Locating the Catalytic Water Molecule in Serine Proteases

The position of a long-sought catalytic water molecule in the serine protease mechanism was recently discussed by P. T. Singer *et al.* (*1*), who applied time-resolved Laue crystallography to view an ordered water (Wat^{1082}) bound in a functioning active site of trypsin. They proposed that this water molecule acts as the catalytic nucleophile in deacylation of the covalent acyl enzyme intermediate. In their experiment, Singer *et al.* reacted trypsin with the slowly hydrolyzed substrate *p*-guanidinobenzoate and obtained Laue data sets at pH 5.5 before hydrolysis and shortly after initiating the reaction by shifting the pH to 8.5. A novel water molecule appearing at the higher pH was interpreted as that which actually performs the nucleophilic attack leading to deacylation of trypsin.

We believe the water molecule Wat^{1082} identified by Singer *et al.* is unlikely to represent the catalytic nucleophile for a number of reasons. First, this water is located 3.9 Å from the reactive electrophilic carbon of the intermediate. This distance is too great to be reconciled with the proximities expected to be found between reactive enzyme and substrate moieties, particularly because the reaction was in progress at the time the data were measured. Second, the *p*-guanidinobenzoate substrate fills only the primary or S1 site of the enzyme (*2*). This is not the situation for peptide and protein substrates, which also fill the (S2, S3 . . . S_n) sites of the amino terminal to the cleaved bond. The filling of these sites by substrate forms a well-defined antiparallel β sheet with enzyme residues 214 to 216 (*3*). When bound to these residues, the extended polypeptide chain of the substrate is coincident with the binding site of the proposed catalytic Wat^{1082}. Displacement of the bound peptide to accommodate Wat^{1082} in the catalytic event is a remote possibility because the peptide is covalently bound. Water must attack the reactive carbon from a different orientation in the case of the larger substrates. Therefore, acceptance of Wat^{1082} as the catalytic nucleophile would imply that trypsin operates by a fundamentally different mechanism on single-residue (as opposed to peptide or protein) substrates.

Mechanistic considerations dictate a further rationale for suggesting that the catalytic water molecule approaches the electronegative acyl enzyme carbon from a different position. Nucleophilic attack by the water oxygen atom is accompanied by donation of one of its protons to Nε of His^{57}. This event directly follows completion of the acylation step of the mechanism, which includes deprotonation of His^{57} Nε and dissociation of the leaving group. Therefore, it is far more likely that the water nucleophile is located adjacent to the Nε of His^{57} on the leaving-group side of the scissile bond, rather than on the amino-terminal side, as proposed by Singer *et al.* Positioning the water on the leaving-group side allows it to approach directly the lone pair of electrons present on His^{57} Nε in the most favorable geometric orientation.

Alternative, more plausible possibilities exist for the way in which the water nucleophile attacks the acyl enzyme. In one scenario, a water from bulk solvent could approach the electrophilic carbon from the leaving-group side in an orientation favorable for proton donation to His^{57} Nε. This mechanism would indicate that there is no specific binding site for the catalytic water on the enzyme. In this case it would not be expected that the water could be visualized in any crystallographic experiment, whether performed by Laue diffraction or with conventional monochromatic radiation.

Inspection of crystal structures of serine proteases also reveals an alternative candidate for an ordered catalytic water nucleophile. This water appears in a conserved position in five crystal structures of peptide transition state analog inhibitor complexes with human neutrophil elastase, trypsin, and the bacterial α-lytic protease (*4*). It forms a hydrogen bond to the backbone carbonyl group at position Phe^{41} and is located 3.5 to 3.7 Å distant from the electrophilic carbon (Fig. 1). The binding site of this water is coincident with that of the P1′ leaving-group residue of the substrate. The position of the carbonyl group of residue 41 is conserved in all trypsin-like serine proteases for which high-resolution structures are available (*5*). In subtilisin, a serine protease that possesses a superimposable catalytic triad within a highly disparate tertiary structure, the backbone carbonyl group of Asn^{218} occupies a similar position (*6*). Locating the catalytic water on the leaving-group side of the scissile bond allows trypsin to cleave single-residue, peptide, and protein substrates by the same mechanism.

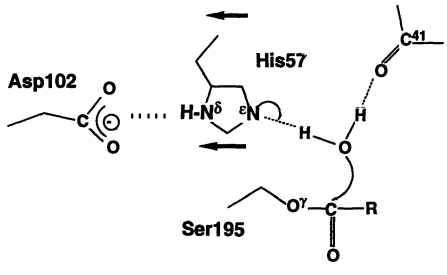

Fig. 1. Role of the catalytic water in deacylation. His^{57} is shown in a shifted position relative to the ground state, which could be adopted during donation of the proton from Nε to the leaving group (*7*). In this position the hydrogen bond with Asp^{102} may be weakened or broken. The catalytic water occupies a position just vacated by the P1′ residue of the leaving group. Large arrows indicate possible movement of His^{57} after acceptance of the proton from the water. After this movement, the hydrogen bond between Nε of His^{57} and Asp^{102} is reformed as the Nε proton is donated to Oγ of Ser^{195}.

Donation of a hydrogen bond by this water to the carbonyl oxygen at position 41 could serve to activate it for nucleophilic attack (Fig. 1). As seen in the inhibitor crystal structures, this water is displaced approximately 1 Å away from an optimal position to directly attack the trigonal carbon of the acyl enzyme. Possibly, the binding site could be viewed as a platform from which rapidly exchanging waters from bulk solvent can traverse the distance to the electrophilic carbon. Alternatively, His^{57} may shift in the direction of this water as it transfers its proton to the leaving group (*7*). In its new position His^{57} might more easily acquire a new proton from the catalytic water. Displacement of the water toward the acyl enzyme target might then be favored because His^{57} moves in the same direction as it retreats to its original position (Fig. 1).

John J. Perona
Charles S. Craik
Robert J. Fletterick
Department of Pharmaceutical Chemistry and Department of Biochemistry and Biophysics, University of California, San Francisco, CA 94143

REFERENCES AND NOTES

1. P. T. Singer, A. Smalås, R. P. Carty, W. F. Mangel, R. M. Sweet, *Science* **259**, 669 (1993).
2. The nomenclature for the substrate amino acid residues is (Pn . . . P2, P1, P1′, P2′ . . .Pn′), where P1 to P1′ denotes the hydrolyzed bond. (Sn . . . S2, S1, S1′, S2′ . . . Sn′) denote the corresponding enzyme binding sites [I. Schechter and A. Berger, *Biochem. Biophys. Res. Commun.* **27**, 158 (1968)].
3. A. Ruhlmann, D. Kukla, P. Schwager, K. Bartels, R. Huber, *J. Mol. Biol.* **77**, 417 (1973); J. J. Perona, C. A. Tsu, C. S. Craik, R. J. Fletterick, *ibid.* **230**, 919 (1993).
4. This water is visualized, in equivalent position, in all available crystal structures of serine proteases complexed with peptidyl boronic acid and peptidyl chloromethyl ketone transition state analogs. In these structures, which provide models of the transition state for deacylation of protein substrates, the reactive carbon or boron has a tetrahedral geometry and makes a covalent bond to the Ser^{195} hydroxyl [R. Bone, S. B. Shenvi, C. A.

Kettner, D. A. Agard, *Biochemistry* **26**, 7609 (1987); M. A. Navia *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 7 (1989); J. J. Perona, L. Hedstrom, W. J. Rutter, R. J. Fletterick, in preparation; R. M. Stroud and J. Finer-Moore, personal communication].

5. The position of the carbonyl group of residue 41 (chymotrypsin numbering system) is conserved in all known structures of trypsin-like serine proteases [Brookhaven Protein Data Bank (PDB) identification codes given]: chymotrypsin (4CHA), pancreatice elastase (3EST), rat mast cell protease II (3RP2), neutrophil elastase (1HNE), tonin (1TON), porcine pancreatic kallikrein A (2PKA), trypsin (3PTN), and the bacterial enzymes α-lytic protease (1P01), *Staphlococcus griseus* trypsin (1SGT), and *S. griseus* proteinase B (3SGB). The carbonyl oxygen atom aligns within 1.5 Å for all of these and within 1.0 Å for the seven mammalian enzymes. The disulfide bond linking residues Cys^{42} to Cys^{58} may be strictly conserved in part because it is required for maintaining the orientation of the carbonyl oxygen at position 41 relative to the catalytic His^{57}.
6. The high-resolution crystal structures of subtilisin BPN' [C. S. Wright, R. A. Alden, J. Kraut, *Nature* **221**, 235 (1969); PDB 1SBT] and of subtilisin Carlsberg complexes with eglin C [W. Bode, E. Papamoukos, D. Musil, U. Seemueller, H. Fritz, *EMBO J.* **5**, 813 (1986); PDB 1CSE] were superimposed onto that of the neutrophil elastase (HNE) with the use of side chain atoms of the catalytic triad serine, histidine, and aspartate residues. The carbonyl oxygen of residue Asn^{218} in the two subtilisin structures is displaced by 2 to 3 Å relative to the equivalent carbonyl oxygen of residue 41 of HNE. However, the orientation and distance of this group with respect to the catalytic triad residues are similar in each enzyme.
7. The issue of whether His^{57} shifts position during leaving-group protonation is still unresolved. Various studies conducted in solution have reached conflicting conclusions [W. W. Bachovchin, *Biochemistry* **25**, 7751 (1986); J. Kraut, *Annu. Rev. Biochem.* **46**, 331 (1977); J. H. Wang, *Proc. Natl. Acad. Sci. U.S.A.* **66**, 874 (1970)]. Crystal structures of α-lytic protease complexes with peptide boronic acid and phosphonate transition state analogs show that the leaving-group oxygen atom (equivalent to the P1'-nitrogen of a protein substrate) is hydrogen-bonded to $N\epsilon$ of His^{57} [R. Bone, A. B. Shenvi, C. A. Kettner, D. A. Agard, *Biochemistry* **26**, 7609 (1987); R. Bone, N. S. Sampson, P. A. Bartlett, D. A. Agard, *ibid.* **30**, 2263 (1991)]. This suggests that little movement of His^{57} may be required for leaving-group protonation.
8. We thank D. Agard for helpful discussions. This work was supported by NSF grant DMB-921986 (to C.S.C.), NIH grant DK-39304 (to R.J.F.), and NIH postdoctoral National Research Service award GM13818-02 (to J.J.P.).

17 February 1993; accepted 6 May 1993

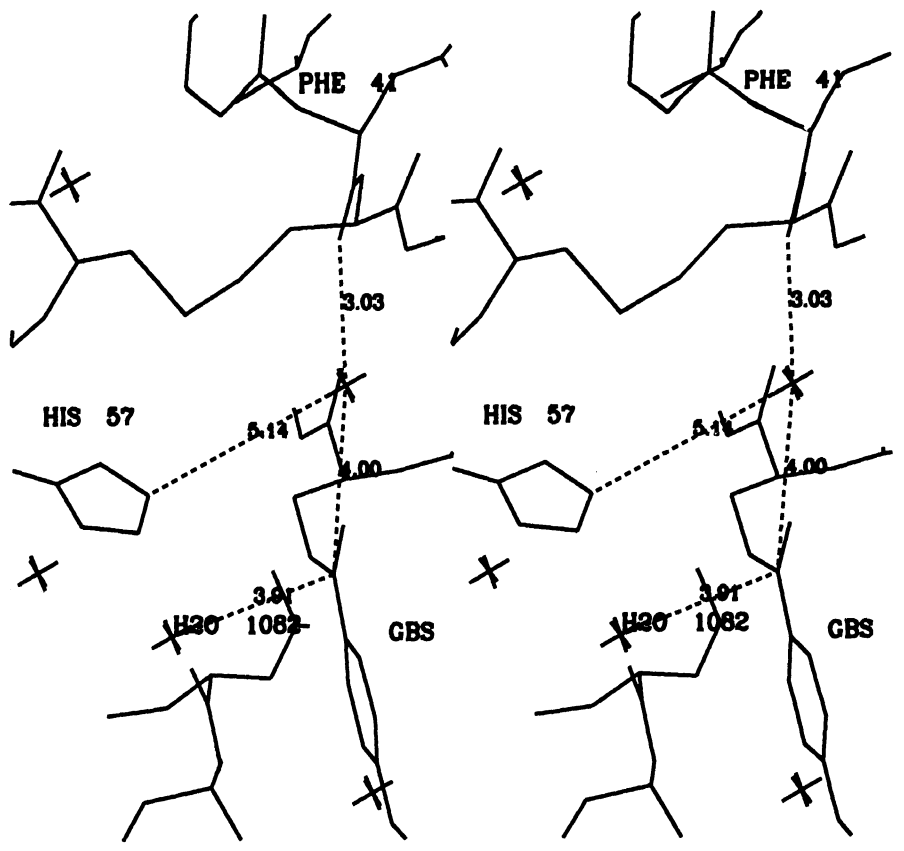

Fig. 1. Crossed-eye stereogram of the active site of GB-trypsin 3 min after the pH has been raised to begin hydrolysis. The view is similar to that of figure 3 in (*2*), except that the water molecule hydrogen-bonded to Phe^{41} O is shown, along with distances to nearby atoms.

Response: The inherent nature of x-ray diffraction, that it takes time to make the measurements, has hampered study of enzyme catalysis by crystallographic methods. We have combated this problem in recent dynamic studies of catalysis by trypsin (*1*, *2*) by use of the Laue method. We measured the data to determine complete three-dimensional structures in a minute or so. In these studies we expanded on our earlier work on a complex between trypsin and the slowly hydrolyzed substrate *p*-guanidinobenzoate (GB) (*3*). We determined structures before and immediately after the enzyme had been activated by a rapid increase in pH to catalyze hydrolysis of this species. Under the second set of conditions, the half-life for the species was roughly 2 hours; we still saw a static but only transiently stable structure. Its dramatic feature was that a water molecule [Wat^{1082} in (*2*)] appeared in a site where it was poised to perform the hydrolysis reaction (Fig. 1). The relatively long distance from Wat^{1082} to the carboxyl carbon atom (3.9 Å) reflects the relatively long lifetime of this species.

The approach of the hydrolytic water molecule should follow standard organic chemical principles and generally accepted mechanisms for serine proteases. What we mean in the first case is that the nucleophile should approach the planar carboxylate group from above or below the plane, attacking the carbon atom directly. In the second case, we mean that the arrangement of atoms, and actually the transfer of momentum to the carboxyl group, should be such that the carbonyl oxygen atom is forced naturally to form hydrogen bonds with the peptide nitrogen atoms of residues 193 and 195 in the "oxyanion hole" (*4*). Wat^{1082} is in approximately the right location to make this approach. Its presence in this transient structure does not so much define the resting site for a water molecule that will always perform ester hydrolysis as it suggests an approach route for the hydrolytic water, flanked by species that will activate the oxygen atom for nucleophilic attack.

Another water molecule we have observed in these structures where the GB group is bound (*1–3*) lies perched on the carbonyl oxygen atom of GB and is hydrogen-bonded to the carbonyl oxygen atom of residue Phe^{41} (shown near the top of Fig. 1). Perona, Craik, and Fletterick kindly pointed out that, beyond being found in these structures of GB trypsin, this water molecule also appears in the five structures mentioned in their comment; we have also found it in two crystal structures of γ-chymotrypsin in which tripeptides are found in the active site (*5*, *6*). It is also likely that this is the water molecule found by Henderson in indole acryloyl chymotrypsin (*7*). The geometric relationship between this water molecule and the active site is revealed in Fig. 1, where distances are given for the time t = 3 min. Laue structure (*2*): 3.0 Å to O of Phe^{41}, 4.0 Å to Cδ of GB-Ser^{195}, and 5.1 Å to $N\epsilon2$ of His^{57}. The distances in other structures are similar, with the water molecule lying perhaps 0.3 Å closer to Cδ of GB-Ser^{195} or $N\epsilon2$ of His^{57} in some structures. That this solvent molecule appears *in stable structures* of several different proteases with a substrate or a product of hydrolysis present in the active site suggests that its role has to do with that stability, not that it can have a further role

in catalysis. Indeed, the distance that this water molecule lies from His^{57} Nε2 in these structures is probably too great for it to play a catalytic role. Also, in the cases of acyl enzymes such as GB trypsin (Fig. 1) and indole acryloyl chymotrypsin, the molecule lies in the plane of the carboxylate that would be its target, making attack unlikely. Finally, approach to His^{57} Nε2 from that direction is obscured by the Cβ atom of Ser^{195}; the face of the His^{57} ring that is approached by Wat^{1082} is open in all of these structures.

Perona, Craik, and Fletterick point out that the site occupied by Wat^{1082} is obscured when a peptide is bound in the active site. The hydrogen bond that is formed from the P1 NH to the O atom of Ser^{214} is rarely a strong one in the complexes that have been observed. It can be as short as 2.85 Å, the ideal distance for a strong hydrogen bond, in neutrophil elastase (8), but is characteristically longer—as long as 3.6 Å in the complex between bovine trypsin and the bovine basic pancreatic trypsin inhibitor (BPTI) (9). One can suppose that, in an acyl enzyme complex formed with a natural substrate, the P1-P2 peptide would spend some time dislodged from this site and wave free in solution. This would provide opportunity for the equivalent of Wat^{1082} to approach in order to achieve hydrolysis. For a covalently bound substrate to leave its active site pocket after a catalytic event is not unusual. Nachman and his co-workers (*10*) labeled bovine pancreatic ribonuclease with an affinity label that required specific recognition by the enzyme for labeling. The nucleoside was not found to be in the active site pocket in the crystal structure. Instead only the linker atoms were present in electron density maps, and the nucleoside atoms were presumed to be disordered in the crystal.

Alternatively, one can consider the effect the presence of the P1-P2 peptide would have on a water molecule approaching in the region of Wat^{1082} and estimate how closely a water molecule could approach the O atom of Ser^{214}. In the trypsin/BPTI complex the carbonyl oxygen atom of inhibitor residue 36 approaches to within 6 Å of Ser^{214} O. This approach is likely close enough for there still to be significant coulombic interaction between the atoms. Evidence for this comes from a recent calculation of nonbonded energy as a function of cutoff distance in the crystalline structure of the cyclic peptide $[Ala\text{-}Pro\text{-}D\text{-}Phe]_2$; Kitson and Hagler showed that the total coulombic energy was seen to change by 63 to 97% of its asymptotic value as the cutoff distance was increased from 8 to 15 Å (*11*).

We reiterate our argument that the Nε atom of His^{57} can serve as a base catalyst as Wat^{1082} approaches the Cδ of GB-Ser^{195} by formation of a hydrogen bond somewhat out of the plane of the imidazole ring. A brief survey of the Brookhaven Protein Data Bank for geometry of hydrogen bond donors to the nitrogen atoms of histidine residues reveals a reasonable distribution of out-of-plane bonds (*12*).

Our study of a rapid transient state points out the polarization induced in the hydrolytic water molecule by the O atom of Ser^{214} and reveals an approach route across the solvent-accessible face of the imidazole ring of His^{57}. This route may be perturbed but is not eliminated by the presence of the P1-P2 peptide link.

P. T. Singer
Biology Department,
Argonne National Laboratory,
Argonne, IL 60439
A. Smalås
Institute of Mathematical and Physical Science,
University of Tromsø,
N-9000 Tromsø, Norway
R. P. Carty
Department of Biochemistry,
State University of New York
Health Science Center,
Brooklyn, NY 11203
W. F. Mangel
R. M. Sweet
Department of Biology,
Brookhaven National Laboratory,
Upton, NY 11973

REFERENCES AND NOTES

1. P. T. Singer *et al.*, *Philos. Trans. R. Soc. London Ser. A* **340**, 285 (1992).
2. P. T. Singer, A. Smalås, R. P. Carty, W. F. Mangel, R. M. Sweet, *Science* **259**, 669 (1993).
3. W. F. Mangel *et al.*, *Biochemistry* **29**, 8351 (1990).
4. J. D. Robertus *et al.*, *ibid.* **11**, 4293 (1972).
5. M. Harel *et al.*, *ibid.* **30**, 5217 (1991); Brookhaven Protein Data Bank (PDB) structure 8GCH.
6. M. M. Dixon, R. G. Brennan, B. W. Matthews, *Int. J. Biol. Macromol.* **13**, 89 (1991); PDB structure 3GCT.
7. R. J. Henderson, *J. Mol. Biol.* **54**, 341 (1970).
8. M. A. Navia *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 7 (1989); PDB structure 1HNE.
9. A. Ruehlmann *et al.*, *J. Mol. Biol.* **77**, 417 (1973); PDB structure 2PTC.
10. J. Nachman *et al.*, *Biochemistry* **29**, 928 (1990); PDB structure 9RSA.
11. D. H. Kitson and A. T. Hagler, *ibid.* **27**, 5246 (1988).
12. S. Bryant, personal communication.

19 March 1993; accepted 6 May 1993