. 2. Schematic representation of the catalytic mechanism of the cysteine proteinases (numbering of residues with respect to papain). Unlike the serine proteinases, the active site exits as a zwitterion ($H159^{+}\cdots Cys25^{-}$). Nucleophilic attack by the ionized thiol of Cys-25 generates a covalent enzyme–substrate intermediate (**A,B,C**). Cleavage of the scissile bond (C) leads to the formation of an acyl-enzyme intermediate, which is subsequently hydrolyzed by a water molecule (**D,E,F**).

the Ser-195 Og-$C_{substrate}$ bond and restoring the enzyme to its native state (Fig. 1).

The catalytic mechanism of the cysteine proteinases shares many common features with the serine proteinases. Initially, the active site exists as a zwitterion; the thiol group of cys-25 is ionized and the imidazole ring of His-159 is protonated. In the first stage of catalysis, the highly reactive thiolate anion of C25 attacks the scissile carbon of the bound substrate yielding a tetrahedral covalent substrate/enzyme intermediate. In an analogous manner to the serine proteinases, the protonated imidazole ring of His-159 then transfers a proton to the scissile nitrogen of the substrate. This initiates the cleavage of the scissile bond; the amine portion of the substrate diffuses away and is replaced by a water molecule which then hydrolyzes the resulting acyl-enzyme intermediate to restore the enzyme to its original state (Fig. 2).

Structural comparisons of the serine and cysteine proteinases have shown that the active site residues of chymotrypsin (Asp$\cdots$His$\cdots$Ser) and papain (Asn$\cdots$His$\cdots$Cys) are virtually superimposable (0.8 Å RMS deviation) (Garavito et al., 1977). This result, coupled with the similarity of the two catalytic mechanisms, has prompted several groups to study the feasibility of converting proteases containing a nucleophilic serine (e.g., subtilisin and trypsin) to a cysteine proteinase by using site-directed mutagenesis. Mutating the catalytic serine to cysteine yields the catalytic triad Cys$\cdots$His$\cdots$Asp, which is a hybrid of the triads found in the serine (Ser$\cdots$His$\cdots$Asp) and cysteine proteinases (Cys$\cdots$His$\cdots$Asn). Because oxygen and sulfur both belong to group 6A in the periodic table and possess similar chemical properties, the resulting thiol enzymes were expected to behave as efficient proteases. However, experiments on the mutant enzymes thiol subtilisin (Neet & Koshland, 1966; Polgar & Bender, 1967) and thiol trypsin (Higaki et al., 1989) have demonstrated that these mutants are inactive toward normal ester and amide substrates.

Several reasons have been postulated for the low reactivity of the mutant serine proteinases, including nonproductive collapse of the tetrahedral intermediate (Wilke et al., 1991), blocking of the oxyanion hole (Wilke et al., 1991), and an unsuitable geometry for the thiolate-imidazolium ion pair (Halasz & Polgar, 1977). However, the most plausible reason may be the inability

of the mutant serine proteinases to form thiolate anions in sufficiently large concentrations. Experimental studies on the acid-catalyzed hydrolysis of trifluoroacetic thiolesters (CF_3COSR) have shown that the leaving group is RS^-; Fersht (1971) has argued that, on the basis of the principle of microscopic reversibility, the reverse reaction (thiolysis of the acid) must occur by nucleophilic attack of the thiolate anion. This result demonstrates that thiolate anions are much more reactive as nucleophiles than the corresponding thiols. Moreover, experimental studies (Lewis et al., 1976, 1981; Creighton et al., 1980; Johnson et al., 1981) on papain have shown that the catalytic triad $Cys^- \cdots His^+ \cdots Asn$) is ionized in the native enzyme, suggesting that a reactive thiolate anion is needed immediately prior to catalysis. Further support for this hypothesis is provided by theoretical and experimental studies on dienelactone hydrolase (DLH). DLH is the only known enzyme with a naturally occurring Cys···His···Asp triad (Pathak & Ollis, 1990). Crystallographic studies on native DLH and several mutant enzymes (Cheah et al., 1993) suggest that the active site of DLH exists in two distinct configurations (Fig. 3A,B). In the native enzyme, the active site adopts a catalytically inert configuration, the neutral thiol group of Cys-123 points away from the active site and forms a hydrogen bond to the carboxylate of E36 (Fig. 3A). When the substrate (dienelactone) binds, several small but important conformational changes occur in the active site. In particular, the carboxylate of E36 moves closer to the side chain of Cys-123, enabling E36 to abstract the thiol proton, thus generating a highly reactive thiolate anion that subsequently attacks the substrate. Kinetic studies on E36N DLH have shown that mutating E36 to alanine results in a catalytically inert enzyme (Cheah et al., 1993). This result demonstrates that E36 plays an essential role in catalysis, and recent ab initio calculations on the active site of DLH also support the substrate-induced activation mechanism described above (Beveridge & Ollis, 1995).

In the present study, Hartree–Fock calculations have been performed on the active sites of papain and S195C rat trypsin to determine whether the active site is neutral (Cys···His···X) or ionized ($Cys^- \cdots His^+ \cdots X$) in each case. Similar calculations on the active site of papain (van Duijnen et al., 1980; Dijkman et al., 1989; Rullman et al., 1989) have been reported, but the results were not completely satisfactory. In particular, the relative stability of the zwitterionic configuration ($Cys^- \cdots His^+ \cdots Asn$) was found to be dependent on both the basis set and the number of active site residues explicitly included in the calculations. The rapid improvement in computer hardware and the development of DIRECT SCF methods has now made it possible to perform more extensive SCF calculations (up to 10 residues) on the active site of papain. The calculations predict that the active site exists predominately as a zwitterion. However, similar calculations on the active site of S195C rat trypsin, a mutant with the same catalytic triad as DLH, predict that the neutral Cys···His···Asp configuration is significantly more stable than the corresponding zwitterionic form ($Cys^- \cdots His^+ \cdots Asp$), suggesting that in the mutant serine proteinases the formation of a reactive thiolate anion prior to catalysis would be extremely difficult. The results obtained for papain and S195C rat trypsin indicate that there are several important structural differences between the active sites of each enzyme. In particular, I have demonstrated that the spatial juxtaposition of the triad residues have been inverted in the serine and cysteine proteinases and, on this basis, it is not possible to mutate a serine proteinase to a cysteine proteinase.

Theoretical methodology

All calculations were performed using the ab initio quantum chemistry package Gaussian92 (Frisch et al., 1992). Correlation energies were calculated using the many body perturbation

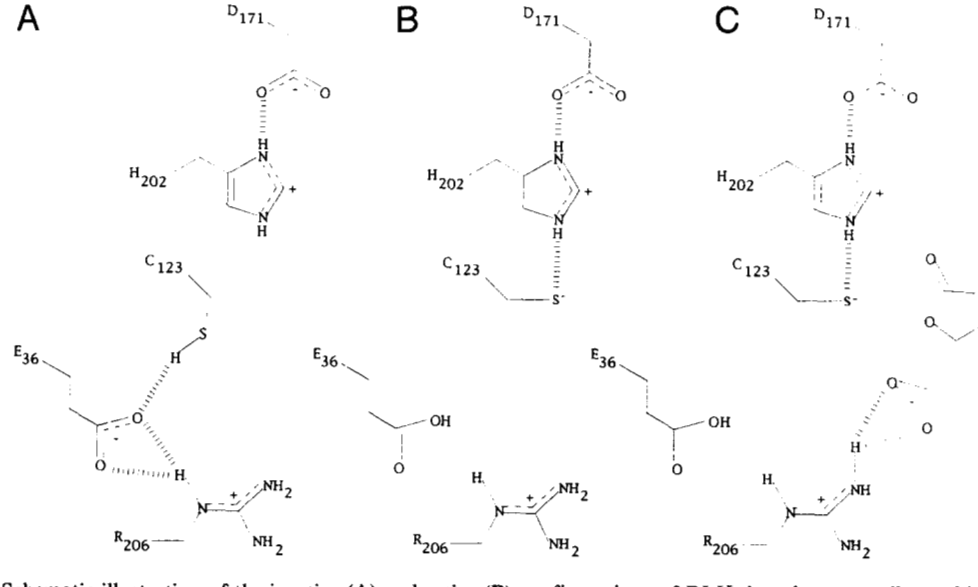

Fig. 3. Schematic illustration of the inactive (**A**) and active (**B**) configurations of DLH, based on crystallographic studies of native DLH and inhibitor/enzyme complexes (Cheah et al., 1993). The active form is generated by $E36^-$ abstracting the C123 thiol proton; the reactive thiolate anion then rotates by ~180 degrees, where it is stabilized by the electropositive field generated by $H202^+$. **C**: Schematic representation of the activated substrate–enzyme complex.

method of Moller and Plesset (1934) at the MP2 level; only the valence electrons were included in the calculations, all core electrons and the corresponding antibonding core orbitals were omitted. All geometry optimizations were performed with the Berny conjugate gradient algorithm in Gaussian92 (Schlegal, 1989).

Calculations were initially performed using two basis sets, 4-31G and 6-31G*. However, in a similar theoretical study on the active site of DLH, Beveridge and Ollis (1995) observed a significant basis set dependency. Hartree–Fock calculations on the native enzyme with the split valence 4-31G basis set consistently favored a zwitterionic active site ($Cys^- \cdots His^+$). A similar basis set dependency was observed in an earlier theoretical study on the active site of papain (Rullman et al., 1989). In this study, thiomethanol, imidazole, and formamide were used to represent the catalytic triad (C25, H159, and N175). Calculations with a double zeta basis set (Roos & Siegbahn, 1970a) favored a zwitterionic active site ($Cys^- \cdots His^+ \cdots Asn$), whereas calculations with an extended basis set, which included polarization functions (Roos & Siegbahn, 1970b), favored the neutral form ($Cys \cdots His \cdots Asn$). Therefore, in the present study, I report only the results from the larger (and more reliable) 6-31G* basis set.

Model construction

Active site of papain

We have used the high-resolution (1.65 Å) crystal structure reported by Kamphuis et al. (1984) to construct models for the active site of papain. This structure corresponds to an inactive form of papain because the thiol group of Cys-25 is oxidized to SO_3 in the X-ray beam. The most significant effect of this oxidation is a distortion of the active site due to nonbonded interactions between the SO_3 group and the imidazole ring of His-159. Drenth et al. (1975) have argued that, in the catalytically active form of papain, the imidazole ring is rotated by about 30° compared with the crystal structure. I have accounted for these effects by optimizing the orientation of the imidazole ring. Active site models were initially constructed using H_2NCO CH_2CH_2Im and $H_2NCOCH_2CH_2ImH^+$ to represent His-159 and the peptide backbone of Ala-160 for the neutral and zwitterionic forms of the active site, respectively. The gas phase geometries of each His-159/Ala-160 moiety were optimized, using a 4-31G basis set, by restraining all torsion angles to correspond to those observed in the crystal structure. Thio-methanol (CH_3SH) and the corresponding thiolate anion (CH_3S^-) were used to represent the side chain of Cys-25. Formamide was used to model the side chain of Asn-175. The optimized geometry of each small molecule was overlayed onto the relevant crystallographic coordinates. Using the 4-31g basis set, the optimum hydrogen bond formed between Cys-25 and His-159 was determined by optimizing the torsion angles corresponding to the imidazole ring of His-159 and the cysteine sulfur. For the neutral form, the torsion angle corresponding to the acidic thiol proton was also optimized. The optimized geometry of the zwitterion shows a significant change in the orientation of the imidazole ring; the ring has rotated by 31 degrees when compared with the crystal structure. This is consistent with the catalytic mechanism proposed by Drenth et al. (1975). The optimized geometry of the neutral structure shows a much smaller change in the orientation of the imidazole ring (7 degrees).

More accurate calculations using larger models for the active site were performed using the larger 6-31G* basis set. In each case, the active site model was constructed by using a least-squares algorithm to overlay a set of small molecules onto the relevant crystallographic coordinates of papain (Kamphuis et al., 1984). The active site models and the set of molecules used in each case are shown in Figure 4 and Table 1. The gas phase geometries of each small molecule were optimized with the 6-31G* basis set. Model 1 was constructed as follows. The optimized gas phase geometries of $H_2NCOCH_2CH_2Im$, $H_2NCOCH_2CH_2$ ImH^+, thio-methanol, and CH_3S^- were used to construct models for the neutral and zwitterionic configurations; however, the torsion angle of the imidazole ring was fixed to the value determined previously in the 4-31g optimizations. The optimum hydrogen bond between Cys-25 and His-159 was then determined by optimizing the torsion angles involving the cysteine sulfur and the acidic thiol proton. A formamide molecule was added to represent the side chain of Asn-175. The results are shown in Table 1.

Previous calculations reported by Rullman et al. (1989) demonstrated that the backbone of Cys-25 also plays an important role in stabilizing the zwitterionic form. They modeled the cysteine backbone using $H_2NC_\alpha HCHO$. Inclusion of the backbone preferentially stabilized the zwitterionic form by 29 kJ mol^{-1} (6.9 kcal mol^{-1}). Inspection of the crystal structure of papain shows that the imido protons of Cys-25 and Trp-26 are both very close to the cysteine sulfur and should therefore help stabilize the zwitterion. I have used $H_2NCO\text{-}C_\alpha H(C_\beta H_2\text{-}S_\gamma^-)\text{-}NHCOH$ (and $H_2NCO\text{-}C_\alpha H(C_\beta H_2\text{-}S_\gamma H)\text{-}NHCOH$) to model Cys-25 and the peptide backbone of Trp-26. The geometries of the Cys-25/Trp-26 moieties were optimized in the gas phase by restraining the torsion angles to correspond to those observed in the crystal structure. The thio-methanol molecules used to represent Cys-25 in model 1 were then replaced by the optimized Cys-25/Trp-26 moieties. The resulting structures were then re-optimized. The

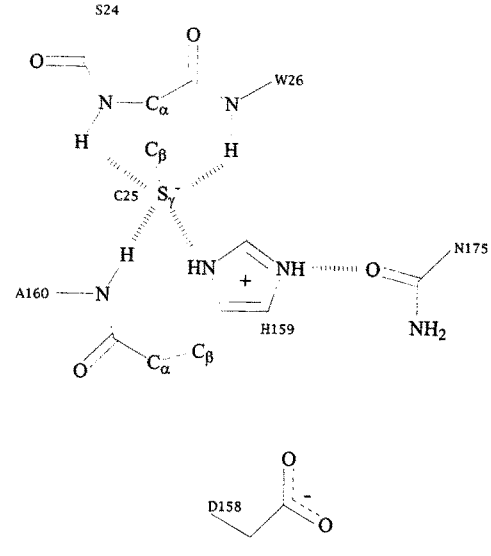

Fig. 4. Schematic representation of the molecules that were used to model the active site of papain.

Table 1. *Calculated SCF energies (a.u) for the active site of papain*[a]

Model	Residues[b]	6-31G*
1	C25, H159/A160, N175	−1,116.35297
		−1,116.30714 (28.8)
2	C25/W26, H159/A160	−1,282.95728
		−1,282.92780 (18.5)
		−1,285.60469[c]
		−1,285.57829[c] (16.6)
3	C25/W26, H159/A160, N175	−1,451.90018
		−1,451.87209 (17.6)
4	C25/W26, H159/A160 N175,D158	−1,640.08699
		−1,640.07626 (6.7)
5	C25/W26, H159/A160, N175, D158, G19	−1,809.02060
		−1,809.01411 (4.1)
6	C25/W26, H159/A160, N175, D158, G19,S176	−1,924.05730
		−1,924.05162 (3.6)
7	C25/W26, H159/A160, N175, D158, G19,S176, Helix-L1	−1,929.16858 (1.1)
		−1,929.24448
		−1,929.24562[d]
		−1,929.24448[d] (0.7)
7a		−1,929.72609[e]
		−1,929.70952[e] (10.4)
8	C25/W26, H159/A160, D175, D158, G19,S176, Helix-L1	−1,948.41365 (14.9)
		−1,948.43745

[a] Energies (kcal mol^{-1}) relative to the most stable configuration are shown in parentheses. The top number corresponds to the neutral form (Cys···His···Asn) and the bottom number refers to the zwitterion (Cys^-···His^+···Asn).

[b] C25 = CH_3SH or CH_3S^-, H159/A160 = H_2NCO-CH_2-CH_2-Im or H_2NCO-CH_2-CH_2-ImH^+, C25/W26 = H_2NCO-$C_\alpha H(C_\beta H_2$-$S_\gamma H)$-$CONH_2$ or H_2NCO-$C_\alpha H(C_\beta H_2$-$S_\gamma^-)$-$CONH_2$, N175 = $CONH_2$, D158 = H_2COO^-, G19 = $CONH_2$, S176 = CH_3OH.

[c] 6-31g*/MP2 calculation using only the valence electrons.

[d] 6-31g**/SCF calculation using 6-31g* optimized geometry.

[e] D158 modeled as formic acid.

bond lengths, angles, and torsion angles of the Cys-25 and Trp-26 imido protons, the torsion angle of Cys-25 S_γ, and the torsion angle of the acidic thiol proton were allowed to vary; all other geometric parameters were fixed. Model 3 was constructed by adding formamide to model 2 to represent the side chain of Asn-175.

Model 4 was constructed by using a formate anion to represent the side chain of Asp-158. Earlier theoretical studies have indicated that Asp-158 plays an important role in stabilizing the zwitterion (Dijkman et al., 1989). Furthermore, kinetic studies on papain suggest that an ionizable group with $pK_a < 4$ proximal to the active site plays an important role in catalysis (Willenbrock & Brocklehurst, 1985); the authors argue that Asp-158 is the most likely candidate. I therefore constructed model 4 by adding a formate anion (representing Asp-158) to model 3. A theoretical model of the enzyme/substrate complex (Bolis et al., 1978) constructed from the crystal structure of a chloromethyl ketone inhibitor complexed with papain (Drenth et al., 1976) indicates that Gly-19 facilitates substrate binding by forming a hydrogen bond to the carbonyl group of the scissile bond. Therefore, model 5 was constructed by adding a formamide molecule to mimic the peptide backbone of Gly-19. The electrostatic field generated by the hydroxyl side chain of Ser-176 (which forms a hydrogen bond to Gly-19) may also influence the relative stabilities of the neutral and zwitterionic configurations, and model 6 was constructed by using methanol to represent the side chain of Ser-176. For model 7, I have used a set of point charges, taken directly from the AMBER united atom force field (Weiner et al., 1984), to represent the effects due to the L1 α-helix. This helix points toward the active site and the dipole moment generated by the helix should stabilize the zwitterion. By representing residues 27–42 as atom-centered point charges, one can simulate the electrostatic and polarization effects due to the helix.

Model 7A was constructed by replacing the formate anion representing the side chain of Asp-158 in model 7 with formic acid. This was done to assess whether the influence of Asp-158 is due to the negative charge or the polar nature of the side chain. In model 8, the effect of "mutating" Asn-175 to an aspartate by replacing the formamide moiety representing Asn-175 in model 7 with a formate anion was investigated.

Active site of native S195C rat trypsin

Active site models for native S195C rat trypsin

Models for the active site were constructed using a high-resolution crystal structure (1.6 Å) of rat trypsin (Higaki et al., 1989). A set of small molecules was used to model the side chains of key active site residues (Fig. 5). Thioethanol and the corresponding thiolate anion were used to model the catalytic cysteine (C195). Imidazole and the imidazolium cation were used to model the triad histidine (H57) and a formate anion ($HCOO^-$) was used to model anionic side chain of Asp-102. Inspection of the crystal structure shows that Ser-214 is hydrogen bonded to D102; therefore, I have modeled the side chain of S214 using methanol. The geometries of each of these small molecules were optimized initially in the gas-phase using a 6-31G* basis set.

The simplest model (model 1) for the active site was constructed using the side chains of the three triad residues. The neutral configuration was constructed by using a least-squares algorithm to overlay the optimized coordinates of imidazole, thioethanol, and the formate anion on the crystallographic coordinates of H57, C195, and D102, respectively. The optimum hydrogen bond between H57 and C195 was then determined by optimizing the torsion angles corresponding to Cys-195 S_γ and the acidic thiol proton; the imidazole ring was also free to rotate around the Cγ-H vector (which corresponds to the Cγ-Cβ

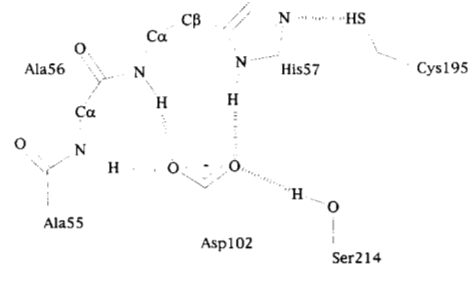

Fig. 5. Schematic representation of the molecules that were used to construct the largest model for the active site of S195C rat trypsin (model 4, Table 2).

bond in the crystal structure). The geometry of the zwitterionic configuration was constructed in a similar manner by using ImH^+ and $C_2H_5S^-$ to represent the side chains of H57 and C195. Inspection of the optimized geometries for model 1 indicates that the orientation of H57 changes very slightly when compared with the crystal structure. For the neutral configuration, the orientation of the imidazole ring is almost identical to the crystal structure, whereas a rotation of about 5 degrees occurred for the zwitterionic form.

Inspection of the crystal structure shows that the anionic side chain of $D102^-$ forms four hydrogen bonds with neighboring residues: the imidazole ring of His-57, the side chain of S214, and the peptide backbones of H57 and A56 (Fig. 5). These hydrogen bonds influence the charge distribution on the carboxyl group of D102, which in turn affects the pK_a of the imidazole ring of His-57. Therefore, the peptide backbones of H57 and A56, and the hydroxyl side chain of S214 should have a marked influence on the relative stability of the zwitterionic configuration. The effect of each of these residues has been investigated by systematically expanding model 1.

To assess the effect of the hydrogen bond formed between the backbone of His-57 and Asp-102^- on the stability of the zwitterionic configuration, model 1 was extended by using HCONH-$C_\alpha H_2$-$C_\beta H_2$-Im(H^+) to model His-57. The gas phase geometries of His-57 and His-57^+ (HCONH-$C_\alpha H_2$-$C_\beta H_2$-Im and HCONH-$C_\alpha H_2$-$C_\beta H_2$-ImH^+) were first optimized by constraining the torsion angles of the heavy atoms to correspond to the crystal structure; the remaining torsion angles and all bond lengths and valence angles were optimized. The geometries for model 2 were then obtained by replacing the imidazole ring in model 1 with the optimized histidine residues.

Model 3 was constructed by adding the side chain of Ser-214 (using MeOH) to the optimized geometries of model 2. The orientation of the S214 hydroxyl proton, which forms a hydrogen bond to $D102^-$, was determined by overlaying a formate anion and methanol on the crystallographic coordinates of D102 and S214; the torsion angle corresponding to the hydroxyl proton was then optimized with a 6-31G* basis set. The geometries for model 3 were then obtained by combining the optimized $HCOO^-$/MeOH moiety with the geometries obtained previously with model 2.

To assess the effect of A56 on the relative stability of the zwitterion, HCONH-$C_\alpha H_2$-CONH-$C_\alpha H_2$-$C_\beta H_2$-Im and HCONH-$C_\alpha H_2$-CONH-$C_\alpha H_2$-$C_\beta H_2$-ImH^+ were used to model His-57 and the backbone of Ala-56. Proceeding as before, the torsion angles of the heavy atoms were fixed to coincide with the crystal structure; only the bond lengths, valence angles, and the remaining torsion angles were optimized. Model 4 was constructed by replacing the histidine residues in model 3 with the optimized HCONH-C_aH_2-CONH-C_aH_2-C_bH_2-Im(H^+) moieties. Finally, model 5 was constructed by using point charges [taken from the AMBER forcefield (Weiner at al., 1984)] to represent all residues within 12 Å of the active site. The SCF energies for models 1–5 are shown in Table 2. MP2 calculations were also performed on models 1, 2, 3, and 5. The core electrons and the corresponding anti-bonding core orbitals were omitted from all MP2 calculations, and the results are shown in Table 2.

Calculated proton affinities for His-57

Inspection of the mechanism shown in Figure 1 indicates that a proton transfer between the substrate and His-57 occurs in two of the steps involved in catalysis. Therefore, intuitively, one would expect the pK_a of His-57 to have an important effect on the barrier associated with each of these proton transfers. Furthermore, as described above, $D102^-$ forms four hydrogen bonds with residues H57, S214, and A56 (Fig. 5). These interactions should have a significant effect on the pK_a of H57, and I have investigated this by calculating the gas phase proton affinities for His-57,

$$\text{His-57}^+ ===== \text{His-57} + H^+, \quad (1)$$

as a function of the surrounding residues. The two steps involving proton transfers (Fig. 2B,D) occur after the side chain of the triad cysteine has formed a covalent bond with the substrate. Therefore, models for the active sites were constructed by tak-

Table 2. *Calculated SCF and MP2 energies (a.u.) for the active site of S195C rat trypsin*[a]

Model		SCF/6-31G*	MP2/6-31G*
1	H57[b],C195,$D102^-$	−889.77182 (N)	−891.29149
		−889.75586 (10.02)	−891.27373 (11.15)
2	H57/A56[c],C195[d],	−1,135.63667	−1,137.86345
	$D102^-$	−1,135.61680 (12.47)	−1,137.84227 (13.29)
3	H57/A56[c],C195,	−1,250.68930	−1,253.22060
	$D102^-$,S214	−1,250.66770 (13.97)	−1,253.19626 (15.28)
4	H57/A56/A55[e],	−1,457.50564 (14.22)	
	$D102^-$,S214	−1,457.48298	
5	H57/A56/A55[e],	−1,473.82307 (11.43)	−1,476.92846 (13.46)
	$D102^-$,S214 & point charges representing residues 12 Å from His-57	−1,473.80486	−1,476.90701

[a] Energies (kcal mol^{-1}) relative to the most stable configuration are in shown parentheses. The top number in each case refers to the neutral configuration.
[b] H57 = Im(H^+).
[c] H57/A56 = HCONH-CH_2-CH_2-Im(H^+).
[d] C195 = C_2H_5SH, $D102^-$ = $HCOO^-$, S214 = CH_3OH.
[e] H57/A56/A55 = HCONH-CH_2-HCONH-CH_2-CH_2-Im(H^+).

ing the partially optimized geometries constructed previously and removing the side chain of C195 in each case. Thus, the simplest model consists of the side chains of the H57/D102 dyad, whereas the largest model consists of H57, the peptide backbone of A56, and the side chains of $D102^-$ and S214. The results are shown in Table 3.

Results

Papain

By making a systematic comparison of the results in Table 1 (models 1–7), one is able to assess the relative effect of each residue on the stability of the zwitterionic and neutral configurations. However, it should be stated that if the active site models had been constructed by adding each residue in a different order, one might possibly expect the net (calculated) effect of each residue to change by as much as 1–2 kcal mol^{-1}. Nevertheless, such a comparison still provides a good semi-quantitative picture of which residues play a major role in stabilizing the zwitterionic configuration and, therefore, enables one to identify those residues that, although not directly involved in the catalytic mechanism, play an important indirect role in catalysis.

Model 1 favors the neutral configuration by 28.8 kcal mol^{-1}. A comparison with model 2 demonstrates that the peptide backbones of Cys-25 and Trp-26 preferentially stabilize the zwitterion by 10.3 kcal mol^{-1}. Examination of the optimized geometries of model 2 indicates that the cysteine sulfur is in close proximity to the imido protons of Cys-25, Trp-26, and Ala-160. These three protons and the hydrogen bond formed between the cysteine sulfur and the imidazole ring clearly help to stabilize the zwitterionic form. Inspection of the optimized zwitterion shows that the Cys-S_γ^- is almost equidistant to the following four protons: Cys-25-NH$\cdots S_\gamma$ = 2.71 Å, Trp-26-NH$\cdots S_\gamma$ = 2.87 Å, Ala-160-NH$\cdots S_\gamma$ = 2.68 Å, His-159-Hd1$\cdots S_\gamma$ = 2.34 Å. These four protons generate a large electropositive field that preferentially stabilizes the zwitterion.

The inclusion of the side chain of Asn-175 (model 3) has a relatively small effect: the zwitterion is preferentially stabilized by 0.9 kcal mol^{-1}. Asp-158, however, has a dramatic effect: the electronegative field generated by the ionized side chain of Asp-158 helps to stabilize the zwitterion by almost 11 kcal mol^{-1}. Gly-19 (model 5) and Ser-176 (model 6) have a relatively small effect and stabilize the zwitterion by 2.6 and 0.5 kcal mol^{-1}, respectively. The effect of helix-L1 (model 7) is also relatively small, stabilizing the zwitterion by 2.5 kcal mol^{-1}. Although the individual effects of Gly-19, Ser-176, and the α-helix are small, the combined effect is significant. Comparing model 7 with model 4 (6-31g* basis), it can be seen that the combined effect preferentially stabilizes the zwitterion by more than 5 kcal mol^{-1}. Larger SCF calculations on model 7 using the 6-31g** basis set are in good agreement with the 6-31g* basis set. The difference between the zwitterion and the neutral form is 0.7 kcal mol^{-1}, in favor of the neutral configuration. In model 7A, I investigated the effect of protonating D158. The formate anion in model 7 was replaced with formic acid. The 6-31g* SCF calculations predict that the neutral configuration is now more stable by 10.4 kcal mol^{-1}.

In model 8, I investigated the effect of "mutating" Asn-175 to aspartate. The formamide molecule representing the side chain of Asn-175 in model 7 was replaced by a formate anion. This mutation is predicted to have a dramatic effect on the stability of the zwitterion. The zwitterion is now 14.9 kcal mol^{-1} more stable than the neutral form. This result was initially a little surprising because, in a similar theoretical study on DLH, the reverse was found to be true. The catalytic triad in DLH is Cys$\cdots$His$\cdots Asp^-$; the presence of an ionized aspartate was predicted to make the formation of a thiolate anion more difficult. In DLH, this can only be achieved by substrate-induced activation of the enzyme (Cheah et al., 1993; Beveridge & Ollis, 1995).

Table 3. *Calculated proton affinities for His-57*

		SCF/6-31G*	MP2/6-31G*
1	H57[a],D102$^-$[b]	−413.03534 (321.24)	−414.19429 (316.31)
		−413.54727	−414.69836
2	H57/A56[c],D102$^-$	−658.90104 (316.18)	−660.76582 (311.18)
		−659.40544	−661.26172
3	H57/A56[c],D102$^-$,	−773.95431 (313.12)	−776.12363 (307.70)
	S214	−774.45329	−776.61397
4	H57/A56/A55[d],	−980.770669 (313.62)	−983.517164 (307.93)
	S214,D102$^-$	−981.270453	−984.007874

* Absolute energies are in a.u.; protonation energies are in parentheses (kcal mol^{-1}). In each case, the bottom number refers to the protonated species ($H57^+$).

[a] H57 = Im(H^+).

[b] $D102^-$ = $HCOO^-$, S214 = CH_3OH.

[c] H57/A56 = HCONH-CH_2-CH_2-Im(H^+).

[d] H57/A56/A55 = HCONH-CH_2-HCONH-CH_2-CH_2-Im(H^+).

S195C rat trypsin

Native S195C rat trypsin

Inspection of the results shown in Table 2 indicates that the S195C mutant exists exclusively in the neutral configuration. The largest calculation (model 5, MP2) predicts that the neutral form is favored by 13.5 kcal mol^{-1}, indicating that it is extremely unlikely that the zwitterionic configuration is present in appreciable amounts in the native enzyme. A comparison of the results for models 1–4 also indicates that the hydrogen bonds formed between the anionic side chain of $D102^-$ and the surrounding residues (S214, A56, and the peptide backbone of H57) favor the neutral configuration by several kcal mol^{-1}. The net effect of these interactions appears to stabilize the negative charge on $D102^-$; this weakens the electrostatic interaction between the imidazole ring of H57 and the carboxylate of $D102^-$, and this in turn destabilizes the zwitterionic configuration. The electrostatic effect of the surrounding residues within 12 Å of the active site has been estimated by modeling each of these residues using atom-centered point charges (model 5, Table 2). The electrostatic field generated by these residues helps to stabilize the zwitterionic configuration by almost 3 kcal mol^{-1} (SCF results, Table 2), but the effect is clearly much too small to have a significant effect on the overall stability of the zwitterionic configuration.

Calculated proton affinities for His-57

The results for the calculated proton affinities of H57 as a function of the neighboring residues are shown in Table 3. The

simplest model uses Im (ImH^+) and $HCOO^-$ to represent the side chains of H57 and $D102^-$. The calculated proton affinities are 321.4 kcal mol^{-1} (SCF) and 316.3 kcal mol^{-1} (MP2). As shown in Figure 5, the peptide backbone of H57 also forms a hydrogen bond with the anionic side chain of $D102^-$. This hydrogen bond polarizes the negative charge on $D102^-$ and thus weakens the electrostatic interaction between $D102^-$ and $H57^+$. This reduces the electrostatic stabilization of the cationic imidazole ring and this is reflected in a reduction of the calculated proton affinities, 316.2 kcal mol^{-1} (SCF) and 311.2 kcal mol^{-1} (MP2) (Table 3). By a similar argument, the hydrogen bond formed between S214 and $D102^-$ also lowers the calculated proton affinities [313.1 kcal mol^{-1} (SCF) and 307.7 kcal mol^{-1} (MP2)]. However, the third hydrogen bond formed between the peptide backbone of A56 and $D102^-$ is predicted to have an almost negligible effect. Nevertheless, the combined effect of these three hydrogen bonds decreases the calculated proton affinities by about 8 kcal mol^{-1}, and this should manifest itself as a significant decrease in the pK_a of the catalytic histidine. Thus, the local environment surrounding the anionic side chain of $D102^-$ could be considered to be an "electropositive hole" that polarizes the charge distribution on the carboxylate of $D102^-$, and, in so doing, modulates the pK_a of His-57. Such an effect may well have important implications for catalysis.

Discussion

Papain

Resting state of native papain

The best SCF calculations using the 6-31G* and 6-31G** basis sets in model 7 predict the neutral form to be slightly more stable than the zwitterion by 1.1 and 0.7 kcal mol^{-1}, respectively. However, I have also estimated the difference in correlation energy by performing MP2 calculations using model 2 (Table 1). The core electrons and the corresponding antibonding orbitals were omitted from the calculations. Comparing the SCF and MP2 results for model 4 (Table 1), there is a difference of 1.9 kcal mol^{-1} in the correlation energies in favor of the zwitterion. If the 6-31G* and 6-31G** SCF results obtained with model 7 are correct by this amount, the zwitterion is predicted to be the most stable by 0.8 and 1.2 kcal mol^{-1}, respectively. This suggests that the zwitterionic and neutral configurations are in equilibrium. The equilibrium is predominately shifted toward the zwitterion by almost a factor of 10. This is compatible with experiments [titration data (Lewis et al., 1976), deuterium isotope effects (Creighton et al., 1980), and methyl thiosulfonate reaction rates (Roberts et al., 1986)] that indicate that the active site is in equilibrium between the zwitterionic and neutral forms. The preferential stability of the zwitterion can be attributed largely to local environment effects. In particular, the peptide backbone of Cys-25 and Trp-26, and the ionized carboxylate of Asp-158, have the greatest effect in stabilizing the zwitterion.

The results of model 7A are very interesting. Protonating D158 (by replacing the formate anion representing D158 in model 7 by formic acid) results in the neutral form becoming the most stable configuration by 10.4 kcal mol^{-1}. A kinetic study of k_{cat} versus pH for the cysteine proteinases cathepsin B and H indicates that formation of a $Cys^- \cdots His^+$ ion pair is initiated by the protonic dissociation of a group with a $pK_a < 4$ (Willenbrock & Brocklehurst, 1985). The theoretical results for models 7 and 7A suggest that D158 may possibly perform a similar function in papain. The electronegative field generated by the carboxylate side chain of Asp-158 appears to be essential to ensure that the active site exists as a zwitterion prior to nucleophilic attack.

N175D mutant

One of the main objectives of the present study was to examine the effects of mutating Asn-175 to an aspartate. This produces the same catalytic triad that is found in DLH and the mutant serine proteases ($Cys \cdots His \cdots Asp$). An earlier study on DLH predicted that the presence of an ionized aspartate would inhibit the formation of a thiolate anion. However, in the case of the "mutant" papain, the reverse appears to be true. This suggests that the N175D papain mutant could possibly be catalytically active. There are two possible reasons why papain has evolved such that the third triad residue is Asn and not Asp. Mutating Asn to Asp may well interfere with protein folding, making it impossible to form a catalytically competent enzyme. Alternatively, the mutation might have adverse effects on the catalytic mechanism. Drenth et al. (1975) have postulated that the first step in catalysis by papain involves nucleophilic attack on the substrate carbonyl by Cys-25 S_γ^-, resulting in the formation of a covalent enzyme/substrate complex. The second step involves the protonated imidazole ring transferring a proton to the scissile nitrogen, resulting in cleavage of the scissile bond and the formation of an acyl-enzyme intermediate. The amine portion of the substrate then diffuses away from the active site and is replaced by a water molecule that reacts with the acyl-enzyme intermediate to form a second tetrahedral complex before breaking the C-S bond and restoring the enzyme in the active state (Fig. 2). The theoretical calculations suggest that mutating Asn-175 to an aspartate would be expected to have a significant effect on the acidity of the protonated imidazole ring. This should inhibit the formation of an acyl-enzyme intermediate by making it more difficult to protonate the initial covalent enzyme/substrate complex. Furthermore, kinetic studies by Willenbrock and Brocklehurst (1985) indicate that acid catalysis (from ImH^+; pK_a 8.3) is essential to complete the acylation step. Therefore, increasing the pK_a of the catalytic histidine should decrease the catalytic efficiency. Further theoretical studies on the catalytic mechanism of papain and the N175D mutant are currently in progress and support the hypothesis that Asn-175 plays an important role in catalysis by modulating the pK_a of His-159. Preliminary semi-empirical calculations using the AM1 and PM3 Hamiltonians have indicated that mutating Asn-175 to Asp increases the pK_a of His-159, making the formation of the initial covalent enzyme/substrate complex much less favorable energetically (A.J. Beveridge, unpubl. results). Recent kinetic studies on the papain mutant N175A (Vernet et al., 1995), have shown that mutating Asn-175 to alanine results in a significant decrease in the intrinsic activity of the enzyme. The authors argue that the loss of a hydrogen bond to His-159 destabilizes the zwitterionic form of the active site ($Cys\text{-}25^- \cdots His\text{-}159^+$), thus making the formation of a reactive thiolate anion prior to catalysis much more difficult. The present calculations support this hypothesis and suggest that Asn-175 stabilizes the zwitterionic form by about 1 kcal mol^{-1} (Table 1); this would effectively increase the relative concentration of the $Cys\text{-}25^-$ by a factor of

10. Vernet et al. (1995) also studied the N175Q mutant; mutating Asn-175 to glutamine was found to have a relatively small effect on the intrinsic activity of the enzyme, demonstrating that the presence of a neutral residue with a side chain that can form a hydrogen bond to His-159 is essential for catalysis. Overall, the calculations demonstrate that several residues play an important role in stabilizing the zwitterionic form of the active site. Mutating any one of these residues may have an undesirable effect on the pK_as of Cys-25 and His-159, and may result in a significant loss of activity.

S195C rat trypsin

Inspection of the active site of S195C rat trypsin (Fig. 5) reveals that residues A56, H57, and S214 form four hydrogen bonds with $D102^-$. Collectively, these three residues form an electropositive hole, which stabilize the anionic side chain of D102. Similarly, the nucleophilic thiolate anion in papain is stabilized by hydrogen bonds from W26, H159, A160, and the peptide backbone of C25 (Fig. 4). Thus, a similar electropositive hole also exists in papain. Moreover, the ab initio calculations on papain demonstrate that these residues play a fundamental role in preferentially stabilizing the reactive thioalate anion. If one compares the two active sites (Figs. 4, 5), it is obvious that the spatial juxtapositions of the nucleophile (cysteine) and the third triad residue have been interchanged and therefore the active sites of the cysteine and serine proteinases have, in effect, been inverted.

On the basis of this observation, one can now explain the inactivity of S195C rat trypsin. In S195C rat trypsin, there are no hydrogen bonding residues in close proximity to the nucleophilic thiol group of Cys-195, which can facilitate the formation of a reactive thiolate anion. The calculations clearly demonstrate that the absence of such hydrogen bonds makes the formation of a reactive thiolate anion in S195C rat trypsin extremely unlikely (Table 2). One could, in principle, attempt to convert rat trypsin to a cysteine-like proteinase with the double mutation Asp-102 → Cys and Ser-195 → Asn, because this would lead to a catalytic triad with the same spatial juxtaposition as the cysteine proteinases. However, although this mutation is more likely to favor the formation of a reactive thiolate anion, the resulting enzyme is almost certain to be inactive because the substrate binding cleft will now be positioned incorrectly to facilitate nucleophilic attack.

This result has important implications for the evolution of enzymes that contain a catalytic triad. The overall topologies of the serine and cysteine proteinases are very different, indicating that each evolved separately as a result of convergent evolution. Furthermore, the active sites of the serine and cysteine proteinases can essentially be regarded as mirror images, and therefore one can argue that evolution has found two alternative solutions for hydrolytic enzymes based on a catalytic triad. One of the main structural differences between the serine and cysteine proteinases can be rationalized by considering the chemistry of the nucleophilic groups (hydroxyl versus thiol) involved in catalysis. Both sulfur and oxygen belong to group 6A in the periodic table, and therefore generally exhibit similar chemical properties; however, the thiol group is a much more effective as a nucleophile when ionized. Nature has exploited this property during the evolution of the cysteine proteinases by ensuring that the local environment of the catalytic cysteine favors the formation of a highly reactive thiolate anion; the imidazole ring of H159 and the peptide backbones of A160, C25, and W26 form an electropositive hole that facilitates the formation of a reactive thiolate anion (Fig. 4).

In contrast, the electropositive hole in the serine proteinases does not interact with the nucleophilic serine, but instead interacts with the third triad residue ($Asp\text{-}102^-$) (Fig. 5). The results in Table 3 demonstrate that the net effect of this interaction is to modulate the proton affinity of the catalytic histidine, which, in turn, should manifest itself as a marked shift in the pK_a of this group. Thus, in the case of the serine proteinases, the electropositive hole may well play an important role in catalysis by modulating the pK_a of the catalytic histidine. Furthermore, if one examines the catalytic mechanism of the serine proteinases (Fig. 1), it is apparent that such an effect would have important implications in catalysis because two of the key steps in the mechanism involve proton transfers between the substrate and the catalytic histidine. This may explain why the third triad residue in papain is Asn rather than Asp. The results for papain predict that active site exists in two forms ($Cys^- \cdots His^+ \cdots$ Asn and $Cys \cdots His \cdots Asn$), with the zwitterionic configuration being favored by about 1 kcal mol^{-1}. However, calculations on the N175D papain mutant demonstrated that, if Asn-175 is mutated to Asp^-, the relative stability of the zwitterion is further enhanced by about 10 kcal mol^{-1}. This result implies that the facile generation of a thiolate anion is not the only important factor in catalysis for the cysteine proteinases. At present, this possibility is being investigated further by calculating the PM3/AM1 energy profiles during the initial stages of catalysis for a model papain/substrate complex. Preliminary results suggest that mutating N175 to Asp^- dramatically increases the barrier associated with the first proton transfer from H159 to the bound substrate (Fig. 2B). This will be the subject a future paper, but the preliminary results of this study suggest that the third triad residue plays an important role in catalysis by modulating the pK_a of the catalytic histidine.

Comparisons with related structures

At present, the only known example of a naturally occurring enzyme that has the same triad (Cys-His-Asp) as S195C rat trypsin is DLH (Pathak & Ollis, 1990). However, crystallographic (Cheah et al., 1993) and theoretical (Beveridge & Ollis, 1995) studies have demonstrated that a fourth residue, E36, plays an essential role in catalysis; E36 abstracts a proton from the nucleophilic cysteine immediately prior to catalysis, thus generating a highly reactive thiolate anion (Fig. 3). Mutating E36 to alanine produces a catalytically inert enzyme, a result that demonstrates that DLH is essentially a catalytic tetrad and cannot therefore be compared directly with the serine and cysteine proteinases. Similarly, there are many examples of enzymes that use a cysteine-histidine dyad [e.g., the interleukin-1β-converting enzyme (Walker et al., 1994; Wilson et al., 1994)]. The calculations on papain have demonstrated that other residues (e.g., D158) proximal to the active site may have a significant influence on the relative stability of the nucleophilic thioalate anion and can therefore play an indirect role in catalysis. Further theoretical studies are needed to determine if catalytic dyads also exist as zwitterions, but this was beyond the scope of the present study.

Recent crystallographic studies of two picornaviral 3C cysteine proteinases have revealed that both of these enzymes have a similar fold to the chymotrypsin-like serine proteinases. The

3C proteinase of hepatitis A virus (HAV-3C) (Allaire et al., 1994) has an aspartate (D84) in close proximity to the catalytic histide (H44), but the side chain points away from the active site and cannot form a hydrogen bond with H44. Although D84 does not form an ion pair with the catalytic histidine, its function may well be similar to that of $D158^-$ in papain, where the theoretical results demonstrate that the electrostatic field due to the anionic side chain is sufficiently strong to stabilize zwitterionic form of the active site ($Cys^-\cdots His^+\cdots Asn$). Thus, this structure should generally be considered to be a catalytic dyad (C172···H44) and again, cannot therefore be compared directly with the serine and cysteine proteinases. The human rhinovirus-14 3C proteinase (Matthews et al., 1994) contains the unusual catalytic triad Cys-146···His-40···Glu-71 and is therefore very similar to that of S195C rat trypsin. Presumably, this triad also exists as a zwitterion ($Cys^-\cdots His^+\cdots Glu^-$), although further calculations would be necessary to determine which structural features stabilize the zwitterionic form. The present theoretical study suggests that the facile generation of a reactive thiolate anion may be an important step in catalysis for many enzymes that contain a nucleophilic cysteine. However, further theoretical work is clearly necessary before one can compare the results that have been obtained for the serine and cysteine proteinases to other similar systems.

Limitations of the present theoretical models

In the present study, I have chosen to neglect the effects of solvation on the active site. Rullman et al. (1989) have estimated the effects of solvation by performing (classical) Monte Carlo simulations on native papain; the simulations were restricted to all residues within 5 Å of the active site cavity (37 in total) and the 14 active site water molecules found in the crystal structure (Kamphuis et al., 1984). Solvation was predicted to favor the the zwitterionic configuration in the native enzyme by 24–28 $kJmol^{-1}$. However, if one assumes that substrate binding displaces all (or most) of the active site water molecules, then these models, which do not include solvent, are more representative of what a substrate "sees" as it approaches the active site. Therefore, it is unlikely that the results would be altered significantly by including effects due to solvation.

Another potential source of error in the present theoretical calculations is the use of point charges to simulate the influence of residues proximal to the active site. In a quantitatively more rigorous approach, Rullman et al. (1989) performed Hartree-Fock-Direct Reaction Field (HF-DRF) calculations on the active site of papain. In this study, the active site residues, Cys-25, His-159, and Asn-175, were modeled using thiomethanol, imidazole, and formamide; the rest of the enzyme was modeled as a set of interacting charges and polarizabilities. Although the HF-DRF formalism provides a more realistic model for the electrostatic and dielectric properties of an enzyme, this study demonstrates that similar results (for papain) could be obtained by using the much simpler point charge representation (used in the present study) for the bulk of the enzyme. On the basis of these observations, I have assumed that modeling the remaining residues using a reaction field methodology would not alter significantly the results obtained with the simpler point charge model. However, in comparison with earlier theoretical calculations, I find that a significant source of error is associated with choosing which residues to include explicitly in the active site. Previous theoretical studies (van Duijnen et al., 1980; Dijkman et al., 1989; Rullman et al., 1989) have been restricted to the triad residues: Cys-25, His-159, and Asn-175. The present calculations demonstrate that several residues proximal to the catalytic triad, in particular, Asp-158 and the peptide backbones of Cys-25 and Trp-26, play an important role in stabilizing the zwitterionic configuration and cannot be modeled correctly by using point charges or a reaction field approach. The present study explicitly includes the most important residues in the active site and therefore the ab initio calculations described in this paper are the most extensive to be reported on the active site of papain to date.

Conclusions

The main conclusion from the present study is the observation that the spatial juxtaposition of the three key residues has been inverted in the serine and cysteine proteinases, and therefore it is not possible to interconvert these two enzymes by site-directed mutagenesis. DLH, an enzyme that contains the triad Cys-123···His-202···Asp-171 (Fig. 3), superficially appears to be an exception to this rule. However, site-directed mutagenesis studies have shown that a fourth residue ($E36^-$) plays a fundamental role in catalysis; mutating E36 to alanine results in a catalytically inert enzyme (Cheah et al., 1993). This result suggests that E36 plays an essential role in catalysis by abstracting the thiol proton to generate a highly reactive thiolate anion (Fig. 3). This hypothesis is further supported by theoretical calculations (Beveridge & Ollis, 1995) that demonstrate that substrate binding facilitates the formation of a reactive thiolate anion by forcing the carboxyl group of E36 to abstract the thiol proton of C123. Thus, DLH is not a genuine catalytic triad, but should instead be classified as a catalytic tetrad. Evolution has only succeeded in activating the hybrid Cys···His···Asp triad by using a fourth residue as a base.

Acknowledgments

I am grateful to Professor R.J. Fletterick for providing the S195C rat trypsin coordinates.

References

Allaire M, Chernaia MM, Malcolm BA, James MNG. 1994. Picornal 3C cysteine proteinases have a fold similar to chymotrypsin-like serine proteinases. *Nature 369*:72–76.

Beveridge AJ, Ollis DA. 1995. A theoretical study of the substrate induced activation of dienelactone hydrolase (DLH). *Protein Eng 8*:135–142.

Bolis G, Clementi E, Ragazzi M, Salvaderi D, Ferro DR. 1978. Preliminary attempt to follow the enthalpy of an enzymatic reaction by ab initio computations: Catalytic action of papain. *Int J Quant Chem 14*:815–838.

Cheah E, Austin C, Ashley GW, Ollis D. 1993. Substrate-induced activation of dienelactone hydrolase: An enzyme with a naturally occurring Cys-His-Asp triad. *Protein Eng 6*:575–583.

Creighton DJ, Gessouron MS, Heapes JM. 1980. Is the thiolate-imidazolium ion pair the catalytically important form of papain? *FEBS Lett 110*: 319–322.

Dijkman JP, Osman R, Weinstein H. 1989. A theoretical study of the effect of primary and secondary structure elements on the proton transfer in papain. *Int J Quant Chem 35*:241–252.

Drenth J, Kalk KH, Swen HM. 1976. Binding of chloromethyl ketone substrate analogues to crystalline papain. *Biochemistry 15*:3731–3738.

Drenth J, Swen HM, Hoogenstraaten W, Sluyterman L. 1975. Mechanism of papain action. *Proc K Ned Acad Wte Ser C 78*:104–110.

Fersht AR. 1971. Acyl-transfer reactions of amides and esters with alcohols

and thiols. A reference system for the serine and cysteine proteinases. Concerning the N protonation of amides and amide–imidate equilibria. *J Am Chem Soc 93*:3504–3515.

Frisch MJ, Head-Gordon M, Trucks GW, Foresman JB, Schlegal HB, Raghavachari K, Robb M, Binkley JS, Gonzalez C, Defrees DJ, Fox DJ, Whiteside RA, Seeher R, Melius CF, Baker J, Martin RL, Kahn LR, Stewart JJP, Topiol S, Pople JA. 1992. Gaussian92. Pittsburgh, Pennsylvania: Gaussian Inc.

Garavito RM, Rossman MG, Argos P, Eventoff W. 1977. Convergence of active site geometries. *Biochemistry 16*:5065–5071.

Halasz P, Polgar L. 1977. Negatively charged reactants as probes in the study of the essential mercaptide-imidazolium ion-pair of thiolenzymes. *Eur J Biochem 79*:491–494.

Higaki JN, Evnin LB, Craik CS. 1989. Introduction of a cysteine proteinase active site into trypsin. *Biochemistry 28*:9256–9263.

Johnson FA, Lewis SD, Shafer JA. 1981. Determination of a low pK for histidine-159 in the S-methylthio derivative of papain by proton nuclear magnetic resonance spectroscopy. *Biochemistry 20*:44–48.

Kamphuis IG, Kalk KH, Swarte MBA, Drenth J. 1984. Structure of papain at 1.65 Å resolution. *J Mol Biol 179*:233–256.

Lewis SD, Johnson FA, Shafer JA. 1976. Potentiometric determination of ionisations at the active site of papain. *Biochemistry 15*:5009–5017.

Lewis SD, Johnson FA, Shafer JA. 1981. Effect of cysteine-25 on the ionisation of histidine-159 in papain as determined by proton nuclear magnetic spectroscopy. Evidence for a His-159···Cys-25 ion pair and its possible role in catalysis. *Biochemistry 20*:48–51.

Matthews DA, Smith WW, Ferre RA, Condon B, Budahazi G, Sisson W, Villafranca JE, Janson CA, McElroy HE, Gribskov CL, Worland S. 1994. Structure of human rhinovirus 3C protease reveals a trypsin-like polypeptide fold, RNA-binding site, and means for cleaving precursor polyprotein. *Cell 77*:761–771.

Moller C, Plesset MS. 1934. Note on and approximation treatment for many-electron systems. *Physical Review 46*:618–622.

Neet KE, Koshland DE. 1966. The conversion of serine at the active site of subtilisin to cysteine: A "chemical mutation." *Proc Natl Acad Sci USA 56*:1606–1611.

Pathak D, Ollis DL. 1990. Refined structure of dienelactone hydrolase at 1.8 Å. *J Mol Biol 214*:497–525.

Polgar L, Bender ML. 1967. The reactivity of thiol-subtilisin, an enzyme containing a synthetic functional group. *Biochemistry 6*:610–620.

Roberts DD, Lewis SD, Ballow DP, Olson ST, Shafer JA. 1986. Reactivity of small thiolate anions and cysteine-25 in papain toward methyl methanethiosulfonate. *Biochemistry 25*:5595–5601.

Roos B, Siegbahn P. 1970a. Gaussian basis sets for the first and second row atoms. *Theoret Chem Acta 17*:209–215.

Roos B, Siegbahn P. 1970b. Polarisation functions for first and second row atoms in Gaussian type MO-SCF calculations. *Theoret Chem Acta 17*: 199–208.

Rullman JAC, Bellido MN, van Duijnen. 1989. The active site of papain. All atom study of interactions with protein matrix and solvent. *J Mol Biol 206*:101–118.

Schlegal HB. 1989. Optimisation of equilibrium geometries and transition structures. *J Comp Chem 3*:214–218.

Thole BT, van Duijnen PTh. 1980. On the quantum mechanical treatment of solvent effects. *Theoret Chim Acta 55*:307–319.

Vernet T, Tessier DC, Chatellier J, Plouffe C, Lee TS, Thomas DY, Storer AC, Menard RJ. 1995. Structural and functional roles of asparagine 175 in the cysteine protease papain. *J Biol Chem 270*:16645–16652.

Walker NPC, Talanian RV, Brady KD, Dang LC, Bump NJ, Ferenz CR, Franklin S, Ghayur T, Hackett MC, Hamill LD, Herzog L, Hugunin M, Houy W, Mankovich JA, McGuiness L, Orlewicz E, Paskind M, Pratt CA, Reis P, Summani A, Terranova M, Welch JP, Xiong L, Moller A, Tracey DE, Kamen R, Wong WW. 1994. Crystal structure of the cysteine protease interleukin-1b-converting enzyme: A $(p20/p10)_2$ homodimer. *Cell 78*:343–352.

Weiner SJ, Kollman PA, Case DA, Singh UC, Ghio C, Alagona G, Profeta S, Weiner P. 1984. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J Am Chem Soc 106*:765–784.

Wilke ME, Higaki JN, Craik CS, Fletterick RJ. 1991. Crystal structure of rat trypsin-s195C at −150 °C. Analysis of low activity of recombinant and semisynthetic thiol proteases. *J Mol Biol 219*:511–523.

Willenbrock F, Brocklehurst K. 1985. A general framework of cysteine-proteinase mechanism deduced from studies on enzymes with structurally different analogous catalytic-site residues Asp-158 and -161 (papain and actinidin), Gly-196 (cathepsin B) and Asn-165 (cathepsin H). *Biochem J 227*:521–529.

Wilson KP, Black JF, Thomson JA, Kim EE, Griffith JP, Navia MA, Murcko MA, Chambers SP, Aldape RA, Raybuck SA, Livingston DJ. 1994. Structure and mechanism of interleukin-1b converting enzyme. *Nature 370*:270–275.