Protein Science (1995), *4*:337–360. Cambridge University Press. Printed in the USA.

REVIEW

Structural basis of substrate specificity in the serine proteases

JOHN J. PERONA[1] AND CHARLES S. CRAIK

Departments of Pharmaceutical Chemistry and Biochemistry & Biophysics, University of California, San Francisco, California 94143-0446

(RECEIVED July 25, 1994; ACCEPTED December 28, 1994)

Abstract

Structure-based mutational analysis of serine protease specificity has produced a large database of information useful in addressing biological function and in establishing a basis for targeted design efforts. Critical issues examined include the function of water molecules in providing strength and specificity of binding, the extent to which binding subsites are interdependent, and the roles of polypeptide chain flexibility and distal structural elements in contributing to specificity profiles. The studies also provide a foundation for exploring why specificity modification can be either straightforward or complex, depending on the particular system.

Keywords: enzyme kinetics; macromolecular recognition; protein engineering; protein–ligand interactions; protein structure; serine protease; site-directed mutagenesis; substrate specificity

Serine proteases were among the first enzymes to be studied extensively (Neurath, 1985). Interest in this family has been maintained in part by an increasing recognition of their involvement in a host of physiological processes. In addition to the biological role played by digestive enzymes such as trypsin, serine proteases also function broadly as regulators through the proteolytic activation of precursor proteins (Neurath, 1984; Van de Ven et al., 1993). Examples of this regulation include the processing of trypsinogen by enteropeptidase to produce active trypsin (Huber & Bode, 1978) and the cascades of zymogen activation that control blood clotting (Davie et al., 1991). Serine proteases have also been recently shown to play essential roles in cell differentiation. For example, the *Drosophila* trypsin-like enzymes Easter and Snake are important components in the specification of ventral and lateral patterns during development (Chasan & Anderson, 1989). Asymmetry of cell fates may be the result of a protease cascade involving both of these enzymes (Smith & DeLotto, 1994).

An alternative rationale for the continued interest in serine proteases has been their emergence as one of the major paradigms for the understanding of enzymic rate enhancements and of structure–activity relationships. Until recently, all of the known enzymes fell into one of two distinct structural classes: the chymotrypsin-like and subtilisin-like families (Matthews, 1977; Fig. 1A,B). However, the crystal structure of wheat serine carboxypeptidase II (Liao & Remington, 1990; Liao et al., 1992; Fig. 1C) reveals conservation of the essential features of the catalytic apparatus within a third distinct protein fold. This homodimeric enzyme possesses the $\alpha+\beta$ fold found also in a number of other enzymes that share hydrolytic activity as their only common feature (Ollis et al., 1992). The fold consists of an 11-stranded mixed β-sheet structure surrounded by 15 helices, with the active site located at the base of a deep bowl-shaped depression in the enzyme surface (Fig. 1C).

The three serine protease classes are distinguished by the absence of any conserved secondary and tertiary motifs, but in

Reprint requests: Charles S. Craik, Departments of Pharmaceutical Chemistry and Biochemistry & Biophysics, University of California, San Francisco, California 94143-0446; e-mail; craik@cgl.ucsf.edu.

[1] Present address: Department of Chemistry and Interdepartmental Program in Biochemistry and Molecular Biology, University of California, Santa Barbara, California 93106.

Abbreviations: APPI, amyloid β-protein precursor inhibitor domain; BAP, *Bacillus alcalophilus* alkaline protease; BLAP, *Bacillus lentus* alkaline protease; BPTI, bovine pancreatic trypsin inhibitor; CMK, chloromethyl ketone; HNE, human neutrophil elastase; hGH, human growth hormone; Nva, norvaline, a linear three-carbon side chain; PAI-1, plasminogen activator inhibitor I; *p*NA, *para*-nitroanilide; PPE, porcine pancreatic elastase; PROK, *Thermus album* proteinase K; RMCPI and RMCPII, rat mast cell proteases I and II; SBPN, *Bacillus amyloliquefaciens* subtilisin BPN′; SCARL, *Bacillus licheniformis* subtilisin Carlsberg; SGPA, *Streptomyces griseus* protease A; SGPB, *S. griseus* protease B; SGPE; *S. griseus* protease E; SSI, *Streptomyces* subtilisin inhibitor; *suc*, succinyl; *suc*-FAHY-*p*NA, tetrapeptide amide substrates varying at the P1 position; *suc*-XAPF-*p*NA, tetrapeptide amide substrates varying at the P4 position; THERM, *Thermus vulgaris* thermitase; TPA, tissue plasminogen activator. Nomenclature for the substrate amino acid residues is P*n*, . . . , P2, P1, P1′, P2′, . . . , P*n*′, where P1–P1′ denotes the hydrolyzed bond. S*n*, . . . , S2, S1, S1′, S2′, . . . , S*n*′ denote the corresponding enzyme binding sites.

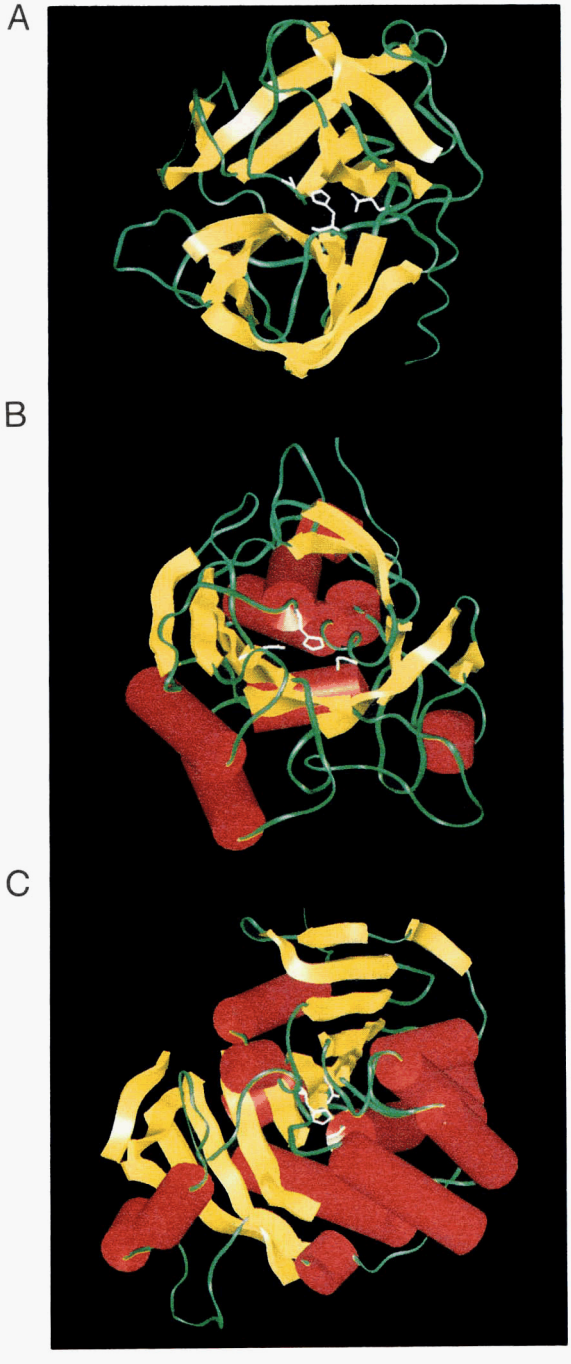


Fig. 1. Diversity of structural motifs in which the common catalytic apparatus of serine protease is embedded. Shown are ribbon drawings of chymotrypsin (**A**), subtilisin BPN′ (**B**), and wheat serine carboxypeptidase (**C**). α-Helices are shown as red cylinders and β-strands as yellow arrows. Secondary structures were determined by the algorithm of Kabsch and Sander (1983). Each enzyme possesses two common residues of crucial importance to catalysis: a nucleophilic Ser and an adjacent His, which functions as a general base (shown in white). Enzymes are oriented identically by superposition of the backbone atoms and Cβ of these two amino acids. A third member of the catalytic machinery is an aspartate residue (shown at left, also in white) not conserved in position relative to the Ser and His (compare serine carboxypeptidase with the other two enzymes). Lack of conservation in position of this residue suggests that the catalytic apparatus may be better viewed as a juxtaposition of Ser–His and His–Asp dyads, rather than as a single catalytic triad.

each case, the catalytic serine and histidine residues maintain an identical geometric orientation (Fig. 1). To a lesser extent, adjacent groups that stabilize the transition state are also similarly arranged (Wright et al., 1969; Robertus et al., 1972a, 1972b; Liao et al., 1992). Thus, it appears that nature has arrived at the same biochemical mechanism by separate avenues: the chymotrypsin, subtilisin, and serine carboxypeptidase families of serine proteases are a classic example of convergent enzyme evolution (Matthews, 1977; Liao et al., 1992). The resemblance of serine carboxypeptidase to other members of the α/β-hydrolase fold family also indicates the operation of divergent evolution within this structural framework (Ollis et al., 1992). Further, a recently generated catalytic antibody has been characterized that catalyzes the stereoselective hydrolysis of norleucine and methionine phenyl esters (Guo et al., 1994). The crystal structure of this enzyme reveals the presence of a Ser–His catalytic dyad structurally similar to those of the other serine protease classes (Zhou et al., 1994). A similar catalytic mechanism is therefore suggested, indicating that the antibody fold may well be a fourth structural framework capable of supporting proteolytic activity in a serine protease-like fashion.

We consider here the structural and kinetic basis for the diversity of substrate specificity in the subtilisin and chymotrypsin-class serine proteases. Emphasis is placed on those systems for which both crystallographic and detailed kinetic measurements are available. After a brief review of the common mechanism of the three classes and the role of mutational analysis in its further elucidation, we concentrate much of our attention on the three enzymes subtilisin BPN′, α-lytic protease, and trypsin. In each case, an extensive structure–function analysis has been applied to address the roles of particular amino acids in contributing to the observed specificity profiles. The wealth of information available on the chemical and kinetic mechanisms of catalysis and the large data base of homologous sequences provide an essential framework that supports these studies. Although the functional and/or structural properties of many of the mutant proteases can be given a relatively straightforward and objective description, there are also many examples where the data cannot be easily encapsulated. In these cases, some subjectivity in the description of kinetic and structural parameters is unavoidable, and other interpretations of the same data could yield different overall conclusions.

The catalytic mechanism

The vast majority of early studies on the serine proteases focused on the elucidation of the chemical and kinetic mechanisms of catalysis (reviewed by Bender & Killheffer, 1973; Blow, 1976; Kraut, 1977; Polgar, 1989). Hydrolysis of ester and amide bonds proceeds by an identical acyl transfer mechanism in all enzymes of the subtilisin and trypsin families (Fig. 2A,B,C). Michaelis complex formation is followed by attack on the carbonyl carbon atom of the scissile bond by the eponymous serine of the catalytic triad, which is enhanced in nucleophilicity by the presence of an adjacent histidine functioning as a general base catalyst. Proton donation by the histidine to the newly formed alcohol or amine group then results in dissociation of the first product and concomitant formation of a covalent acyl-enzyme complex. The deacylation reaction occurs via the same mechanistic steps, with the attacking nucleophile provided by a water molecule that approaches from the just-vacated leaving group

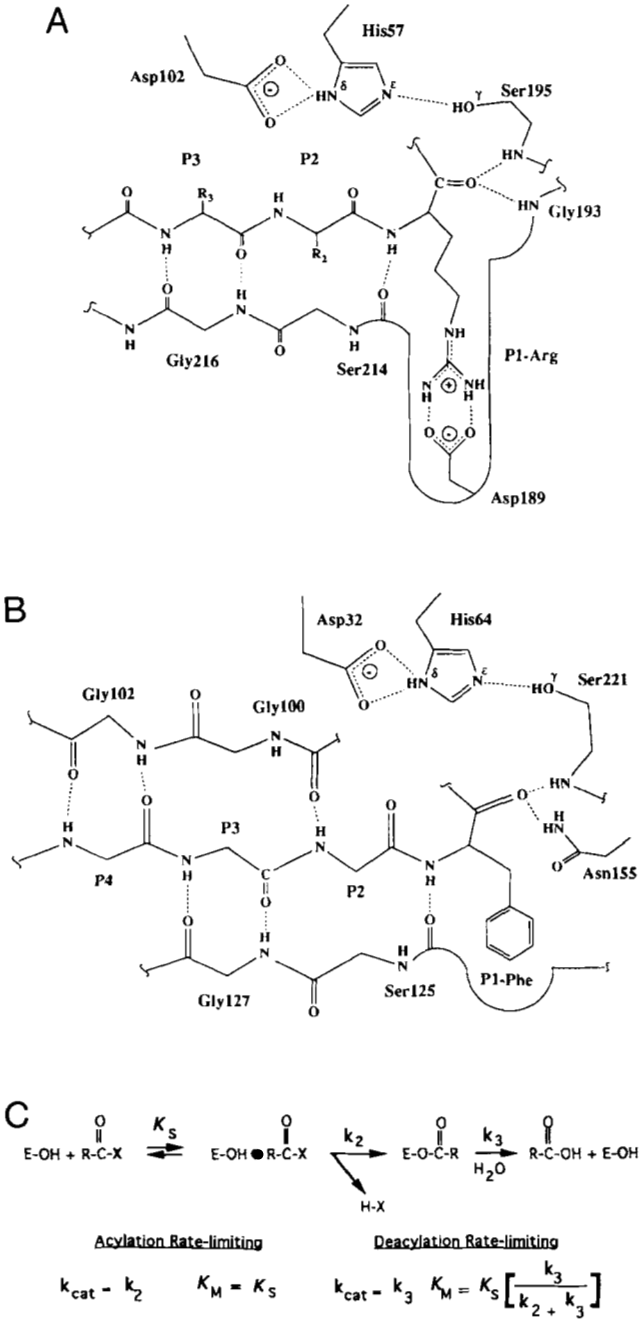


Fig. 2. Chemical and kinetic mechanisms of catalysis for serine proteases. The catalytic groups of trypsin (**A**) and subtilisin (**B**) are shown interacting with an oligopeptide substrate binding to the P1–P4 sites. (Nomenclature for the substrate amino acid residues is P*n*, . . . , P2, P1, P1′, P2′, . . . , P*n*′, where P1–P1′ denotes the hydrolyzed bond. S*n*, . . . , S2, S1, S1′, S2′, . . . , S*n*′ denote the corresponding enzyme binding sites [Schechter & Berger, 1968].) Note the distinction in residues that form the oxyanion hole; in subtilisin, part of the interaction is made by an enzyme side chain. The binding site for the oligopeptide also differs; in subtilisin it forms the central strand of a three-stranded antiparallel β-sheet. The S1 site of trypsin and the S1 and S4 sites of subtilisin are the major sites where mutagenesis has been used to probe specificity. **C**: Common kinetic mechanism of catalysis for serine proteases indicating the meaning of the mechanistic rate constants and their relationship to the Michaelis parameters. The correct interpretation of k_{cat} and K_m differs depending on the rate-limiting step in catalysis, which varies among the different enzymes as well as among differing substrates of the same enzyme.

side. Each step proceeds through a tetrahedral intermediate, which resembles in structure the high-energy transition state for both reactions. This mechanism is capable of accelerating the rate of peptide bond hydrolysis by a factor of more than 10^9 relative to the uncatalyzed reaction (Kahne & Still, 1988).

Extensive structural evidence obtained from X-ray crystallographic and NMR investigations has provided conclusive corroboration of the essential features of this mechanism (reviewed by Steitz & Shulman, 1982). The investigations have been favored by the availability of good ground-state and transition-state substrate analogs, which have been used to obtain high-resolution images of these interactions. The scissile bond of the substrate is bound directly adjacent to the Ser–His catalytic couple in all the complexes studied. A strong hydrogen bond between these two amino acids, necessary to subsequent proton transfer, is formed only after substrate is bound. A binding site for the oxyanion of the intermediate is formed by the Gly 193 and Ser 195 backbone amide nitrogens in the chymotrypsin-like enzymes (Fig. 2A), by one amide nitrogen and the Asn 155 side chain in the subtilisin family (Fig. 2B), and by the backbone amides of Tyr 147 and Gly 53 in the serine carboxypeptidases (Liao et al., 1992). The interactions made in the S1–S4 enzyme sites (see Fig. 2 legend for substrate nomenclature) by the P1–P4 positions of substrate form an antiparallel β-sheet hydrogen bonding arrangement in the chymotrypsin and subtilisin families. Because the active site of wheat serine carboxypeptidase II does not possess similarly exposed peptide backbone groups, it seems likely that substrate binding N-terminal to the scissile bond will occur in a different fashion in this family (Liao et al., 1992). Another unique structural feature of carboxypeptidase is an extensive hydrogen bonding network, which interacts with the C-terminal carboxylate of the substrate, essential to its activity as an exopeptidase (Mortenson et al., 1994).

Mutational analysis of both subtilisin and trypsin has confirmed the essential roles of Ser 195 and His 57 in providing rate acceleration. Replacement of the catalytic Ser 221 and His 64 residues of subtilisin with alanine results in decreases of 10^4–10^6-fold in k_{cat} (Carter & Wells, 1987, 1988). A decrease of 10^6-fold when the two residues are simultaneously replaced with alanine showed that the two catalytic moieties function in a highly cooperative manner: mutation of either component reduces activity to a baseline level. Similar results were obtained by analogous mutations of Ser 195 and His 57 in rat trypsin (Corey & Craik, 1992). This study also showed that enzyme variants such as H57K and H57E, which might provide an alternative general base, were ineffective, further underscoring the importance of the native catalytic triad geometry. These experiments, as well as others involving replacement of Ser 195 with a Cys (Higaki et al., 1989; McGrath et al., 1989) and engineering a metal-actuated activity switch involving His 57 (Higaki et al., 1990; McGrath et al., 1993), clarify the role of these active-site moieties. The mutational data are in agreement with early chemical modification experiments, which also indicated that Ser 195 and His 57 play crucial roles in catalysis (Dixon et al., 1956; Shaw et al., 1965).

The residual activity remaining in subtilisin after removal of the catalytic moieties was attributed to remaining binding determinants that stabilized the transition state complex. One such determinant is provided by a hydrogen bonding interaction of Asn 155 with the oxyanion intermediate. Mutation of Asn 155 to a variety of other amino acids resulted in 10^2–10^3-fold de-

creases in k_{cat}/K_m (Bryan et al., 1986; Wells et al., 1986; Carter & Wells, 1990). This provides support for the proposals made on the basis of crystallographic studies, which suggested that a weak hydrogen bond to Asn 155 in the Michaelis complex is strengthened in the transition state (Robertus et al., 1972b; Poulos et al., 1976). Interestingly, mutation of Thr 220 of subtilisin showed that it stabilizes the transition state by 2 kcal/mol despite the fact that the side-chain O^{γ} lies 4.0 Å from the oxyanion, too far for a direct interaction (Braxton & Wells, 1991). One explanation for the influence of Thr 220 was proposed to be that dynamic fluctuations of the protein structure (Rao et al., 1987) cause transient direct interactions to occur. An alternative suggestion was that the oriented Thr 220 side-chain dipole may stabilize the transition state at a distance, by influencing the electrostatic potential in the active site. Significant perturbation of the pK_a of the catalytic His 64 results from mutation of charged surface residues some 12–20 Å distant from the active site (Russell et al., 1987; Loewenthal et al., 1993). Similar mutation of distant charged residues affects the stability of complex formation with a transition-state analog inhibitor (Jackson & Fersht, 1993). These observations support the hypothesis that long-range electrostatic interactions may play a small but significant role in stabilizing the catalytic transition state.

Considerable controversy has surrounded the role of an additional component of the catalytic apparatus, a conserved buried aspartate residue first described in the crystal structure of chymotrypsin (Matthews et al., 1967; Blow et al., 1969). Mutation of this residue confirmed its essential role, because all variants of trypsin and subtilisin in which the aspartate is absent are decreased in catalytic efficiency by at least a factor of 10^4 (Craik et al., 1987; Sprang et al., 1987; Carter & Wells, 1988; Corey & Craik, 1992). The early suggestion of a two-proton transfer model, in which the Asp accepts a proton to become uncharged in the transition state, now appears to be unsupported by the bulk of the experimental (Bachovchin & Roberts, 1978; Markley, 1979; Kossiakoff & Spencer, 1981) as well as theoretical (Warshel et al., 1989) evidence. One role for the conserved Asp appears to be ground-state stabilization of the required tautomer and rotamer of the catalytic His (Craik et al., 1987; Sprang et al., 1987). In addition, because the His imidazole ring acquires a proton in the transition state, the Asp carboxylate can provide compensation for the developing positive charge. Its role may therefore be considered similar to that of the hydrogen bond donor groups in the oxyanion hole, which compensate the developing negative charge on the substrate carboxyl oxygen atom (Warshel et al., 1989; Fig. 2A,B). Experimental evidence for the role of electrostatic stabilization of the trypsin transition state has been obtained by mutation of the conserved Ser 214, which forms a solvent-inaccessible hydrogen bond to Asp 102, to various charged and uncharged amino acids (McGrath et al., 1992). Decreases in the free energies of catalysis were in agreement with electrostatic calculations, based on crystal structures of the mutants, which predicted these losses of activity.

Comparative analysis of the structures of chymotrypsin, subtilisin, and serine carboxypeptidase shows that the precise geometric orientation of the Asp is not conserved relative to the Ser–His catalytic diad (Liao et al., 1992; compare Fig. 1A,B,C). In contrast to chymotrypsin and subtilisin, the plane of the Asp carboxylate in carboxypeptidase is tilted far out of the plane of the His imidazole, such that the His–Asp hydrogen bond is 45° out of the carboxylate plane. This geometry is unfavorable for proton transfer from His to Asp and provides further evidence against the double proton-transfer mechanism. A detailed analysis of high-resolution subtilisin structures also showed differences in the Asp–His hydrogen bonding relative to trypsin (McPhalen & James, 1988). It now appears that the Asp can occupy virtually any position relative to the Ser–His diad. Therefore, it may be more accurate to regard the operation of the serine protease catalytic machinery as two diads—Ser–His and His–Asp—that operate in concert, rather than as a single catalytic triad (Liao et al., 1992). In this context, it is of interest to note that relocation of the Asp 102 carboxylate group to position 214 in trypsin significantly reconstitutes the activity lost in the variants D102S and D102N (Corey et al., 1992). The crystal structure of this mutant shows that Asp 214 still interacts with His 57, but in an altered geometric orientation in which the plane of the carboxylate is displaced from that of the imidazole ring by 40°. The relatively high catalytic efficiency of this variant thus supports the view of the catalytic apparatus as a juxtaposition of two diads.

Substrate specificity in the subtilisin family

The catalytic machinery and substrate binding clefts of the subtilisin-class serine proteases are embedded in a single-domain molecule (Wright et al., 1969; McPhalen & James, 1988). Six crystal structures are available in this family: *Bacillus amyloliquefaciens* subtilisin BPN′ (Novo) (Wright et al., 1969; McPhalen & James, 1988), *Bacillus licheniformis* subtilisin Carlsberg (Bode et al., 1986a; McPhalen & James, 1988), *Thermus vulgaris* thermitase (Gros et al., 1989), *Thermus album* proteinase K (Betzel et al., 1988), *Bacillus lentus* alkaline protease (Betzel et al., 1992), and *Bacillus alcalophilus* alkaline protease (van der Laan et al., 1992). The central core of the globular heart-shaped molecule is formed by a seven-stranded parallel β-sheet (Fig. 1B). Nine α-helices are packed against the sheet in a mostly antiparallel fashion relative to the β-strands; seven of these are on the same face and form the larger of two subdomains defined on either side (McPhalen & James, 1988). A two-stranded antiparallel β-sheet is also formed in the larger subdomain near the C-terminus of the chain. The active site is located in the larger subdomain adjacent to the central β-sheet; the catalytic Ser 221 is found near the amino-terminus of a long α-helix, which follows the small antiparallel sheet (Fig. 1B; McPhalen & James, 1988; numbering system for SBPN is used throughout).

Nearly all of the secondary structure elements of the enzymes are very highly conserved. A central core of 194 amino acids has been defined by comparison of the known structures, which contains nearly all of the conserved α-helices and β-strands (Siezen et al., 1991). The fungal-derived PROK deviates most significantly in structure but still superimposes these equivalent Cα atoms with RMS deviation of about 0.9 Å (the other prokaryotic enzymes superimpose at 0.4 Å to 0.65 Å; Siezen et al., 1991). If PROK is omitted, a more extended core of 232 amino acids can be defined among the bacterial species of known structure. An extensive sequence comparison of 47 subtilisin-class enzymes showed a subdivision into two subclasses, based on conserved differences in certain parts of the alignment. SBPN, SCARL, THERM, BAP, and BLAP are members of subclass I; the structurally divergent PROK is a representative of subclass II (Siezen et al., 1991). Although the homologous catalytic core of some

270 amino acids is found in all subtilisins, some of the enzymes possess large insertions in this domain, and many also possess C-terminal extensions resulting in polypeptide chains as long as 1,775 amino acids. This large database of sequence information forms the basis for homology modeling of those enzymes for which no tertiary structure is available (Siezen et al., 1991, 1993).

Crystal structures of enzyme–inhibitor complexes have identified substrate binding determinants extending over nine amino acids, from P6 to P3′. The structures include several peptide chloromethyl ketone complexes, in which subsites P1–P3 are occupied (Robertus et al., 1972a; Poulos et al., 1976), as well as complexes of SCARL with the protein inhibitor eglin C (Bode et al., 1986a; McPhalen & James, 1988), SBPN with eglin C, chymotrypsin inhibitor 2 and *Streptomyces* subtilisin inhibitor, (Bode et al., 1986a; McPhalen & James, 1988; Takeuchi et al., 1991a, 1991b), THERM complexed to eglin C (Gros et al., 1989), and PROK complexed with peptide inhibitors (Betzel et al., 1993). In each of these complexes, the inhibitor chain binds in a surface channel of the enzyme, which accommodates six residues from P4 to P2′. On the N-terminal side of the scissile bond, the P1–P4 residues of the substrate main chain are invariably inserted between two β-strands of the enzyme at positions 125–127 and 100–102 (Fig. 2B). The substrate thus forms the central strand of a three-stranded antiparallel sheet unique to the subtilisins; in the chymotrypsin-like proteases, this structure is not formed because only the strand corresponding to residues 125–127 is present (Fig. 2A).

Subtilisins in general show broad substrate specificity profiles and often display a preference for large hydrophobic groups at position P1 (Markland & Smith, 1971). At this position specificity arises from a broad open S1 binding cleft formed on one side by the two β-strands, which interact with the P1–P4 substrate residues, and on the other by a loop comprising residues 155–166 (Fig. 3). This loop varies in size among members of the family (Siezen et al., 1991). In SBPN, two different modes of binding exist to accommodate either P1-Phe or P1-Lys substrates (Robertus et al., 1972a; Poulos et al., 1976). The Phe ring binds deeply in the S1 cleft near Gly 166, whereas the charged Lys extends across the cleft to form a salt bridge with Glu 156. A prominent hydrophobic cavity is also present for binding of the P4 substrate side chain (Fig. 3). These two sites have been the focus of much of the work on substrate specificity. Interactions made in the more distal sites influence catalytic efficiency markedly, and there is evidence for nonadditivity of mutational effects suggesting a functional communication between sites (Grøn & Breddam, 1992).

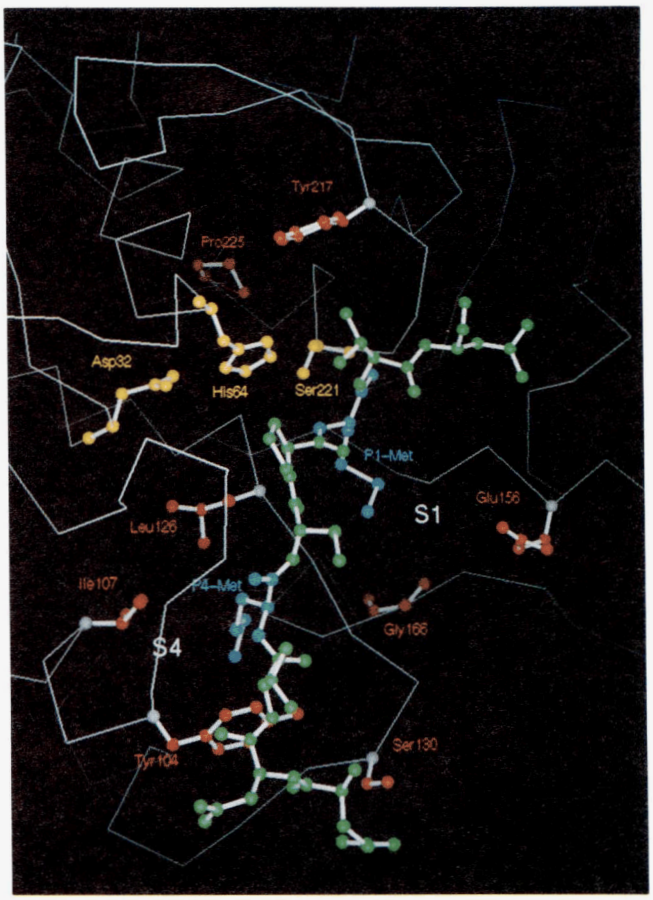


Fig. 3. Structure of the S1 and S4 sites of subtilisin BPN′ showing binding of a peptide derived from the cocrystal structure with *Streptomyces* subtilisin inhibitor. An α-carbon trace of the protein is shown in thin blue lines. Catalytic residues are in yellow, and the inhibitor chain is in green with the P1 and P4 side chains labeled in blue. Locations of amino acids at which the S1 and S4 sites have been mutated are indicated in red. In the subtilisin family, both the S1 and S4 sites are generally specific for hydrophobic side chains, but Glu 156 in the S1 site of subtilisin BPN′ provides activity toward P1-Lys side chains as well. At both sites, specificity alteration is readily achievable by the substitution of a small number of residues directly in contact with substrate. Modulation of the hydrophobic specificity profiles has been achieved at both sites, and altered specificity toward charged residues has been achieved in the S1 pocket.

Interactions in the S1 site

The most intensively studied member of the subtilisin family is SBPN, which has been the subject of extensive protein engineering investigations (reviewed in Wells et al., 1987b; Wells & Estell, 1988). The enzyme efficiently cleaves peptidyl amide substrates possessing a broad range of P1 amino acids, with the k_{cat}/K_m value showing a linear dependence on the hydrophobicity of the substrate side chain. The preference of the enzyme at this position is roughly Tyr, Phe > Leu, Met, Lys > His, Ala, Gln, Ser ≫ Glu, Gly (Estell et al., 1986; Wells et al., 1987c). To investigate the role of hydrophobicity more closely, 12 different amino acids were substituted for Gly 166, which lies at the base of the pocket (Fig. 3). Analysis of the mutants showed that an increase in the side-chain volume at this position, which consequently decreases the size of the S1 cleft, caused substantial reductions (up to 5,000-fold) in k_{cat}/K_m toward large P1 amino acids. This presumably occurs due to steric repulsion, which predominates over the favorable effect of a more hydrophobic pocket. Catalytic efficiencies toward small P1 side chains were increased by up to 10-fold in these variants. An optimal combined volume for the S1 and P1 side chains of 160 Å^3 was estimated from these data. It appears that hydrophobicity of the S1 site is the main driving force for specificity, whereas other effects, such as attractive van der Waals forces and hydration of polar side chains, have a lesser though still significant role.

Because these studies showed that specificity is easily modulated by replacing amino acids directly contacting substrate, it seemed plausible that more distant portions of the enzyme struc-

ture might be of little importance. This idea was further explored by a mutational study in which several amino acids from the related SCARL enzyme were exchanged for those in SBPN (Wells et al., 1987a). Although these two enzymes differ by 31% in sequence, only three substitutions lie within 7 Å of the S1 pocket. Two of these, at positions 156 and 217 (Fig. 3), directly contact substrate (residue 217 is in the S1′ site). A third residue at position 169 is positioned behind the loop comprising residues 156–166, which forms one side of the S1 pocket. In SBPN the amino acids are Ser 156, Ala 169, and Leu 217; these replaced the analogous Glu 156, Gly 169, and Tyr 217 of SCARL. The wild-type enzymes differ by factors of 6–60-fold in their k_{cat}/K_m values toward peptidyl amide substrates possessing P1-Glu, Met, Phe, Gln, or Ala; in each case, SBPN is more efficient (Wells et al., 1987a).

The triple mutant E156S/G169A/Y217L was found to exhibit a substrate specificity profile very similar to that of SCARL. Cleavage at each of the P1 amino acids tested occurred with efficiencies within threefold of the target protease (Wells et al., 1987a). These data demonstrate that, of the 86 amino acid differences between the two enzymes, three alone are largely sufficient to determine the differences in specificity. Further, analysis of singly and doubly substituted variants showed that the E156S mutation is alone almost entirely responsible for the shift in specificity profile. Because the activity of the E156S/Y217L enzyme was found to be within twofold of the triple mutant, it appears P1 substrate specificity is in fact locally determined to a significant degree.

The behavior of the E156S variant is similar to that of other mutant SBPN enzymes also possessing electrostatic substitutions in the S1 site (Table 1; Wells et al., 1987c). Sixteen variants were constructed at positions 156 and 166, each of which altered the electrostatic potential of the S1 site by introducing or removing Arg, Lys, Glu, or Asp residues at one or both sites. Analysis of the mutants showed that increases as high as 10^3-fold in k_{cat}/K_m toward complementary charged substrates could be achieved. To assess the contribution of electrostatic free energy to the stabilization of the transition-state complex, parallel substitutions of roughly isosteric but uncharged residues (Met replacing Lys; Gln replacing Glu) were also made. For example, it was found that increasing the positive charge in the S1 site increases k_{cat}/K_m much more for P1-Glu than for P1-Gln substrates. In this way, substrate binding effects associated solely with the charge–charge interaction could be isolated.

Table 1. *Engineering electrostatic interactions in subtilisin*[a]

	Net charge	P1-Glu	P1-Lys
E156D166	−2	–	16,200
E156N166	−1	40	17,800
E156Q166	−1	16	12,600
S156D166	−1	17	17,400
E156G166 (wt)	−1	35	39,800
Q156G166	0	620	1,070
Q156N166	0	110	5,600
E156R166	0	810	1,550
Q156K166	+1	66,000	1,700
S156K166	+1	16,200	5,400

[a] Substrate: *suc*-Ala-Ala-Pro-Glu/Lys-*p*NA. k_{cat}/K_m, s^{-1} M^{-1}.

Several of the S1-site specificity variants were also utilized in a different study that addressed the ability of SBPN to function as a peptide ligase (Abrahmsen et al., 1991). This reaction occurs when peptides bearing a free amino-terminal group can compete effectively with water for attack on the acyl-enzyme intermediate. The intrinsic low level of ligase activity normally present in SBPN was enhanced by substitution of the active-site Ser 221 by Cys, which shifts the relative preference toward aminolysis by more than 10^3-fold (Nakatsuka et al., 1987). The additional mutation P225A improves ligase activity by an additional 10-fold (Abrahmsen et al., 1991). The usefulness of this SBPN variant (referred to as subtiligase) for the synthesis of proteins was improved by introducing specificity variants G166I, G166E, and E156Q/G166K into the S221C/P225A framework. Preferred ligation of P1-Glu, P1-Phe, P1-Lys, and P1-Arg esters was achieved; the specificity for ligation mirrored that for cleavage of peptidyl amide substrates (Estell et al., 1986; Wells et al., 1987c). The ability to modulate the S1-site specificity thus provides greater flexibility in the choice of ligation junctions. Subtiligase has been used to synthesize ribonuclease A and active-site variants of this enzyme by stepwise ligation of six esterified peptide fragments 12–30 residues long (Jackson et al., 1994).

Substrate-assisted catalysis

Substrate-assisted catalysis represents a strategy for enhancing the specificity of proteolytic cleavage. Subtilisins lacking the catalytic His 64 can be reconstituted by including a histidine residue within the substrate (Carter & Wells, 1987; Carter et al., 1989, 1991). By placing a His at the P2 position of peptidyl amide substrates, specificity of up to 400-fold was achieved relative to analogous P2-Gln and P2-Ala substrates. The increased specificity at position P2 occurs within the context of a compromised enzyme: H64A subtilisin is reduced 10^6-fold in k_{cat}/K_m, and H64A in the presence of a P2-His substrate remains 5,000-fold less efficient than the wild-type enzyme (Carter & Wells, 1987). Mutation of Ser 221, Asp 32, and Asn 155 in the context of H64A suggested that interactions of the catalytic His with the Ser and Asp residues are severely compromised when the His is present in the substrate (Carter et al., 1991). By contrast, the oxyanion hole interactions appear much less disrupted. Model-building of P2-His substrates indicates that the imidazole ring can occupy roughly the same position as that of His 64 in the native enzyme, although some deviation in hydrogen bond distances and angles exists, which may partially explain the reduced activity.

The large database of S1-site specificity variants was again used to enhance the selectivity of proteolytic cleavage by the prototype H64A enzyme (Carter et al., 1989). For example, an improvement of 20-fold in cleavage of *suc*-FAHY-*p*NA was observed by introducing the S1 and S1′-site mutations E156S, G169A, and Y217L (Estell et al., 1986; Wells et al., 1987c), which increase catalytic efficiency toward P1-Phe and P1-Tyr substrates. The additional mutation G166A enhanced specificity for P1-Phe but not P1-Tyr substrates, as expected because the C^β of Ala 166 appears to cause steric hindrance to the binding of the larger Tyr side chain. Little specificity was observed on the

C-terminal side of the peptide bond in the cleavage of peptide substrates. The mutant subtilisins have been shown to selectively cleave designed target sites in fusion proteins, even under adverse conditions, making them a useful additional tool in the repertoire of protein chemists (Carter et al., 1989).

Further insight into substrate-assisted catalysis was provided by a novel approach using phage display technology (Matthews & Wells, 1993; Fig. 4A). A randomized target substrate sequence for an improved H64A subtilisin (Carter et al., 1989) was inserted between an amino-terminal affinity domain representing a variant of human growth hormone, and the carboxy-terminal domain of the M13 phage gene III coat protein. A collection of phage particles bearing different substrate sequences is bound to immobilized hGH-binding protein and cleaved by subtilisin, so that phage bearing good substrate sequences are eluted and those bearing poor sequences remain bound. Propagation of the phage further enriches for efficient or inefficient cleavage sites. Analysis of the sequences that were efficiently cleaved revealed that P1′-His as well as P2-His-containing substrates could function in substrate-assisted catalysis. Analysis of cleavage of fusion proteins linked to alkaline phosphatase, which provides an easily assayed activity, suggested that P1′-His-mediated cleavage was comparable in efficiency to P2-His cleavage. Further study of P1′-His cleavage would be informative because release of the leaving group after formation of the acyl-enzyme implies that no catalytic His is present to assist in deacylation. Molecular modeling has shown that a P1′-His can also occupy the position vacated by His 64 in an H64A variant (Matthews & Wells, 1993).

The P4–S4 interactions

Considerable specificity toward substrate residues distant from the scissile bond exists in the subtilisin-class family. A thorough mapping of the preferences of two enzymes—SBPN and BLAP—shows that the most marked distal interaction occurs on the N-terminal side of the substrate at the S4 enzyme site (Grøn et al., 1992). Mutational analysis at this position has been applied to three of the enzymes of known structure: SBPN (Eder et al., 1993; Rheinnecker et al., 1993, 1994), BLAP (Bech et al., 1992, 1993; Sørensen et al., 1993), and BAP (Teplyakov et al., 1992). The S4 site is formed from the juxtaposition of two structural elements: residues 100–107 at the amino-terminus of an α-helix in the small subdomain and residues 125–132 in an adjacent surface loop. Substrate interactions include both the main-chain β-sheet hydrogen bonds as well as contacts with the side chains of residues 104, 107, 126, and 135, which line the sides and base of the site (Fig. 3). Of the amino acids shaping the cleft, only Gly 127 is invariant in the family (Siezen et al., 1991).

In SBPN, the amino acid side chains in the S4 site are Tyr 104, Ile 107, and Leu 126, which create a large hydrophobic pocket. Accordingly, the substrate preferences follow the series Phe > Leu, Ile, Val > Ala for cleavage of peptidyl amide substrates (Rheinnecker et al., 1993). Slightly different preferences following the same general trend were observed toward long peptides occupying subsites S5–S5′ (Grøn et al., 1992). However, the range of k_{cat}/K_m values varies only over a three- to sixfold range. It was suggested that the small variability might be due to compensatory shrinkage of the S4 site upon binding of smaller side chains (Takeuchi et al., 1991a). Efficiencies toward polar resi-

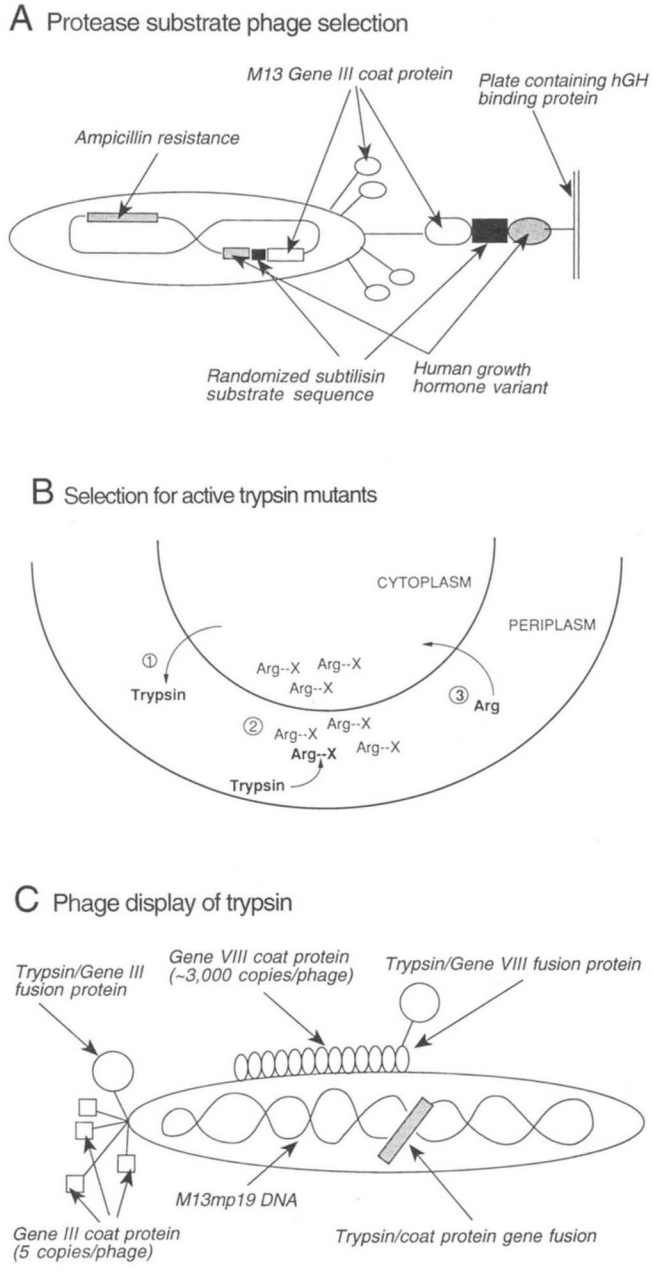


Fig. 4. Randomization methodologies employed in isolation of serine protease substrate specificity mutants. **A:** "Substrate phage" approach applied to subtilisin. In this method, the sequence of the substrate rather than the enzyme is varied to explore the substrate specificity at many of the subsites. By using H64A subtilisin as the cleaving protease, it was discovered that substrate-assisted catalysis functions when the substrate His is present at the P1′ as well as the P2 position. Note that in phage display systems, the phage particle provides a "package" in which the mutant DNA and variant protein are physically linked. This facilitates analysis after enrichment of those phage bearing good substrate sequences. **B:** Genetic selection for the isolation of trypsin variants. Periplasmic expression of a variant trypsin capable of cleaving the nonnutritive Arg-X substrate (1, 2) leads to release of free Arg (3), which enters the cytoplasm and relieves auxotrophy. Twenty variant trypsins possessing altered Arg/Lys specificity ratios have been isolated in this manner. **C:** Phage display approach for the isolation of trypsin variants. A wild-type trypsin gene fused to the M13 gene III coat protein specifically binds immobilized ecotin, a dimeric protein inhibitor of mammalian serine proteases that is found in the bacterial periplasm.

dues are decreased by more than 100-fold relative to hydrophobic amino acids (Grøn et al., 1992).

Tyr 104, Ile 107, and Leu 126 were mutated singly and in combination to amino acids that in every case were smaller than the wild-type residue. The following variant enzymes were characterized kinetically toward amide substrates of the form *suc*-XAPF-*p*NA: Y104F, Y104A; I107G, I107A, I107V; L126G, L126A, L126V, and the double mutants I107G/Y104A, I107G/L126A, I107G/L126V (Rheinnecker et al., 1993, 1994). These alterations test the effects of enlarging the P4 pocket as well as the consequences of deleting a hydrogen bond present between the side chains of Tyr 104 and Ser 130.

It was found that the Tyr 104–Ser 130 hydrogen bond has little effect on enzyme efficiency or specificity: Y104F SBPN hydrolyzes P4-Ala, Val, Ile, Leu, and Phe substrates nearly identically to the wild-type enzyme. The effect of introducing Ala at this position is similar to that caused by decreasing the size of Ile 107: in each case specificity is increased for residues possessing large side chains at P4. Among the single mutants at positions 104 and 107, the largest improvements in the relative specificity for P4-Phe relative to P4-Ala are roughly 200-fold for both Y104A and I107G. For these variants, the effects are achieved by maintaining approximately wild-type levels of k_{cat}/K_m toward Phe and sharply decreasing efficiencies toward Ala and the other smaller substrate residues. Mutation of Leu 126 had smaller effects on relative specificities, but large decreases in the range of $10–10^4$-fold were observed in k_{cat}/K_m, with decreased efficiency correlated with decreasing size of the side chain.

The three double mutants also showed strong preference for large side chains at position P4 (Rheinnecker et al., 1994). Among these enzymes, the mutant I107G/L126V improves the P4-specificity for large side chains to 340-fold relative to P4-Ala, but in this case the maximal discrimination was achieved with P4-Leu rather than P4-Phe. The other two double mutants similarly exhibited a maximal preference for P4-Leu. In all cases, nonadditivity was observed relative to the single mutants, as expected from the close proximity of the three side chains. Kinetic parameters were also measured toward the single-residue substrate acetyl-tyrosine ethyl ester, which might be considered as a probe measuring the extent to which S4-site mutants affect the functioning of the S1 site. Large decreases of up to 60-fold were observed, with the largest effects occurring for the double mutants. However, the same variants exhibit comparable efficiencies to wild-type when measured toward favored *suc*-XAPF-*p*NA substrates. This suggests that less productive binding may occur in the absence of the subsite interactions, particularly because the ester substrate is more easily cleaved owing to the better leaving group.

The substrate preference of BLAP at the P4 substrate position is also toward large hydrophobic side chains (Grøn et al., 1992). A broader range of specificities exists than in SBPN: in this case, a 24-fold (rather than sixfold) increase in k_{cat}/K_m when progressing from small to large hydrophobic amino acids is observed. The individual subsite interactions do not affect the overall catalytic efficiencies in an additive manner, suggesting that functional communication occurs and is mediated by structural elements of the protein (Grøn & Breddam, 1992). For example, modest substrate preferences at some sites are masked if the optimal P1-Phe and/or P4-Phe residues are present. These amino acids dominate the cleavage efficiency such that an upper limit in k_{cat}/K_m is reached even when other subsites are filled by nonpreferred residues. These other sites are therefore less important when a good substrate rather than a poor substrate is bound. This study underlines an important principle: optimal subsite mapping of subtilisins (and other proteases) should be carried out using sets of matched substrates where the interdependency of binding sites is not manifested. In the case of BLAP, the presence of an anthraniloyl group at P5 and a Pro at P2 apparently disrupts the P1-Phe and P4-Phe interactions, such that a substrate series containing these nonoptimal groups permits distribution of P1′ site preferences over a 15-fold range. Only a 50% difference between the most and least favored P1′ amino acid is observed in the absence of the nonoptimal groups, which prevents accurate mapping of the true subsite preference (Grøn & Breddam, 1992).

The structure of the BLAP S4 pocket is similar to that of SBPN. The side chains of Val 104, Ile 107, Leu 126, and Leu 135 form the base and one side of the pocket, whereas Ser 128, Ser 130, and Ser 132 are situated along the outside rim with each of the side-chain hydroxyl groups pointing inward. The substitution of Val 104 for the Tyr present in SBPN allows Leu 135 access to the substrate in BLAP. The only other difference in the pocket between the two enzymes is the presence of Gly 128 rather than Ser 128 in SBPN. A total of 21 mutants in the BLAP S4 site have been constructed and analyzed (Bech et al., 1992, 1993; Sørensen et al., 1993). At position 104 it was found that bulky hydrophobic side chains produced enzymes that preferentially cleaved small hydrophobic side chains, and conversely, smaller amino acids increased specificity toward large substrates. This behavior is reminiscent of the effects caused by increasing the size of residue Gly 166 in the S1 site of SBPN (Estell et al., 1986; see above). Mutations at other positions in the BLAP S4 site often also showed these effects, but in many cases complex specificity profiles not immediately interpretable in simple terms were obtained. What does appear clear is that both steric and hydrophobic effects play important roles in determining the S4 specificity profile (Bech et al., 1993; Sørensen et al., 1993). For some mutants it was further suggested that structural flexibility is also critical.

Distinguishing the degree to which hydrophobicity, steric exclusion, and substrate-induced conformational changes function to determine specificity profiles requires high-resolution structural information on the mutant enzymes. Such information has begun to be obtained in the study of BAP variants (Teplyakov et al., 1992). Substitution of Val 104 in this enzyme with Trp increased activity toward *suc*-AAPF-*p*NA by 12-fold. The crystal structure of the uncomplexed variant showed that no other structural change occurs and that the S4 site is now blocked off such that a modeled P4-Ala residue makes a good van der Waals contact with Trp 104. Trp 104 in this variant is oriented nearly identically to Trp 104 in THERM, which also exhibits high activity toward *suc*-AAPF-*p*NA.

Comparison of the structures of SSI and a P4-Met to Gly mutant of SSI complexed to SBPN showed that the S4 site undergoes a substantial shrinkage upon binding of P4-Gly (Takeuchi et al., 1991b). The structural flexibility in this enzyme raises the possibility that a capacity for such rearrangement may exist in other members of the family as well. Required for an assessment of the degree of flexibility, and the extent to which amino acid alterations affect this property, are crystal structures of wild-type and mutant enzymes complexed to substrate analogs pos-

sessing small and large side chains at the P4 position. In the case of BAP, for example, it would be of interest to determine the catalytic efficiencies of the wild-type and V104W enzymes toward larger hydrophobic P4-side chains and then to carry out a systematic structural analysis of complexes of each enzyme with analogous inhibitors. Such an analysis for the chymotrypsin-like α-lytic protease has yielded substantial insight into the structural basis for enzyme flexibility (Bone et al., 1991; see below).

Together these mutational alterations within the subtilisin S1 and S4 sites allow two important conclusions: (1) only the local environment of amino acids directly contacting substrate need be considered in designing specificity changes; (2) there is no important distinction between hydrophobic and polar enzyme–substrate interactions because each type is manipulatable to generate new specificity profiles while maintaining high activity. The importance of these generalizations to protein design in other systems depends upon the extent to which the structural design of the binding cleft, and the nature of the reaction being catalyzed, are crucial parameters. As we shall see, structural context can have great influence in mediating the extent to which specificity alteration is straightforward. A clue to its important role can be seen in the dependence of catalytic efficiency on the extent to which subsites are filled. The signal that distal portions of substrate are bound is transmitted over large distances and must in some way be mediated by the intervening protein structure. Long-range effects are key in the chymotrypsin family of enzymes, both in terms of filling subsites as well as in determining specificity at a single site (Corey et al., 1992; Hedstrom et al., 1992, 1994a, 1994b; Perona et al., 1995; see below).

Prohormone convertases: Specificity toward paired dibasic residues

Tissue-specific processing of precursor proteins in mammalian cells is accomplished by a subfamily of subtilisin-class enzymes known as prohormone convertases. The need for this cleavage event to release bioactive products provides a crucial regulatory step for the cell. Early protein sequencing studies of various peptide hormones suggested that the dibasic sequences Lys–Lys and Lys–Arg provided the sites of cleavage (reviewed by Lazure et al., 1983). The first protease isolated in this class was the yeast kexin, which cleaves with high selectivity both synthetic peptide and protein substrates possessing Lys–Arg at the P2 and P1 sites, respectively (Fuller et al., 1989; Brenner & Fuller, 1992). Following isolation of the yeast enzyme a number of mammalian species have been cloned including furin (Van den Ouweland et al., 1990), PC1/PC3 and PC2 (Smeekens et al., 1991), and more recently the enzymes PC4, PC5, and PACE4 (Rehemtulla et al., 1993). The enzymes possess pro-domains and must therefore themselves be processed prior to activation. Maturation has been shown to occur in an autocatalytic fashion in the cases of PC2 (Matthews et al., 1994) and of furin (Creemers et al., 1993). These studies have now shown that most cleavage takes place either at Lys–Arg and Arg–Arg dibasic sites, or at an Arg-X-Lys-Arg consensus site, depending on the intracellular pathway of localization.

Mature prohormone convertases are large enzymes that typically possess 600–800 amino acids. In addition to the subtilisin-like catalytic domain, they also variously possess other structural elements such as transmembrane anchors, Ser/Thr-rich regions, glycosylation sites and Cys-rich regions (Seidah et al., 1991). Based on homology modeling, it was predicted that these enzymes possess a greatly increased number of negatively charged residues near the substrate binding cleft. Many of these amino acids are highly conserved (Siezen et al., 1991; Fig. 5). Their importance was tested by site-directed mutagenesis of furin, using processing of a peptide hormone in vivo as the functional assay (Creemers et al., 1993). The following residues were mutated: Asp 33, Asp 61, Glu 101, Asp 104, Glu 107, Glu 129, Asp 130, Asp 131, Asp 165, and Asp 209. Cleavage was assayed toward the wild-type hormone precursor as well as toward three mutants in which one of the positively charged amino acids in the cleavage site sequence P4-Arg–P3-Ser–P2-Lys–P1-Arg was altered to Gly or Ala. The ability of mutants to carry out autoproteolytic activation was also assessed.

Mutation of the P1-Arg in this sequence gave rise to prohormones that could not be processed either by wild-type or by any of the mutant furins, suggesting that a basic residue at this position is critical to recognition (Creemers et al., 1993). Several of the mutants possessed preferences for one of the three mutant prohormone substrates, implicating the Asp or Glu at that enzyme position in recognition of the substrate residue that was altered. Thus, Asp 33 is implicated in P2-site binding and Glu 107 in P4-site binding, in accord with modeling that predicts their locations adjacent to these substrate positions (Siezen et al., 1991). Mutation of Asp 165, predicted to lie at the base of the S1 site, abolished activity, as did removal of the negative charge

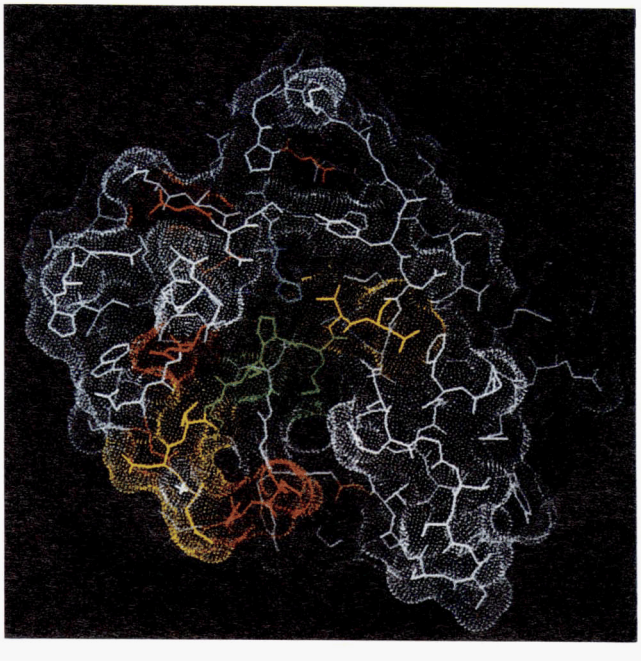


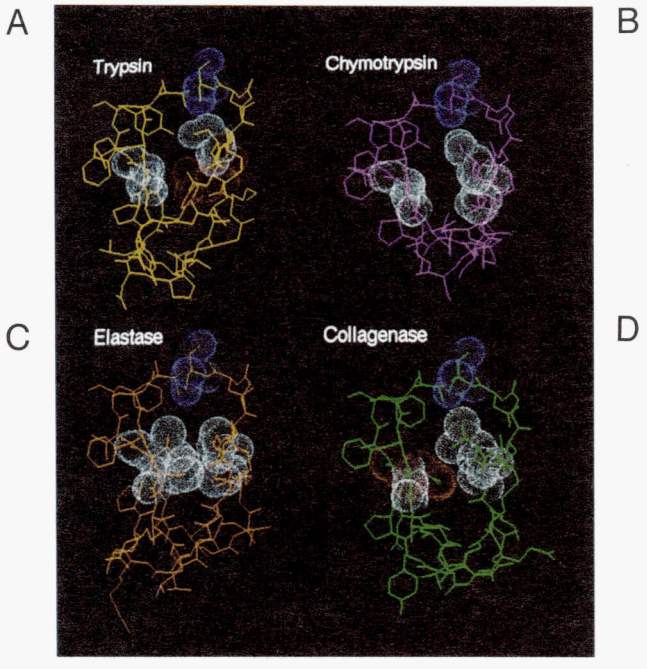
Fig. 5. A distinct subclass of the subtilisin family of serine proteases, the prohormone convertases, are involved in prohormone processing in a number of important physiological contexts. The specificity of processing is toward sites possessing 2–4 Arg and Lys residues at the P1–P4 positions. Shown is a solvent-accessible protein surface on which are mapped the binding determinants specifying prohormone processing by furin. The structure is that of subtilisin BPN′ complexed to SSI because no three-dimensional structure is yet available in this subclass. A large number of negatively charged amino acids is found on the substrate binding face of the enzyme (red). The catalytic triad is in blue and the substrate is in yellow, with the P1–P4 amino acids in green.

from positions Glu 129, Asp 130, or Asp 131 putatively near the P4 site. Interestingly, the roughly isosteric mutant D209L abolished activity, despite being located some distance from the binding cleft. By contrast, other substitutions nearer to the substrate could be introduced without loss of activity. These furin mutants provide the first mapping of structural determinants affecting prohormone processing. An obvious need now exists for an accurate three-dimensional structure of an enzyme in this class. Together with detailed kinetic analysis of synthetic substrates, this would provide substantial insight into the structural determinants of this most interesting specificity.

Substrate specificity in the chymotrypsin family

As in the subtilisin family of enzymes, the diversity of substrate specificity among the chymotrypsin-like proteases rests upon small differences in structure in the substrate-binding cleft. All of the chymotrypsin-like enzymes are composed of two juxtaposed β-barrel domains, with the catalytic residues bridging the barrels (Fig. 1A; Kraut, 1977; Steitz & Shulman, 1982; Bazan & Fletterick, 1990). Crystal structures are available for bovine chymotrypsin (Matthews et al., 1967), porcine pancreatic elastase (Watson et al., 1970), bovine, rat, and *Streptomyces griseus* trypsins (Ruhlmann et al., 1973; Sprang et al., 1987; Read & James, 1988), rat tonin (Fujinaga & James, 1987), kallikrein (Bode et al., 1983), rat mast cell protease II (Remington et al., 1988), human neutrophil elastase (Navia et al., 1989), thrombin (Bode et al., 1989a), factor Xa (Padmanabhan et al., 1993), and complement factor D (Narajana et al., 1994). Additionally, structures are available for four microbial enzymes: *S. griseus* proteases A, B, and E (SGPA, Delbaere et al., 1979; SGPB, Moult et al., 1985; SGPE, Nienaber et al., 1993), and the *Lysobacter enzymogenes* α-lytic protease (Brayer et al., 1979). The microbial enzymes share the chymotrypsin-like bilobal β-barrel structure but are more distantly related as evidenced by their shorter sequences and substantial structural differences in surface loops (James, 1976). *S. griseus* trypsin, on the other hand, is an example of a microbial enzyme that is more homologous to mammalian serine proteases than to its bacterial counterparts (Read & James, 1988).

Molecular modeling methods have been used to create a structure-based sequence alignment of the chymotrypsin-like serine proteases (Greer, 1990), which is very useful in assessing substrate preferences. The specificity is usually most pronounced at the S1-sites of the enzymes, where the majority of sequences group into one of three subclasses definable by inspection of a small number of crucial amino acids. Position 189, located at the base of the S1 pocket, is very highly conserved as an Asp in enzymes with trypsin-like specificity toward Arg- and Lys-containing substrates (Fig. 6; chymotrypsin numbering system is used throughout—see Greer, 1990). It is found as a Ser or other small amino acid in chymotrypsin and elastase-class enzymes, which manifest specificity toward aromatic and small hydrophobic amino acids, respectively. The amino acid side chains at positions 190 and 228 extend into the base of the pocket as well and play an additional role to modulate the specificity profile. Amino acids at positions 216 and 226 are usually Gly in both trypsin and chymotrypsin-like enzymes; larger amino acids at these positions partially or fully block access of large substrate side chains to the base of the pocket (Fig. 6). Accordingly, elastases possess larger, usually nonpolar residues at these positions,



E

	182	195	214	228
Bovine Chymotrypsin	CAG-ASGV-**S**SCMGDS	...	SW**G**SST-CS-TSTP**G**V**Y**	
Rat Trypsin	CVGFLEGGK**D**SCQGDS	...	SW**G**YG--CALPDNP**G**V**Y**	
Porcine Elastase	CAGGDG-VR**S**GCQGDS	...	SF**V**SRLGCNVTRKP**T**V**F**	
Crab Collagenase	CID-STGGK**G**TCDGDS	...	SF**G**AAAGCEA-GYP**D**A**F**	
	------		-----	
	Loop 1		Loop 2	

Fig. 6. Common architecture of the S1 site of four members of the chymotrypsin-like class of serine proteases, with the eponymous Ser 195 catalytic residue shown in blue. An early paradigm for substrate specificity was derived from a comparison of the S1-site structures of trypsin (**A**), chymotrypsin (**B**), and pancreatic elastase (**C**). Amino acids at positions 216 and 226 (left side of the pocket) and at 189 and 190 (right side) are indicated by van der Waals surfaces colored white for uncharged and red for negatively charged residues. The shape and electrostatic character of each site corroborate the specificities toward Arg/Lys, Phe/Tyr/Trp, and Ala, respectively. Fiddler crab collagenase (**D**) possesses a negatively charged Asp in an altered position relative to trypsin. Although it might be predicted that this enzyme possesses a trypsin-like specificity profile, it is instead capable of efficiently cleaving P1-side chains of substrates specific to each of the three other proteases. Amino acid sequence alignment of these four enzymes (**E**) showing the distinction in primary specificity residues (bold) and secondary determinants (underlined). Positions in the sequence of two adjacent surface loops are also shown (see Figs. 7, 11, 13).

providing a platform for interaction with small hydrophobic substrate P1-amino acids. The shapes of the S1 pockets of trypsin, chymotrypsin, and elastase thus appear to readily explain the observed specificities, leading to the canonical view that substrate preferences are in fact determined by this limited set of amino acids (Stroud, 1974). However, as discussed below, this perspective has now been shown to be incorrect by the discovery that other structural elements distant from the substrate binding site are also crucial determinants of specificity.

Kinetic measurements of substrate preferences for the two mammalian elastases of known structure (PPE and HNE) permit a more detailed appraisal of structure–function relationships

(Bode et al., 1989b). Both enzymes possess bowl-shaped hydrophobic S1 binding sites that accommodate small hydrophobic substrates (Watson et al., 1970; Navia et al., 1989). However, the S1 site of PPE has been described as slightly less hydrophobic and marginally smaller than that of HNE (Bode et al., 1989b). PPE cleaves peptide bonds preferentially at small P1-Ala and Nva side chains (Harper et al., 1984), whereas HNE manifests substantial activity toward the branched-chain Val, Ile, and Leu residues (Harper et al., 1984; Stein et al., 1987). These preferences are in accord with the smaller S1 site of PPE, but the small difference in size is insufficient to account for the altered profiles. The identity of the amino acids that line the S1 pockets differ substantially in the two enzymes, most notably by the presence of the charged Asp 226 in HNE, which is present as a Thr in PPE. In HNE, Asp 226 is buried by Val 216 and Val 190, and the carboxylate group points away from substrate into a network of buried water molecules (Navia et al., 1989). One possible explanation for the superior ability of HNE to cleave branched-chain substrates could thus be that the S1-site possesses greater intrinsic flexibility as a consequence of its different construction and interaction with surrounding portions of the structure (Bode et al., 1989b). A small shrinkage of the S1 site is in fact observed upon binding Val relative to Leu in this position (Bode et al., 1986b; Wei et al., 1988).

Cleavage of peptide substrates adjacent to the acidic Asp and Glu residues is the hallmark of an additional subclass of enzymes. Recognition of the negatively charged carboxylate is accomplished by means of a His residue at position 213 in a number of microbial enzymes including the *Staphylococcus aureus* V8 protease (Drapeau, 1978), SGPE (Svendsen et al., 1991), and two epidermolytic toxins of *S. aureus* (Dancer et al., 1990). Recently, the crystal structure of SGPE complexed with the tetrapeptide Ala-Ala-Pro-Glu has been determined at 2.0 Å resolution (Nienaber et al., 1993). The structure reveals that the Glu carboxylate is indeed bound directly by His 213 as well as by the side chains of Ser 192 and Ser 216. The structure of the enzyme also shows that His 213 is hydrogen bonded in series to two other His residues at positions 199 and 228 to form a solvent-inaccessible His triad that penetrates through the core of the enzyme. This remarkable structural feature is postulated to play a role in substrate charge compensation, by delocalizing the substrate negative charge through proton transfer across the His residues (Nienaber et al., 1993). No other serine protease is known to possess the His triad. An alternative to the use of His 213 is found in a protease from cytotoxic T-lymphocytes, which possesses an Arg at position 226 (Murphy et al., 1988). This enzyme is unusual in its preference for cleavage at Asp rather than Glu residues (Odake et al., 1991). Mutation of Arg 226 to Gly, followed by qualitative assay of crude lysates in which the variant was expressed, showed lowered activity toward peptidyl P1-Asp thiobenzyl ester substrates and increased activity toward analogous P1-Phe substrates (Caputo et al., 1994).

Virtually all chymotrypsin-like serine proteases share a common feature: an S1-site specificity that is restricted to a relatively narrow subset of the naturally occurring amino acids. It therefore came as some surprise when one enzyme, the collagenolytic serine protease I from the fiddler crab *Uca pugilator*, was shown to possess high catalytic activity toward each of trypsin, chymotrypsin, and elastase-like substrates (Grant & Eisen, 1980). The specificity profile of this enzyme has recently been reexamined in detail (Tsu et al., 1994). Crab collagenase exhibits 5% of elastase, 10% of chymotrypsin, and 65% of trypsin activity, as assessed by k_{cat}/K_m values toward peptidyl amide substrates possessing Ala, Phe, and Arg, respectively, at the P1 position. k_{cat} values toward each of these amino acids are extremely high. Additionally, it is the most efficient chymotrypsin-like enzyme known toward P1-Leu and P1-Gln amide substrates, manifesting 6-fold and 50-fold greater activities than does chymotrypsin toward these substrates (Tsu et al., 1994). Therefore, the chymotrypsin-like scaffold can maintain an S1 binding pocket that accommodates a very broad range of amino acids without sacrificing catalytic efficiency.

Crab collagenase exhibits an interesting rearrangement of a negative charge at the base of the S1 site: residues Asp 189 and Gly 226 of trypsin are altered to Gly 189 and Asp 226 in collagenase (Grant et al., 1980; Fig. 6). However, this predicts a strict specificity for P1-Lys and Arg substrates: the amino acids at positions 190 and 216 are Thr and Gly, respectively, which allows access of the substrate to Asp 226. As discussed above, Asp 226 of human neutrophil elastase is buried by Val 216, leading to a hydrophobic specificity profile (Navia et al., 1989). A possible explanation for the ability of crab collagenase to accommodate hydrophobic as well as positively charged substrate residues is provided by a recently refined 2.5-Å crystal structure of the enzyme complexed with the dimeric serine protease inhibitor ecotin (J.J. Perona, C.A. Tsu, C.S. Craik, & R.J. Fletterick, submitted for publication). The structure shows that one carboxylate oxygen of Asp 226 is accessible to substrate, but that the P1-methionine residue of ecotin does not enter the S1-site and binds instead on the surface of the enzyme adjacent to the disulfide bond at positions 191–220. Modeling shows that the pocket can provide multiple binding sites that accommodate diverse amino acid side chains in distinct positions. Therefore, S1-site flexibility does not appear to be utilized as a structural determinant in the broad specificity of crab collagenase.

α-Lytic protease: Exploring the role of structural plasticity in substrate specificity

α-Lytic protease, an extracellular enzyme produced by the soil bacterium *L. enzymogenes*, has been the subject of intensive analysis aimed at relating structure to catalytic activity. This microbial protease, while possessing the chymotrypsin-like fold comprising two β-barrels (Brayer et al., 1979), nevertheless displays large insertions and deletions relative to the pancreatic enzymes, resulting in an overall RMS deviation in the positions of structurally equivalent α-carbons of 1.36 Å for 110 of 198 amino acids, when compared with chymotrypsin (Fujinaga et al., 1985). By comparison, the equivalent pairwise fits with the bacterial proteases SGPA and SGPB yield RMS deviations of roughly 0.7 Å, a value very similar to that which relates the mammalian pancreatic enzymes to each other. The S1 pockets of α-lytic protease and trypsin are particularly divergent in structure (Fig. 7). An insertion of two amino acids causes Met 192 of α-lytic protease to occupy a position similar to Ser 190 of trypsin. More strikingly, an adjacent surface loop at positions 185–188 is deleted in α-lytic protease, and a second nearby loop at positions 217–225 is enlarged by eight amino acids. A consequence of these differences is that, although both enzymes possess a disulfide bond linking the conserved residues Cys 191 and Cys 220, the positions of the sulfur atoms are displaced by 7–8 Å (Fig. 7).

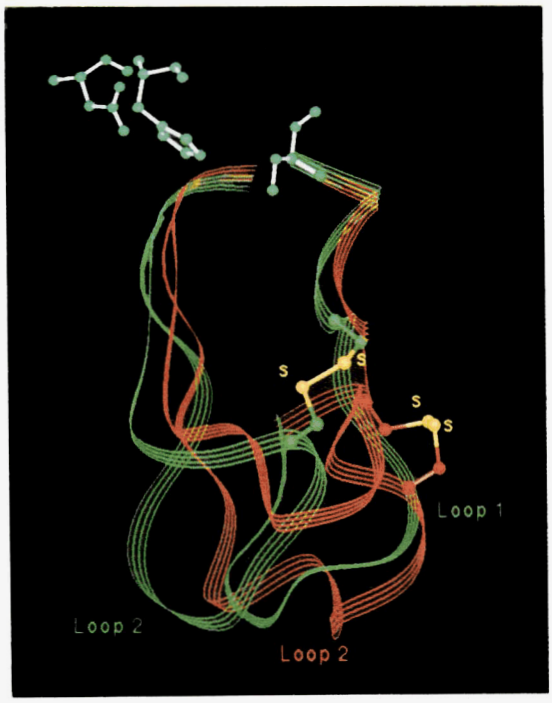


Fig. 7. Diversity in S1-site structure between the mammalian and the microbial trypsin-like enzymes is illustrated by a superposition of trypsin (green) and α-lytic protease (red). Although the mammalian enzymes such as trypsin possess two well-defined loops (loop 1 and loop 2) joining the β-strands of the specificity pocket, in α-lytic protease and other microbial enzymes loop 1 is absent, whereas loop 2 is greatly enlarged. Conserved disulfide bonds of each enzyme (Cys 191–Cys 220; yellow) are displaced some 7 Å from each other. The catalytic triad is shown at the top in green.

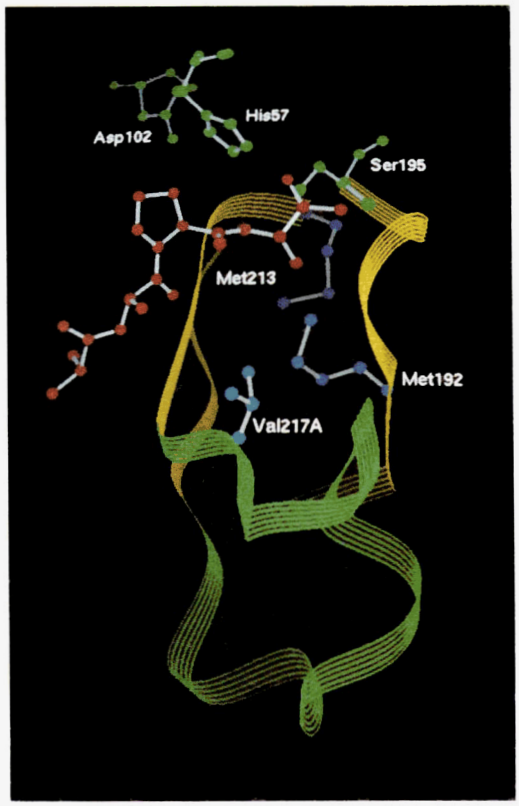


Fig. 8. Structure of the S1 site of α-lytic protease bound to the substrate analog *suc*-Ala-Ala-Pro-Ala-boronic acid (red), showing the positions of the hydrophobic amino acids Met 192, Met 213, and Val 217a, which form a platform for binding of small hydrophobic side chains. The three β-strands of the S1 site are shown in yellow and the large connecting ω-loop is in green. Catalytic groups are also in green (top). Mutation of either Met 192 or Met 213 to Ala creates variant enzymes possessing greatly broadened specificities toward hydrophobic amino acids, without sacrificing catalytic efficiency.

Kinetic data show that α-lytic protease possesses a hydrophobic specificity profile for substrate residues in the P1 position. The preference of the enzyme at P1, as described by relative k_{cat}/K_m values, is roughly Ala > Met, Val, Gly > Nle > Leu > Phe for hydrolysis of tetrapeptide amide substrates (Bauer et al., 1981; Bone et al., 1991). The structural elements that interact with the P1-substrate side chains comprise the three hydrophobic side chains Met 192, Met 213, and Val 217a, which together form a shallow depression in the enzyme surface (Brayer et al., 1979; Fujinaga et al., 1985; Fig. 8). More recently, six crystal structures of the enzyme complexed with peptidyl boronic acid inhibitors of the general structure R-boroX (where R is methoxysuccinyl-Ala-Ala-Pro and boroX is the α-aminoboronic acid analog of Ala, Val, Ile, Nle, Leu, or Phe) have been determined at resolutions between 2.0 and 2.5 Å (Bone et al., 1987, 1989a, 1991). Boronic acids are tight-binding (K_i's in the nanomolar range) reversible inhibitors of serine proteases (Kettner & Shenvi, 1984) that form covalent, nearly tetrahedral adducts with Ser 195 (Bone et al., 1987). They represent good structural analogs of the high-energy tetrahedral intermediate present on the actual catalytic pathway.

The crystal structures of the boronic acid complexes confirm that covalent tetrahedral adducts are formed with Oγ of Ser 195 for the P1-Ala, Val, Ile, Leu, and Nle inhibitors. The large P1-Phe side chain cannot fit into the S1-site, leading to the formation of an unusual trigonal adduct that includes His 57 (Bone et al., 1989a). The interactions of the inhibitor among these structures are nearly identical with the exception of the way in which the P1 side-chains interact with Met 192, Met 213, and Val 217a. These side chains adjust conformation in response to the differing sizes and shapes of the inhibitor amino acids. Small shifts in the position of adjacent main-chain atoms in the S1 and S2 specificity sites occur in the complexes with the larger Nle and Phe. Particular importance has been ascribed to the rearrangements at positions 217a–217d (Bone et al., 1989a, 1991; see below). Low activity toward the larger Leu and Phe side chains appears to arise solely from steric considerations, whereas Met is preferred to Leu presumably owing to its greater flexibility. Although the structural basis for the preference of Ala relative to Val was not unambiguously clear, it was proposed that strong binding to the oxyanion hole, required in the transition state, is prevented for the Val substrate on steric grounds. Differences in the electronic character of the boronate inhibitor, relative to a true transition state, do not allow for a complete mimicking of the latter (Bone et al., 1989a).

The substrate specificity profile of α-lytic protease was altered dramatically by the introduction of either of two single-site mutations in the S1 site: M192A or M213A (Bone et al., 1989b; Ta-

ble 2; Figs. 8, 9). In each case, high activity toward Ala was retained, but the increased size of the S1 pocket allowed accommodation of P1-side chains as large as Phe, with catalytic efficiencies k_{cat}/K_m increased up to 15-fold relative to wild-type cleavage at P1-Ala. For M192A, improved catalytic efficiencies toward P1-Met and P1-Val resulted mainly from lowered K_m values, whereas the P1-Leu and P1-Phe substrates were improved in both k_{cat} and K_m. The catalytic activity toward P1-Leu and P1-Phe substrates was improved by 10^4–10^6-fold, respectively, relative to wild type. However, the wild-type preference of nearly 10^5-fold for P1-Ala/Phe was decreased to 30-fold in M192A and nearly completely eliminated in M213A (Table 2). Complicating a straightforward interpretation of the profiles of these variants were two factors: (1) the dependence of k_{cat}, K_m, and k_{cat}/K_m was not correlated with the size or hydrophobicity of the P1 side-chain; (2) enlargement of the pocket by the same volume in the two mutants gave rise to considerably different functional effects. Therefore, extensive structural analysis of the mutant enzymes complexed with the boronic acid inhibitors was carried out to understand which factors cause the altered specificities (Bone et al., 1989b, 1991).

The principle rationale for the exceptionally broad specificity profiles of M192A and M213A is that the S1 site possesses structural plasticity, which encompasses a combination of alternate side-chain conformations as well as deformability of the main chain (Bone et al., 1989b; Fig. 9). For example, accommodation of the P1-Phe side chain by M192A results from a substrate-induced conformational change, in which the side chain of Val 217a rotates to remove one carbon from the pocket, and the main chain from Val 217a to Val 217d shifts by 0.5–0.8 Å. This permits the large inflexible aromatic ring to be nearly completely buried in the specificity pocket. In this case, some of the binding energy is presumably used to drive the conformational change in the protein, a phenomenon that is also observed to lesser extents in other mutant–inhibitor complexes. In general, hydrogen bond lengths, buried hydrophobic surface area, unfilled cavity volume, and the magnitude of conformational changes vary significantly among the various mutant and wild-type complexes (Bone et al., 1991). The energetic consequences of these differences were quantified (see Bone & Agard [1991] for a review of the energetics of intermolecular interactions) and correlated with free energies of catalysis for the various mutant–substrate combinations.

The analysis has led to an increased understanding of the way in which the different energetic terms can contribute to the stabilization of the enzyme–substrate complex, although no single factor has been found that consistently correlates well with either activity or inhibition (Bone et al., 1991). Thus, the wild-type enzyme has a relatively limited ability to adapt to large side chains, so that the specificity profile is driven primarily by steric exclusion. M192A, however, is improved in its ability to hydrolyze large side chains in part because the degree of conformational change required for their accommodation is reduced; further, it also possesses the ability to shrink so that P1-Ala substrates are hydrolyzed well. By contrast, the M213A pocket cannot contract, leading to a sharply reduced activity toward P1-Ala as well as a reduced discrimination relative to P1-Gly (Bone et al., 1991). In both mutants, however, the broad specificities depend on the ability of the main chain and side chain atoms at positions 217a–217d to readjust (Fig. 9). This flexibility is proposed to arise from a large adjacent surface loop, which begins at res-

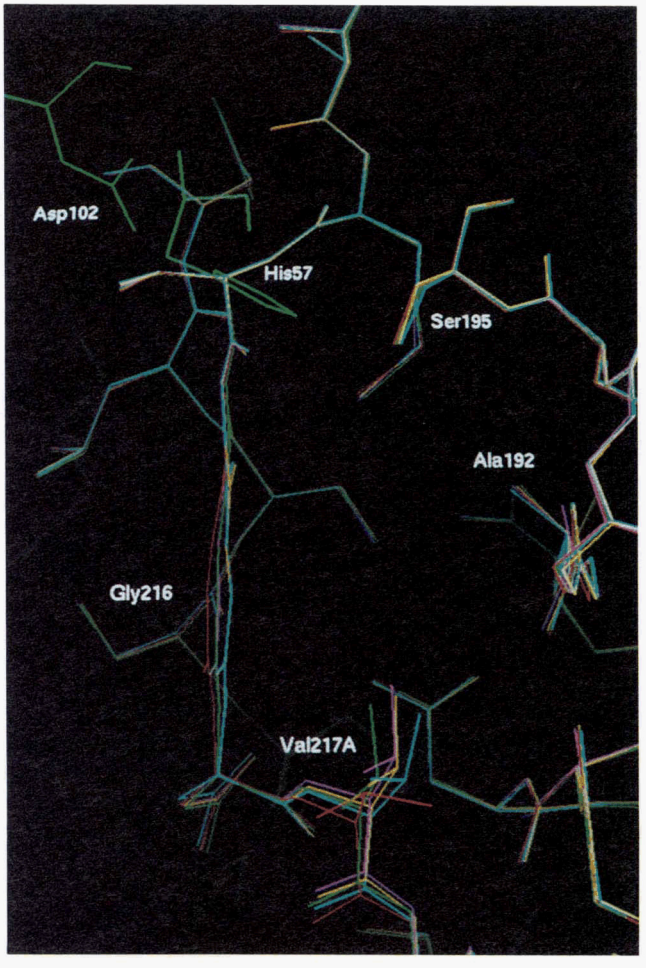


Fig. 9. Principal rationale for the ability of α-lytic protease mutants to exhibit greatly enhanced specificities toward new substrate side chains is structural plasticity of the S1 site. Shown is a superposition of five structures of the M192A variant of the enzyme (the new Ala 192 side chain is at the right side). Each enzyme is complexed with a peptidyl boronate inhibitor (not shown for clarity) possessing a particular hydrophobic P1-side chain (see Fig. 8 for inhibitor binding). The conformation of the active site adjusts to the different substrates at position Gly 216 and in the following loop region (bottom). Both side-chain and main-chain rearrangements are important components of active-site plasticity. The ability of the active site to adjust in this manner may be an important factor in the ability to effect specificity modification by mutation at only a single site.

Table 2. *Broadening the specificity of α-lytic protease*[a]

X	Wild type	M192A	M213A
Ala	21,000	10,000	600
Val	790	3,000	340
Met	1,800	35,000	980
Leu	4.1	11,000	160
Phe	0.38	31,000	340

[a] Substrate: *suc*-Ala-Ala-Pro-X-*p*NA. k_{cat}/K_m, s^{-1} M^{-1}.

idue 217a (Figs. 7, 8), and which appears to be able to absorb structural changes in the preceding residues. The energies of interaction of the S1 site with this and other peripheral structural elements thus also play a significant role in determining the specificity profiles.

Another recent study of α-lytic protease used random mutagenesis of four residues in the substrate binding pocket, coupled to an activity screen using synthetic substrates, to identify new variants with altered specificities (Graham et al., 1993). A library was constructed beginning with the M192A variant, with randomization of positions Gly 192a, Arg 192b, Met 213, and Val 217a. Screening and qualitative characterization of 47 active variants revealed that a majority of the enzymes retained a specificity profile similar to that of the parent M192A. Also emerging from the screen was a subclass of enzymes capable of cleaving P1-His-containing substrates. All mutants possessing this ability contained His 213, an amino acid heretofore correlated with P1-Glu specificity in other microbial enzymes (Nienaber et al., 1993). In general, residue 213 appears to play a significant role as a primary specificity determinant in several microbial enzymes. Although this amino acid has not yet been mutated in any mammalian protease, it appears very unlikely that it will assume a similar role. Clearly the divergence in structure of the S1 site in the two subclasses (Fig. 7) has led to a more prominent role for this residue in the bacterial enzymes, despite the fact that its position relative to the Ser 195/His 57 catalytic couple does not vary.

Kinetic data indicate that α-lytic protease makes substrate binding interactions over at least six subsites from P2′ to P4 (Bauer et al., 1981). Interestingly, the crystal structure shows that a small hydrophobic pocket exists beyond the P4 side chain of the tetrapeptide boronic acid inhibitor, formed from residues Leu 227, Leu 180, Val 167, Ala 169, and Ser 225 (Bone et al., 1987). Although extension of a substrate side chain to fill the S5 site does not have a significant influence on kinetic parameters (Bauer et al., 1981), it is possible that additional binding energy from interactions in the hydrophobic pocket cannot be realized in catalysis unless a P6 side-chain is also bound. Little specificity has been observed at the other subsites, although a preference for Pro at position P2 has been noted in binding of the peptide boronic acid inhibitors (Bone et al., 1987). Although the S2 enzyme site is hydrophobic, adjacent side-chain hydroxyl groups of Ser 214 and Tyr 171 participate in a hydrogen bonding network, which includes the carboxylate of Asp 102. Introduction of the mutations S214A and Y171F caused decreases in both k_{cat} and K_m, and the data were used to infer that the role of the two hydroxyl groups in the native enzyme is to facilitate catalysis by maintaining the S2 site in an optimal configuration (Epstein & Abeles, 1992).

Mutational analysis of trypsin: Combining structural genetics, classical enzymology, and X-ray crystallography

Trypsin represents the third serine protease that has been the subject of extensive mutational analysis aimed at an understanding of substrate specificity. These studies have focused largely on the origins of specificity at the primary S1 site. At this position, trypsin hydrolyzes amide substrates containing P1-Lys and P1-Arg amino acids by factors of 10^5 or greater relative to the next-preferred residues (Graf et al., 1988; Evnin et al., 1990). The preference of the enzyme is 2–10-fold in favor of Arg- relative to Lys-containing substrates (Craik et al., 1985; Perona et al., 1993c). As might be expected from their structural disparity, Lys and Arg interact in a differential manner with the primary determinants Asp 189 and Ser 190 (Ruhlmann et al., 1973; Bode et al., 1984; Fig. 10). The guanidinium group of P1-Arg substrates makes an ion-pair interaction with Asp 189, whereas the interaction of P1-Lys is solely by a water-mediated contact. Both Arg and Lys substrate side chains also interact with Ser 190.

An early study assessed the precision with which the S1 site is constructed by introducing small perturbations: the Gly residues at positions 216 and 226 were converted to Ala, resulting in the three trypsin mutants G216A, G226A and G216A/G226A (Craik et al., 1985; Fig. 10). Relative specificities for tripeptide amide P1-Arg/Lys substrates, as assessed by the ratio of k_{cat}/K_m values, were altered by up to 20-fold. Catalytic efficiencies were decreased by 40-fold to 10^4-fold, and these effects involved significant decreases in k_{cat} as well as higher K_m values. The differential effects of the k_{cat} and K_m values resulted in enzymes that were more Arg specific (G216A) and more Lys specific (G226A) than the wild-type enzyme. Subsequent crystal structure determinations of trypsins G226A (Wilke et al., 1991) and G216A (M.E. McGrath & R.J. Fletterick, unpubl. results)

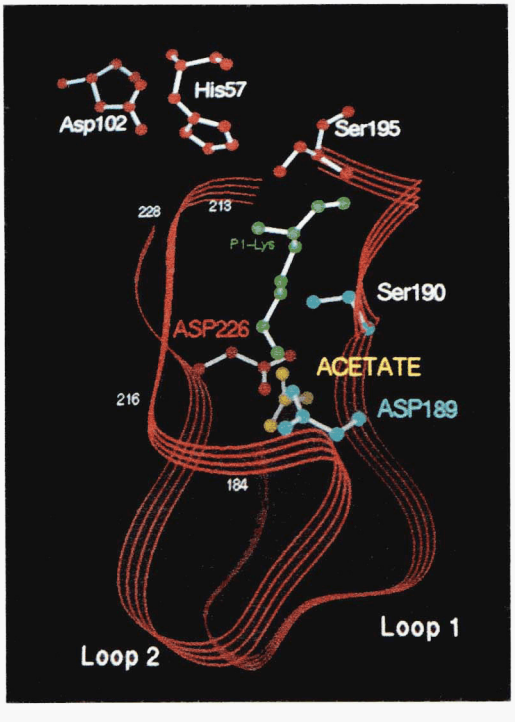


Fig. 10. Role of the position of the negative charge at the base of the trypsin S1 site has been probed by random and site-directed mutagenesis coupled to crystal structure analysis of variants. Shown is the structure of the S1 binding pocket of trypsin, indicating the positions at which the negatively charged amino acid has been determined by X-ray crystal structures. Blue, wild-type trypsin at position 189; red, trypsin D189G/G226D at position 226; yellow, exogenously added acetate ion in trypsin D189S (acetate reconstitutes activity toward P1-Arg and P1-Lys-containing substrates). Wild-type amino acids at positions 216 and 226 are each Gly, permitting access of the large P1-Lys (green) and P1-Arg side chains to Asp 189.

complexed with benzamidine showed that the alanine substitutions produced no structural perturbations beyond the immediate vicinity of the mutated residues. Because the catalytic triad Ser 195, His 57, and Asp 102 amino acids are unaffected by these binding pocket alterations, it is highly probable that the decreases in k_{cat} are attributable to altering the catalytic register of the scissile bond. These data thus provided an early demonstration that substrate binding and catalytic turnover are interrelated functions in trypsin, and that they can be affected differentially to alter the function of the enzyme.

A series of studies have addressed the role of the negatively charged Asp 189 residue in binding and catalysis. These investigations have made use of both site-directed mutagenesis as well as a genetic selection approach for the isolation of new variants (Fig. 4B). The selection is based on expression of a library of trypsin variants into the periplasmic space of an *E. coli* strain that is auxotrophic for arginine or lysine (Evnin et al., 1990). Cells are plated on minimal media containing a nonnutritive substrate analog of one of these amino acids; active trypsins cleave the analog, liberating free amino acid and thereby relieving the auxotrophy (Evnin et al., 1990; Perona et al., 1993a).

Twenty variant trypsins have been isolated from a library of 400 possible mutants encompassing the amino acids at positions 189 and 190 at the base of the S1 site. Kinetic characterization of these enzymes, as well as of the variants D189K (Graf et al., 1987) and D189S (Graf et al., 1988), indicates that the presence of a negative charge at the base of the binding pocket is essential to high-level catalysis by trypsin. Variants lacking the negative charge are compromised in k_{cat}/K_m toward peptidyl Arg- or Lys-containing amide substrates by a factor of 10^5 or greater. Activity toward these substrates is partially restored by the presence of an Asp or Glu residue at positions 189 or 190. The variants span a range of catalytic efficiencies ranging from wild type to decreases of 10^6-fold (Evnin et al., 1990; Perona et al., 1993a).

A framework for the interpretation of these data is provided by kinetic and crystallographic investigation of two other variants: trypsins D189G/G226D (Perona et al., 1993b, 1993c) and D189S (Perona et al., 1994). The structure of each mutant enzyme was determined complexed with the protein inhibitors APPI and/or BPTI, which are analogs of the substrate Michaelis complexes possessing Arg and Lys, respectively, at the P1 position (Perona et al., 1993b). This allows for the direct comparison of substrate-like interactions of Arg and Lys side chains in the binding pockets of wild-type and mutant enzymes. Trypsin D189G/G226D is equally reduced (10-fold) in binding affinity toward Lys and Arg substrates and is sharply lowered (10^3-fold) in k_{cat} toward Arg. The crystallographic analysis showed that Asp 226 is partially sequestered from substrate by intramolecular interactions made with Ser 190 and Tyr 228, such that only a single carboxylate oxygen is available for substrate binding. Further, comparisons with the wild-type interactions indicated no correlation between the binding affinities of either Lys and Arg substrates and the number of direct contacts made with Asp 226. Therefore, it appears that substrate binding affinity to trypsin depends upon the accessibility of the negative charge to substrate and not upon the formation of direct interactions. This observation implies that direct electrostatic hydrogen bonding interactions between the substrate Lys/Arg and the enzyme carboxylate group do not significantly improve the free energy of binding relative to indirect water-mediated interactions (Perona et al., 1993c).

The crystal structure of trypsin D189S revealed that an acetate ion from the crystallization buffer was trapped at the base of the binding pocket, such that its carboxylate group was partially oriented toward substrate (Perona et al., 1994; Fig. 10). Exogenously added acetate provided up to 300-fold rate enhancements to trypsin D189S toward Arg- and Lys-containing substrates, but catalytic activity remained diminished relative to wild-type trypsin. This structure thus provides a second example showing that optimal placement of the negative charge in the binding pocket is critical to catalysis. Significantly, the diminished activities of both trypsins D189G/G226D and D189S/acetate are reflected in k_{cat} as well as K_m. Measurement of activities toward analogous ester as well as amide substrates by these enzymes allows calculation of the mechanistic parameters K_s, k_2, and k_3 (Zerner & Bender, 1964; Fig. 2C), removing the ambiguity in interpretation of the steady-state Michaelis–Menten parameters. This analysis shows that the role of the Asp 189 carboxylate in trypsin is twofold: it provides both tight binding affinity K_s as well as high acylation rate k_2 (Perona et al., 1994). Therefore, the precise location of the negatively charged group within the trypsin S1 site is critical to positioning the scissile bond in catalytic register with Ser 195 and His 57.

Analysis of the kinetic properties of the 20 variants isolated from the genetic selection corroborates these hypotheses regarding the operation of the S1 site. Although the binding constants of the enzymes vary widely, it is significant that relative affinities for Lys versus Arg substrates remain very similar (Perona et al., 1993a). The negatively charged carboxylate in these mutants is provided by either Asp or Glu at positions 189 or 190, and the partner to this residue is 1 of 10 different amino acids. Thus, it is very unlikely that equal reductions in affinity toward Lys versus Arg substrates can in most cases be attributed to an equal loss of hydrogen bonding or electrostatic interactions. Instead, binding affinity is likely to be better correlated with accessibility of the negative charge to substrate; barring substrate-induced conformational changes, this accessibility will be the same for both Lys and Arg substrates. Binding affinities are then predicted to be weaker when the carboxylate is partially sequestered from substrates, as seen in the structures of the mutants D189G/G226D and D189S/acetate. Crystal structures of additional variants from the selection pool should enable a quantitative correlation between binding affinity and accessibility of the negative charge. These experiments also explain the rationale for conservation of the Asp at position 189 in the vast majority of trypsin homologs, because other locations result in partial sequestration of the negative charge.

In a second set of experiments, site-directed mutagenesis has been used to convert trypsin into a chymotrypsin-like protease possessing high selectivity for cleavage adjacent to large hydrophobic amino acids (Hedstrom et al., 1992, 1994a, 1994b). The structures of the S1 pockets of the two enzymes are very similar (Figs. 6, 11A), so it was expected that specificity modification might be straightforward as in subtilisin and α-lytic protease. However, when the amino acids directly in contact with substrate were exchanged into trypsin, the resulting variants D189S and D189S/Q192M/I138T/T218 failed to exhibit significant improvement in cleavage of P1-Phe amide substrates (Graf et al., 1988; Hedstrom et al., 1992; Table 3). Poor efficiency was also shown toward trypsin substrates, as expected because the pocket lacks a negative charge. The crystal structure of trypsin D189S showed that only very local structural changes were introduced

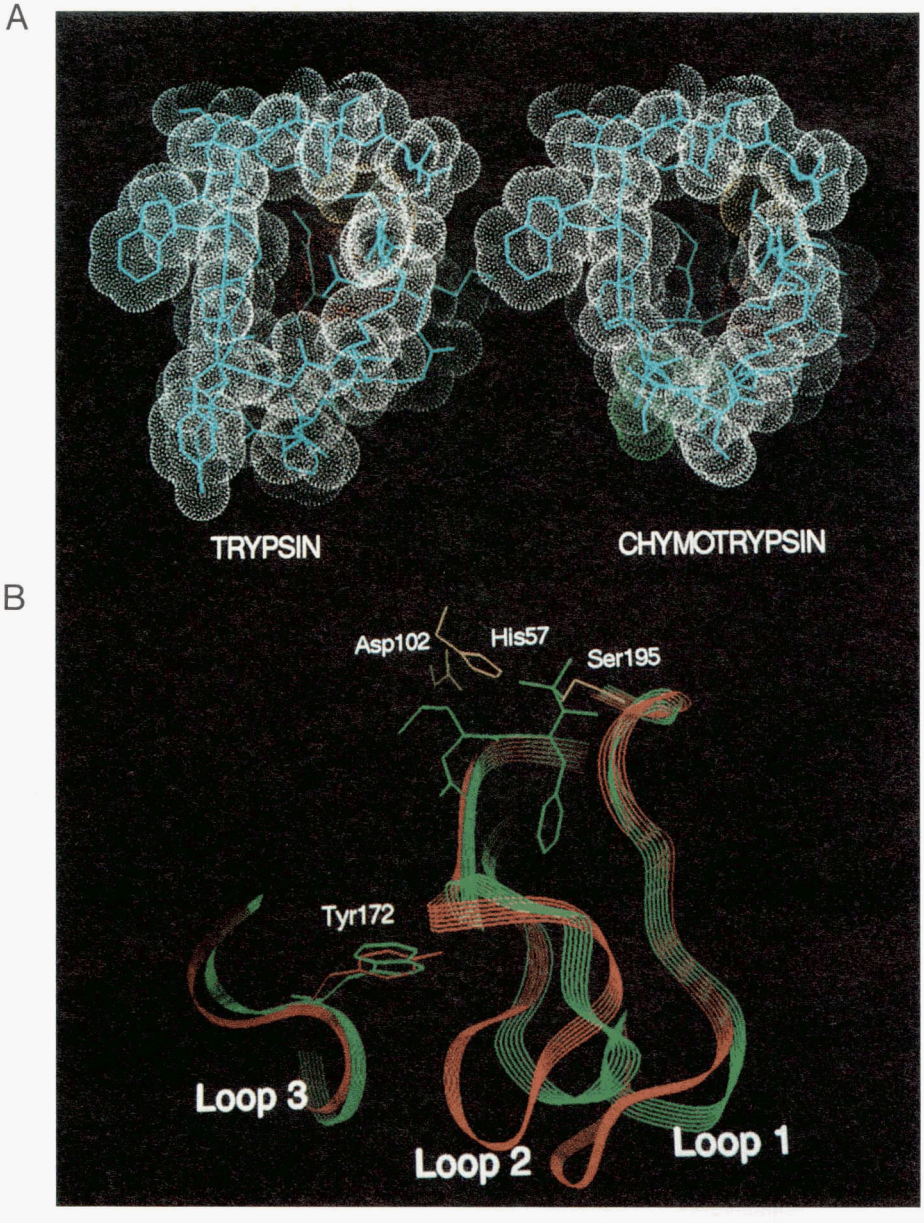


Fig. 11. A: Comparison of the S1 sites of trypsin and chymotrypsin. Van der Waals surfaces of each enzyme are shown with the position-189 amino acid (Asp in trypsin; Ser in chymotrypsin) indicated in red. In yellow is the conserved Ser 190, which is oriented into the S1 pocket in trypsin but rotates out in chymotrypsin. The inserted Thr 218 in chymotrypsin is shown in green. Two other amino acids directly in or adjacent to the S1 site are Ile 138 (Thr 138 in chymotrypsin), and Gln 192 (Met 192 in chymotrypsin). Although a high degree of structural similarity is clear, exchange of these four amino acids fails to transfer chymotryptic specificity to trypsin. **B:** Structural determinants required to exchange substrate specificity include two adjacent surface loops (loop 1 and loop 2) and an amino acid (Tyr 172 in trypsin) in a third adjacent segment (loop 3). None of these structural elements directly contact substrate (shown at top in thin green lines). Trypsin is shown in red and chymotrypsin in green.

as a consequence of the substitution; the binding pocket maintains a trypsin-like conformation (Perona et al., 1994). This confirms that the small structural differences between trypsin and chymotrypsin in the S1 site (Fig. 11A) must be critical determinants of the specificity and must rely on more distant parts of the structure for maintenance of their particular conformations.

Exchange of the two surface loops, loop 1 and loop 2 (Fig. 11B), resulted in the hybrid enzyme Tr→Ch[S1+L1+L2], which exhibited an acylation rate constant k_2 equal to that of chymotrypsin toward peptidyl P1-Phe amide substrates (Hedstrom et al., 1992; Table 3). However, the enzyme was still reduced by nearly 10^3-fold in k_{cat}/K_m because of a very weak substrate binding affinity. The mechanistic kinetic parameters K_s, k_2, and k_3 were calculated for cleavage of both single-residue and peptidyl P1-Phe amide substrates for the enzymes trypsin, chymotrypsin, D189S and Tr→Ch[S1+L1+L2]. These data showed that, like chymotrypsin, the hybrid trypsin was able to use the binding energy obtained by occupancy of the S2–S4 enzyme sites to increase the acylation rate. They also demonstrated that, among this series of enzymes, the key mechanistic step that determines substrate specificity is not binding affinity, but instead the chemical step of acylation (Hedstrom et al., 1992, 1994a).

Further mutations were sought to improve catalytic efficiency toward chymotryptic substrates by increasing binding affinity. The additional mutation Y172W in a third adjacent surface loop (Fig. 11B) produced the hybrid enzyme Tr→Ch[S1+L1+L2+Y172W], which improves the activity of Tr→Ch[S1+L1+L2] by 20–50-fold, creating an enzyme with up to 15% of the activity of chymotrypsin (Hedstrom et al., 1994b; Table 3). The improvement toward a tetrapeptide P1-Phe amide substrate is manifested almost entirely in tighter binding affinity. The relative catalytic efficiencies measured toward Trp, Tyr, Phe, and Leu P1-amide substrates also more closely mimic chymotrypsin (Hedstrom et al., 1994b).

Table 3. *Conversion of trypsin to chymotryptic specificity*[a]

	K_s (M)	k_2 (s^{-1})	k_3 (s^{-1})
Trypsin	>0.25	>0.2	36
D189S	0.015	0.29	33
Tr→Ch[S1+L1+L2]	0.011	20	37
Tr→Ch[S1+L1+L2+Y172W]	5.0×10^{-4}	41	63
Chymotrypsin	1.5×10^{-3}	850	52

[a] Substrate: *suc*-Ala-Ala-Pro-Phe-*p*NA.

Fig. 12. A correlation is observed between the backbone conformation of residue 216 and the S1 site substrate preference among all of the trypsin-, chymotrypsin-, and elastase-like proteases of known structure. Shown is a superposition of seven mammalian serine proteases (color-coded), indicating the structure at this position that is most easily visualized in the orientation of the carbonyl oxygen atom. Specific trypsin-like, chymotrypsin-like, and elastase-like ϕ–ψ backbone angles are observed. Residue 216 binds the P3 position of the substrate in all the enzymes. Extended peptide binding to residue 216 is required both to acheive full catalytic potency as well as to obtain a maximal level of P1-site discrimination among alternative amino acids. Conversion of the substrate specificity of trypsin to that of chymotrypsin requires reorientation of Gly 216 to a chymotrypsin-like conformation. Thus, the position-216 backbone is strongly suggested as an essential specificity determinant in the mammalian trypsin-like proteases.

The structural basis for the activities of the two hybrid trypsins was elucidated by determination of their crystal structures complexed with the transition-state inactivator *suc*-Ala-Ala-Pro-Phe-chloromethyl ketone (*suc*-AAPF-CMK; Perona et al., 1995). Loop 2 of Tr→Ch[S1+L1+L2] adopts a conformation identical to that which it possesses in chymotrypsin. However, amino acids at positions 185–187 within Loop 1 are disordered. The structure of Tr→Ch[S1+L1+L2+Y172W] showed improved order in Loop 1 and a rearrangement of solvent structure and Ser 217 side-chain orientation, each of which more closely mimicked the structure of chymotrypsin. No other changes were present between the two hybrid enzymes, implicating these structural elements as important determinants of K_s in chymotrypsin.

Both hybrid enzymes possess wild-type chymotrypsin-like acylation rates k_2 toward peptidyl P1-Phe amide substrates, and each utilizes binding of the extended peptide (substrate sites P2–P4) to increase this rate. In fact, the 10^6-fold specificity of chymotrypsin relative to trypsin for cleavage at P1-Phe is manifested solely in extended peptidyl substrates; only a 10^2-fold level of discrimination exists for single-residue substrates (Hedstrom et al., 1994b). In all available crystal structures of the enzymes, including those of the trypsin hybrids, two hydrogen bonds are formed in an antiparallel β-sheet fashion with the backbone amide group of Gly 216 (Perona et al., 1995). The backbone conformation at Gly 216 differs between trypsin and chymotrypsin; the hybrid enzymes adopt a chymotrypsin-like conformation. This suggests that the Gly 216 backbone is a critical specificity determinant because it directly binds a portion of substrate responsible for a 10^4-fold preference at position P1. The mechanism by which Gly 216 functions is likely to be through promoting accurate scissile bond positioning (Perona et al., 1995). Because Asp 189 of trypsin also plays a critical role in this function, it appears that the identity of the amino acid at position 189, and the backbone conformation at Gly 216, must be matched in order to permit efficient and specific catalysis by trypsin and chymotrypsin.

Structural comparisons among a number of the chymotrypsin-like proteases, including both PPE and HNE, showed a striking correlation between the P1-site specificity and the backbone conformation at position 216 (Perona et al., 1995). Three structural classes were delineated, which correspond to trypsin, chymotrypsin, and elastase-like enzymes (Fig. 12). The role of Gly 216 in promoting accurate substrate positioning may thus be a feature of many enzymes in the family. In this context it is relevant to note that the kinetic phenomenon observed for both trypsin (Perona et al., 1993c) and chymotrypsin (Hedstrom et al., 1992), namely that subsite occupancy causes large increases in the rates of the chemical steps of catalysis, is also common to other trypsin-like enzymes including PPE (Thompson & Blout, 1970), HNE (Stein et al., 1987), SGPA (Bauer et al., 1976; Bauer, 1978), SGPB (Bauer, 1978), and α-lytic protease (Bauer et al., 1981; also see above). The significance of the recent kinetic analysis (Hedstrom et al., 1992) is that it shows that both the catalytic rate toward cognate substrates, as well as the degree of specificity at the P1-position, are dependent on the filling of subsites, which themselves exhibit little amino acid preference.

The crystal structures of the trypsin hybrids also address another fundamental question in enzyme catalysis: the role of the global protein structure. Distal structural elements such as Trp 172 and loops 1 and 2 play a key role in specifying the conformation of residues that do interact directly with substrate. Thus, their role is not solely to provide an inert platform that stabilizes the amino acids that interact directly with substrate. These elements of the global architecture play an active role in determining substrate specificity as well, which should thus be viewed as a more distributed property of the protein fold. An alternative mechanism for the way in which global protein folds may influence specificity is by modulating the degree of backbone flexibility of the S1 site, as exemplified in the α-lytic protease studies (Bone et al., 1991).

Exchange of the S1-site residues of HNE into trypsin also fails to convert the specificity of trypsin and results, as in the case of the mutants D189S and D189S/Q192M/I138T/T218, in a poor nonspecific protease (J.J. Perona & C.S. Craik, unpubl.

obs.). Similarly, introduction of Lys, Arg, or His residues into the trypsin S1 site has failed to generate specificity toward Asp or Glu residues (Graf et al., 1987; Willett et al., 1995; J.J. Perona & C.S. Craik, unpubl. obs.). A better mutational strategy for specificity modification in trypsin may be the construction of libraries that instead span the distal structural elements. When coupled to strategies such as the genetic selection (Evnin et al., 1990; Perona et al., 1993a) or phage display (Corey et al., 1993; Fig. 4C) systems, it should be possible to search a large number of different structures for those providing altered specificity.

Surface loops determine subsite specificity in the trypsin-class enzymes

We have seen that the best-studied members of the chymotrypsin-like class of serine proteases each manifest primary specificity at the P1 site directly adjacent to the cleaved bond. However, there are also several enzymes in the class that possess significant specificity toward substrate residues at a greater distance in both the N- and C-terminal directions. Sequence alignments of these enzymes reveal that a number of surface loops flanking the catalytic residues are very likely to play crucial roles in determining this extended recognition selectivity (Fig. 13).

One enzyme manifesting an extended subsite specificity that is also of known tertiary structure is RMCPII (Woodbury et al., 1978a, 1978b), a member of a homologous subclass of trypsin-like serine proteases expressed also in other granulocytes (Salvesen et al., 1987) as well as in lymphocytes (Lobe et al., 1986). RMCPII and the related RMCPI (which possess 73% amino acid sequence identity; LeTrong et al., 1987b) each manifest a chymotrypsin-like primary substrate specificity but also exhibit preferences for hydrophobic amino acids in positions P2 and P3 (Yoshida et al., 1980; Powers et al., 1985). RMCPI also has been shown to prefer hydrophobic residues at position P1′ in polypeptide substrates, although the extent of the selectivity has not been established quantitatively (LeTrong et al., 1987a).

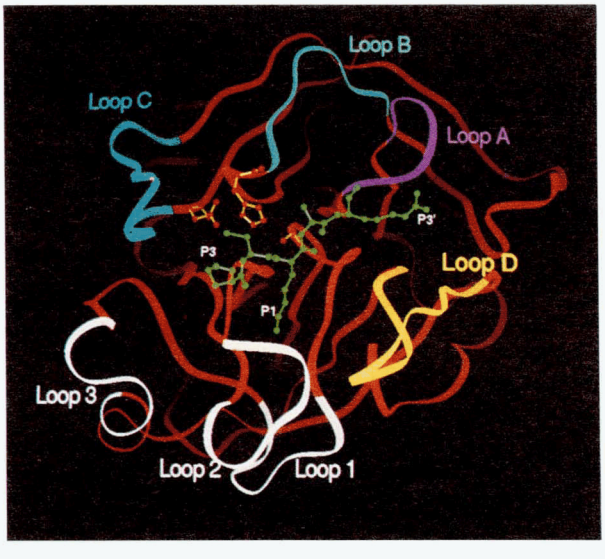


Fig. 13. Structure of trypsin, highlighting the positions of four surface loops (loops A, B, C, D) involved in determining subsite preferences among a number of the enzymes in the family. The location of these loops relative to the catalytic machinery and binding cleft may be contrasted with the position of the three loops (loops 1, 2, 3) that combine to influence specificity in the S1 site. A polypeptide substrate chain is shown in green and the catalytic triad is in yellow. It is clear that loop C is positioned to interact with substrate residues N-terminal to the scissile bond, whereas loops A and D are positioned to interact with the C-terminal amino acids on the leaving-group side of the scissile bond.

The crystal structure of uncomplexed RMCPII has been determined at a resolution of 1.9 Å (Remington et al., 1988). This structure suggests that the enhanced substrate selectivity of the homologous RMCPI at the P1′ position is likely to be provided by the presence of a large cleft not found in the other chymotrypsin-like proteases of known structure. The cleft is formed as a consequence of an unusual conformation adopted by two surface loops that lie adjacent to the catalytic residues (Remington et al., 1988). The loops comprise residues 34–41 (loop A) and 59–64 (loop B) and are positioned such as to be capable of interacting directly with substrate residues C-terminal to the scissile bond (Fig. 13). Modeling of a substrate complex with RMCPII suggests that loop A is most likely to directly contact the P1′-P2′ substrate sites, whereas loop B plays a structural role in helping to form the cleft.

The subclass of serine proteases to which RMCPII belongs is distinguished by the absence of the otherwise well-conserved disulfide bond linking residues 191 and 220 (LeTrong et al., 1987b). In the other enzymes, this disulfide bridges the two walls of the S1 site and likely provides a degree of structural rigidity to the cavity (Fig. 7). RMCPII possesses a Phe residue at position 191 and a shortened loop L2 (residues 217–225) relative to chymotrypsin; each of these features is conserved within the subclass (LeTrong et al., 1987b). Modeling of a tripeptide substrate possessing Phe at position P3 shows that the aromatic ring is readily sandwiched between the side chains of Met 192 and Pro 221A and also makes van der Waals interactions with Phe 191 (Remington et al., 1988). This small hydrophobic pocket is absent in chymotrypsin owing to the presence of the Cys 191–Cys 220 disulfide bond. Thus, the crystal structure provides a plausible rationale explaining the 100-fold preference of RMCPI and RMCPII for Phe relative to Gly at position P3 (Yoshida et al., 1980).

A second example of extended binding site specificity is provided by the enzyme enteropeptidase (enterokinase), which functions in vivo to cleave the zymogen trypsinogen at position Ile 16, generating the new N-terminus required for trypsin activity (reviewed in Huber & Bode, 1978). This enzyme hydrolyzes the peptide bond directly C-terminal to the sequence $(Asp)_4Lys$ in trypsinogen, and consequently possesses a trypsin-like specificity toward positively charged amino acids in the P1 position. The bovine and porcine enzymes exist as glycosylated disulfide-linked heterodimers comprising a heavy chain of 115 kDa and a light chain of 43 kDa (Magee et al., 1977; LaVallie et al., 1993). Chemical modification studies established that the catalytic activity and specificity of the enzyme resides in the light chain (Light & Fonseca, 1984). Most recently, cloning and expression of the light chain has revealed it to possess 35–40% sequence identity to the trypsin-like class of serine proteases (LaVallie et al., 1993). This study also demonstrated that this subunit possesses full activity toward the fluorogenic peptide substrate $(Asp)_4Lys$-β-naphthylamide. The presence of the heavy chain, however, endows the holoenzyme with 100-fold greater catalytic efficiency toward the cognate trypsinogen substrate (LaVallie et al., 1993).

Native enteropeptidase is capable of cleaving the $(Asp)_4Lys$ sequence in trypsinogen with a catalytic efficiency roughly 10^4-fold greater than trypsin (Maroux et al., 1971). Mapping the sequence of the light chain of the enzyme onto the structure of trypsin indicates that the peptide Lys 96–Arg 97–Arg 98–Lys 99 (KRRK) is well positioned to play a direct role in interacting with the negatively charged aspartates occupying positions P2–P5 (LaVallie et al., 1993). This peptide comprises a portion of a surface loop located adjacent to Asp 102 (loop C; Fig. 13), on the opposing side of the catalytic triad relative to the loops A and B that form the cleft important to P1′ recognition by RMCPI.

The kinetic basis for the improved specificity of enteropeptidase relative to trypsin for recognition of the $(Asp)_4Lys$ sequence is not yet known. By analogy with the known operation of the pancreatic proteases, it would be predicted that the specificity arises at least partly from the ability of enteropeptidase to selectively accelerate the acylation rate of $(Asp)_4Lys$-β-naphthylamide relative to other peptidyl or to single-residue substrates. It is tempting to speculate that enteropeptidase may use a distinct structural mechanism, involving specific interactions with the aspartates, to convert substrate binding energy into a high catalytic rate. Inspection of the sequence alignment with trypsin reveals further differences at positions 215–219 at the lip of the S1 site, as well as the insertion of a residue in loop L3 (Fig. 13), each of which may be of importance to precise orientation of the $(Asp)_4Lys$ substrate. Additionally, enteropeptidase possesses a striking 10-residue insertion between residues 58 and 59, in the surface loop B that lies directly behind the KRRK sequence of loop C (LaVallie et al., 1993; Fig. 13). Although loops B and C do not contact each other in trypsin, the much larger loop B in enteropeptidase would be capable of making interactions conceivably of importance to maintaining correct orientation of the KRRK residues.

A third example of the importance of surface loops in these enzymes relates to the inhibition of the trypsin-like tissue plasminogen activator by plasminogen activator inhibitor I (Ny et al., 1986). The interaction between TPA and PAI-1 is of importance in the regulation of the cascade of activities involved in blood clotting (Davie et al., 1991). Surface loop A of TPA (Fig. 13) possesses a high density of positively charged amino acids (residues Lys 296–His 297–Arg 298–Arg 299) that have been shown to be critical to its interaction with a negatively charged region of PAI-1 (Madison et al., 1990). This was confirmed in an elegant experiment in which loop A in the homologous enzyme thrombin was replaced with that of TPA, endowing PAI-1 susceptibility onto thrombin (Horrevoets et al., 1993). Thus, both the extended substrate specificity as well as the specificity of interaction with physiologically important inhibitors can arise from contacts with the same surface loops.

An important activity of crab collagenase is the ability to cleave native triple-helical collagen, a property not exhibited by the canonical pancreatic proteases (Eisen et al., 1973; Tsu et al., 1994). Cleavage occurs within domains of the triple-helical substrate that are relaxed from the strict Gly-Pro-Xaa repetitive sequence. Detailed examination of the cleavage sites by protein sequencing has shown that proteolysis of collagen occurs at positions that mirror the P1-site selectivity (Tsu et al., 1994). Sequence alignments of a range of serine collagenases from diverse species fails to elucidate clear amino acid similarities that might be correlated to the triple-helical specificity (Sinha et al., 1987; Sellos & Van Wormhoudt, 1992). However, the crystal structure of collagenase complexed with the dimeric protein inhibitor ecotin has allowed construction of a model of collagen interacting with the enzyme (J.J. Perona, C.A. Tsu, R.J. Fletterick, & C.S. Craik, in prep.). Several surface loops, including loops A and D (Fig. 13), may play crucial roles in recognition of the triple helix.

Recently, a novel assay has been introduced that provides the possibility of assaying relative preferences at positions on the leaving-group side of the scissile bond (Schellenberger et al., 1993). In an initial study, the S1′ subsite specificities of trypsin and chymotrypsin from cow and rat were determined by monitoring the reverse reaction of peptide hydrolysis. Acyl transfer was measured to a mixture of 21 peptide nucleophiles of the general structure H-Xaa-Ala-Ala-Ala-Ala-NH_2; the *decrease* in concentration of each nucleophile was monitored by HPLC and represents a measure of the ability of that substrate to compete with water for attack on the acyl enzyme. Chymotrypsin hydrolyzes substrates possessing Arg and Lys at the substrate P1′ position roughly 10-fold more rapidly than does trypsin; this selectivity is attributable to the presence of additional negatively charged residues in two adjacent surface loops (see below). Trypsin exhibits a slight preference for hydrophobic amino acids at this position, relative to chymotrypsin. The data confirm the relative lack of specificity of each enzyme at this position. Application of the methodology to crab collagenase showed a 30-fold preference for P1′-Arg residues; an Arg is also found on the C-terminal side of several of the collagen cleavage sites of the enzyme (Tsu et al., 1994). Data have also been obtained for specificities at the subsites S1′–S3′ for trypsin, chymotrypsin, α-lytic protease, and the cercarial protease from *Schistosoma mansoni*; in these cases, relative cleavage rates varied by factors of up to 10^2-fold (Schellenberger et al., 1994).

It is clear from the many known structures of chymotrypsin-like serine proteases that loop C is invariably positioned to directly contact the extended substrate on the N-terminal side, whereas loops A and D interact on the leaving group side. By contrast, loop B appears less likely to be involved in direct contacts but instead is positioned to stabilize the primary interactions made by the more forward loops (Fig. 13). Depending on the size and conformation of this loop in different enzymes, it might in principle be able to stabilize either loop A or C. A final example of specificity modification in this class involves loop D: introduction of histidine residues at the N- and C-terminal ends of this loop confers metal-dependent specificity for histidine at the P2′ substrate position onto rat trypsin (Willett et al., 1995). In general, because subsite specificity of chymotrypsin-like proteases is modulated by surface loops rather than by core secondary structure elements, the prospects for engineering novel specificities, and for the development of "restriction proteases" that might recognize substrate sites from P5 to P2′, seem hopeful.

Conclusions and future directions

One of the questions addressed in these studies is the role of water molecules in mediating enzyme–ligand interactions. Crystal structures of wild-type and variant enzymes complexed with substrate analogs, together with the measurement of affinity constants, allows deduction of the importance of particular interactions. In the recognition of basic Lys and Arg substrate side chains by Asp 189 of trypsin, the conclusion is that a water-mediated interaction can provide a comparable free energy gain to a direct contact (Perona et al., 1993c). These studies have im-

plications to understanding protein-nucleic acid interactions. For example, the crystal structures of the *trp* repressor-operator complex, and of the uncomplexed operator DNA, suggest a crucial role for water-mediated interactions in providing DNA sequence specificity because no direct contacts with base functional groups are observed (Otwinowski et al., 1988; Shakked et al., 1994). Although a second-site reversion analysis of the operator DNA further implied a key role for the intervening waters, it was clear that a structural analysis of the modified complexes is still required (Joachimiak et al., 1994). Such an analysis of the charge-charge interactions in the trypsin S1 site shows more definitively that a specificity-determining role for solvent is in principle possible. A similar study of the *trp* repressor and of other systems is warranted, to address the extent to which this phenomenon may be dependent on the local structural context.

Another fundamental question concerns the design of enzyme structures to provide different degrees of flexibility to the substrate binding site. The comparison of trypsin and α-lytic protease offers an excellent opportunity to address this issue. Thus far, it appears from both kinetic and structural analysis of mutants that the trypsin pocket may be considerably more rigid. However, the two structures are homologous so that the degree of difference in the surrounding scaffolds is relatively small. Thus, the problem may be manageable: which specific interactions bridging the primary and secondary shell residues are most critical for determining flexibility? Are residues located even more distant also important? An excellent test of our understanding would be the construction of a trypsin variant with chymotryptic specificity, which possessed far fewer than the 15 alterations of Tr→Ch[S1+L1+L2+Y172W]. If indeed the conformation of Gly 216 is crucial to P1-site specificity, then the problem reduces to adding certain key mutations to D189S such that Gly 216 can reorient upon substrate binding, as it is observed to do in α-lytic protease (Bone et al., 1991; Fig. 9). A deeper understanding of flexibility would have clear application to protein folding and stability as well (Rose & Creamer, 1994).

The degree to which a substrate binding cleft is inherently deformable may be an important parameter governing the ease with which specificity modification can be effected. Prior to the advent of site-directed mutagenesis, it appeared possible that even conservative amino acid changes might cause highly deleterious long-range structural effects. We now know that most substitutions are absorbed locally and that the majority of protein structural contexts therefore have some ability to deform. Protein folding and stability often are not greatly perturbed even by very challenging mutations. The sensitivity of enzyme activity to precise substrate positioning might alternatively suggest that mutation of the binding site would usually result in low catalytic activity. However, this appears not to be the case: among the well-studied binding pockets considered here, the subtilisin S1 and S4 sites, as well as the α-lytic protease S1 site, each are readily modified to alter specificity with only limited local substitutions. Only the trypsin S1 site requires extensive nonlocal changes.

Another reason for the difficulty in modifying trypsin substrate specificity could be that the charge-charge interactions present in a trypsin transition-state complex require a precise electrostatic environment not readily altered (Hwang & Warshel, 1988). The electrostatic potential is presently the least understood force shaping enzyme structure and activity; it is also the only one that operates over large distances. Considerable efforts are underway to improve empirical forcefields, so that catalytic free energies can be accurately estimated directly from structural models. Serine proteases are a favored system in these studies owing to the large database of structure-activity information (Bash et al., 1987; Rao et al., 1987; Caldwell et al., 1991; Mizushima et al., 1991; Wilson et al., 1991). Further mutational analysis will thus also be invaluable in providing a testbed for new algorithms. Greater insight into the connection between structure and energetics will lead to much better predictive ability regarding the consequences of mutation. This improved insight, together with the innovative technologies for the generation and screening of large libraries, may soon result in the creation of new, highly efficient proteases possessing a broad range of useful properties.

Acknowledgments

We thank Prof. R. Fletterick for helpful discussions. This work was supported by NSF grants MCB-9219806 and BCS-9119237 to C.S.C. and NIH postdoctoral NRSA award GM 13818-03 to J.J.P.

References

Abrahmsen L, Tom J, Burnier J, Butcher KA, Kossiakoff A, Wells JA. 1991. Engineering subtilisin and its substrates for efficient ligation of peptide bonds in aqueous solution. *Biochemistry 30*:4151-4159.

Bachovchin WW, Roberts JD. 1978. Nitrogen-15 nuclear magnetic resonance spectroscopy. The state of histidine in the catalytic triad of α-lytic protease. Implications for the charge-relay mechanism of peptide-bond cleavage by serine proteases. *J Am Chem Soc 100*:8041-8047.

Bash PA, Singh UC, Langridge R, Kollman PA. 1987. Free energy calculations by computer simulation. *Science 236*:564-566.

Bauer CA. 1978. Active centers of *Streptomyces griseus* protease 1, *Streptomyces griseus* protease 3, and α-chymotrypsin: Enzyme-substrate interactions. *Biochemistry 17*:375-380.

Bauer CA, Brayer GD, Sielecki AR, James MNG. 1981. Active site of α-lytic protease. Enzyme-substrate interactions. *Eur J Biochem 120*:289-294.

Bauer CA, Thompson RC, Blout ER. 1976. The active centers of *Streptomyces griseus* protease 3 and α-chymotrypsin: Enzyme-substrate interactions remote from the scissile bond. *Biochemistry 15*:1291-1295.

Bazan JF, Fletterick RJ. 1990. Structural and catalytic models of trypsin-like viral proteases. *Semin Virol 1*:311-322.

Bech LM, Sørensen SB, Breddam K. 1992. Mutational replacements in subtilisin 309. Val 104 has a modulating effect on the P4 substrate preference. *Eur J Biochem 209*:869-874.

Bech LM, Sørensen SB, Breddam K. 1993. Significance of hydrophobic S4-P4 interactions in subtilisin 309 from *Bacillus lentus*. *Biochemistry 32*: 2845-2852.

Bender ML, Killheffer JV. 1973. Chymotrypsins. *CRC Crit Rev Biochem 1*:149-199.

Betzel C, Klupsch S, Papendorf G, Hastrup S, Branner S, Wilson KS. 1992. Crystal structure of the alkaline proteinase savinase from *Bacillus lentus* at 1.4 Å resolution. *J Mol Biol 223*:427-445.

Betzel C, Pal GP, Saenger W. 1988. Three-dimensional structure of proteinase K at 0.15-nm resolution. *Eur J Biochem 178*:155-171.

Betzel C, Singh TP, Visanji M, Fittkau S, Saenger W, Wilson KS. 1993. Structure of the complex of proteinase K with a substrate analogue hexapeptide inhibitor at 2.2-Å resolution. *J Biol Chem 268*:15854-15858.

Blow DM. 1976. Structure and mechanism of chymotrypsin. *Acc Chem Res 9*:145-152.

Blow DM, Birktoft JJ, Hartley BS. 1969. Role of a buried acid group in the mechanism of action of chymotrypsin. *Nature 221*:337-340.

Bode W, Chen Z, Bartels K, Kutzbach C, Schmidt-Kastner G, Bartunik H. 1983. Refined 2 Å X-ray crystal structure of porcine pancreatic kallikrein A, a specific trypsin-like serine proteinase—Crystallization, structure determination, crystallographic refinement, structure and its comparison with bovine trypsin. *J Mol Biol 164*:237-282.

Bode W, Mayr I, Baumann U, Huber R, Stone SR, Hofsteenge J. 1989a. The refined 1.9 Å crystal structure of human α-thrombin: Interaction

with D-Phe-Pro-Arg chloromethyl ketone and significance of the Tyr-Pro-Pro-Trp insertion segment. *EMBO J 8*:3467–3475.

Bode W, Meyer E Jr, Powers JC. 1989b. Human leukocyte and porcine pancreatic elastase: X-ray crystal structures, mechanism, substrate specificity, and mechanism-based inhibitors. *Biochemistry 28*:1951–1963.

Bode W, Papamokos E, Musil D, Seemuller U, Fritz H. 1986a. Refined 1.2 Å crystal structure of the complex formed between subtilisin Carlsberg and the inhibitor eglin C. Molecular structure of eglin and its detailed interaction with subtilisin. *EMBO J 5*:813–818.

Bode W, Walter J, Huber R, Wenzel HR, Tschesche H. 1984. The refined 2.2 Å X-ray crystal structure of the ternary complex formed by bovine trypsinogen, valine–valine and the Arg 15 analogue of bovine pancreatic trypsin inhibitor. *Eur J Biochem 144*:185–190.

Bode W, Wei AZ, Huber R, Meyer EF Jr, Travis J, Neumann S. 1986b. X-ray crystal structure of the complex of human leukocyte elastase (PMN elastase) and the third domain of the turkey ovomucoid inhibitor. *EMBO J 5*:2453–2458.

Bone R, Agard DA. 1991. Mutational remodeling of enzyme specificity. *Methods Enzymol 202*:643–671.

Bone R, Frank D, Kettner CA, Agard DA. 1989a. Structural analysis of specificity: α-Lytic protease complexes with analogues of reaction intermediates. *Biochemistry 28*:7600–7609.

Bone R, Fujishige A, Kettner CA, Agard DA. 1991. Structural basis for broad specificity in α-lytic protease mutants. *Biochemistry 30*:10388–10398.

Bone R, Shenvi AB, Kettner CA, Agard DA. 1987. Serine protease mechanism: Structure of an inhibitory complex of α-lytic protease and a tightly bound peptide boronic acid. *Biochemistry 26*:7609–7614.

Bone R, Silen JL, Agard DA. 1989b. Structural plasticity broadens the specificity of an engineered protease. *Nature 339*:191–195.

Braxton S, Wells JA. 1991. The importance of a distal hydrogen bonding group in stabilizing the transition state in subtilisin BPN′. *J Biol Chem 266*:11797–11800.

Brayer GD, Delbaere LTJ, James MNG. 1979. Molecular structure of the α-lytic protease from *Myxobacter 495* at 2.8 Å resolution. *J Mol Biol 131*:743–775.

Brenner C, Fuller RS. 1992. Structural and enzymatic characterization of a purified, prohormone-processing enzyme: Secreted, soluble Kex2 protease. *Proc Natl Acad Sci USA 89*:922–926.

Bryan P, Pantoliano MW, Quill SG, Hsiao HY, Poulos T. 1986. Site-directed mutagenesis and the role of the oxyanion hole in subtilisin. *Proc Natl Acad Sci USA 83*:3743–3745.

Caldwell JW, Agard DA, Kollman PA. 1991. Free energy calculations on binding and catalysis by α-lytic protease: The role of substrate size in the P1 pocket. *Proteins Struct Funct Genet 10*:140–148.

Caputo A, James MNG, Powers JC, Hudig D, Bleackley RC. 1994. Conversion of the substrate specificity of mouse proteinase granzyme B. *Struct Biol 1*:364–367.

Carter P, Abrahmsen L, Wells JA. 1991. Probing the mechanism and improving the rate of substrate-assisted catalysis in subtilisin BPN′. *Biochemistry 30*:6142–6148.

Carter P, Nilsson B, Burnier JP, Burdick D, Wells JA. 1989. Engineering subtilisin BPN′ for site-specific proteolysis. *Proteins Struct Funct Genet 6*:240–248.

Carter P, Wells JA. 1987. Engineering enzyme specificity by "substrate-assisted catalysis." *Science 237*:394–399.

Carter P, Wells JA. 1988. Dissecting the catalytic triad of a serine protease. *Nature 332*:564–568.

Carter P, Wells JA. 1990. Functional interactions among catalytic residues in subtilisin BPN′. *Proteins Struct Funct Genet 7*:335–342.

Chasan R, Anderson KV. 1989. The role of Easter, an apparent serine protease, in organizing the dorsal-ventral pattern of the *Drosophila* embryo. *Cell 56*:391–400.

Corey DR, Craik CS. 1992. An investigation into the minimum requirements for peptide hydrolysis by mutation of the catalytic triad of trypsin. *J Am Chem Soc 114*:1784–1790.

Corey DR, McGrath ME, Vasquez JR, Fletterick RJ, Craik CS. 1992. An alternate geometry for the catalytic triad of serine proteases. *J Am Chem Soc 114*:4906–4907.

Corey DR, Shiau AK, Yang Q, Janowski B, Craik CS. 1993. Trypsin display on the surface of bacteriophage. *Gene 128*:129–134.

Craik CS, Largman C, Fletcher T, Roczniak S, Barr PJ, Fletterick RJ, Rutter WJ. 1985. Redesigning trypsin: Alteration of substrate specificity. *Science 228*:291–297.

Craik CS, Roczniak S, Largman C, Rutter WJ. 1987. The catalytic role of the active site aspartic acid in serine proteases. *Science 237*:909–913.

Creemers JWM, Siezen RJ, Roebroek AJM, Ayoubi TAY, Huylebroeck D, Van de Ven WJM. 1993. Modulation of furin-mediated proprotein processing activity by site-directed mutagenesis. *J Biol Chem 268*:21826–21834.

Dancer SJ, Garratt R, Saldanha J, Jhoti H, Evans R. 1990. The epidermolytic toxins are serine proteases. *FEBS Lett 268*:129–132.

Davie EW, Fujikawa K, Kisiel W. 1991. The coagulation cascade: Initiation, maintenance and regulation. *Biochemistry 30*:10363–10370.

Delbaere LTJ, Brayer GD, James MNG. 1979. The 2.8 Å resolution structure of *Streptomyces griseus* protease B and its homology with α-chymotrypsin and *Streptomyces griseus* protease A. *Can J Biochem 57*:135–144.

Dixon GH, Go S, Neurath H. 1956. Peptides combined with ^{14}C-diisopropyl phosphoryl following degradation of ^{14}C-DIP-trypsin with α-chymotrypsin. *Biochim Biophys Acta 19*:193–200.

Drapeau GR. 1978. The primary structure of staphylococcal protease. *Can J Biochem 56*:534–544.

Eder J, Rheinnecker M, Fersht AR. 1993. Hydrolysis of small peptide substrates parallels binding of chymotrypsin inhibitor 2 for mutants of subtilisin BPN′. *FEBS Lett 335*:349–352.

Eisen AZ, Henderson KO, Jeffrey JJ, Bradshaw RA. 1973. A collagenolytic protease from the hepatopancreas of the fiddler crab, *Uca pugilator*. Purification and properties. *Biochemistry 12*:1814–1822.

Epstein DM, Abelès RH. 1992. Role of serine 214 and tyrosine 171, components of the S2 subsite of α-lytic protease, in catalysis. *Biochemistry 31*:11216–11223.

Estell DA, Graycar TP, Miller JV, Powers DB, Burnier JP, Ng PG, Wells JA. 1986. Probing steric and hydrophobic effects on enzyme-substrate interactions by protein engineering. *Science 233*:659–663.

Evnin LB, Vasquez JR, Craik CS. 1990. Substrate specificity of trypsin investigated by using a genetic selection. *Proc Natl Acad Sci USA 87*:6659–6663.

Fujinaga M, Delbaere LTJ, Brayer GD, James MNG. 1985. Refined structure of α-lytic protease at 1.7 Å resolution. Analysis of hydrogen bonding and solvent structure. *J Mol Biol 183*:479–502.

Fujinaga M, James MNG. 1987. Rat submaxillary gland serine protease, tonin—Structure solution and refinement at 1.8 Å resolution. *J Mol Biol 195*:373–396.

Fuller RS, Brake A, Thorner J. 1989. Yeast prohormone processing enzyme (KEX2 gene product) is a Ca^{2+}-dependent serine protease. *Proc Natl Acad Sci USA 86*:1434–1438.

Graf L, Craik CS, Pathy A, Roczniak S, Fletterick RJ, Rutter WJ. 1987. Selective alteration of substrate specificity by replacement of aspartic acid-189 with lysine in the binding pocket of trypsin. *Biochemistry 26*:2616–2623.

Graf L, Jancso A, Szilagyi L, Hegyi G, Pinter K, Naray-Szabo G, Hepp J, Medzihradszky K, Rutter WJ. 1988. Electrostatic complementarity in the substrate binding pocket of trypsin. *Proc Natl Acad Sci USA 85*:4961–4965.

Graham LD, Haggett KD, Jennings PA, LeBrocque DS, Whittaker RG, Schober PA. 1993. Random mutagenesis of the substrate binding site of a serine protease can generate enzymes with increased activities and altered primary specificities. *Biochemistry 32*:6250–6258.

Grant GA, Eisen AZ. 1980. Substrate specificity of the collagenolytic serine protease from *Uca pugilator*: Studies with noncollagenous substrates. *Biochemistry 19*:6089–6095.

Grant GA, Henderson KO, Eisen AZ, Bradshaw RA. 1980. Amino acid sequence of a collagenolytic protease from the hepatopancreas of the fiddler crab, *Uca pugilator*. *Biochemistry 19*:4653–4659.

Greer J. 1990. Comparative modeling methods: Application to the family of the mammalian serine proteases. *Proteins Struct Funct Genet 7*:317–334.

Grøn H, Breddam K. 1992. Interdependency of the binding sites in subtilisin. *Biochemistry 31*:8967–8971.

Grøn H, Meldal M, Breddam K. 1992. Extensive comparison of the substrate preferences of two subtilisins as determined with peptide substrates which are based on the principle of intramolecular quenching. *Biochemistry 31*:6011–6018.

Gros P, Betzel C, Dauter Z, Wilson KS, Hol WGJ. 1989. Molecular dynamics refinement of a thermitase-eglin-c complex at 1.98 Å resolution and comparison of two crystal forms that differ in calcium content. *J Mol Biol 210*:347–367.

Guo J, Huang W, Scanlan TS. 1994. Kinetic and mechanistic characterization of a highly active hydrolytic antibody: Evidence for the formation of an acyl intermediate. *J Am Chem Soc 116*:6062–6069.

Harper JW, Cook RR, Roberts CJ, McLaughlin BJ, Powers JC. 1984. Active site mapping of the serine proteases human leukocyte elastase, cathepsin G, porcine pancreatic elastase, rat mast cell proteases I and II, bovine chymotrypsin Aα and *S. aureus* protease V-8 using tripeptide thiobenzyl ester substrates. *Biochemistry 23*:2995–3002.

Hedstrom L, Farr-Jones S, Kettner CA, Rutter WJ. 1994a. Converting trypsin to chymotrypsin: Ground state binding does not determine substrate specificity. *Biochemistry 33*:8764–8769.

Hedstrom L, Perona JJ, Rutter WJ. 1994b. Converting trypsin to chymotrypsin: Residue 172 is a substrate specificity determinant. *Biochemistry 33*:8757–8763.

Hedstrom L, Szilagyi L, Rutter WJ. 1992. Converting trypsin to chymotrypsin: The role of surface loops. *Science 255*:1249–1253.

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

Higaki JN, Haymore BL, Chen S, Fletterick R, Craik CS. 1990. Regulation of serine protease activity by an engineered metal switch. *Biochemistry 29*:8582–8586.

Horrevoets AJG, Tans G, Smilde AE, van Zonnefeld AJ, Pannekoek H. 1993. Thrombin-variable region 1. Evidence for the dominant contribution of VR1 of serine proteases to their interaction with plasminogen activator inhibitor 1. *J Biol Chem 268*:779–782.

Huber R, Bode W. 1978. Structural basis of the activation and action of trypsin. *Acc Chem Res 11*:114–122.

Hwang JK, Warshel A. 1988. Why ion pair reversal by protein engineering is unlikely to succeed. *Nature 334*:270–272.

Jackson DY, Burnier J, Quan C, Stanley M, Tom J, Wells JA. 1994. A designed peptide ligase for total synthesis of ribonuclease A with unnatural catalytic residues. *Science 266*:243–247.

Jackson SE, Fersht AR. 1993. Contribution of long-range electrostatic interactions to the stabilization of the catalytic transition state of the serine protease subtilisin BPN'. *Biochemistry 32*:13909–13916.

James MNG. 1976. Relationship between the structures and activities of some microbial serine proteases. II. Comparison of the tertiary structures of microbial and pancreatic serine proteases. In: Ribbons DW, Brew K, eds. *Proteolysis and physiological regulation*. New York: Academic Press. pp 125–142.

Joachimiak A, Haran TE, Sigler PB. 1994. Mutagenesis supports water mediated recognition in the *trp* repressor–operator system. *EMBO J 13*: 367–372.

Kabsch W, Sander C. 1983. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers 22*:2577–2637.

Kahne D, Still WC. 1988. Hydrolysis of a peptide bond in neutral water. *J Am Chem Soc 110*:7529–7534.

Kettner CA, Shenvi AB. 1984. Inhibition of the serine proteases leukocyte elastase, pancreatic elastase, cathepsin G and chymotrypsin by peptide boronic acids. *J Biol Chem 259*:15106–15114.

Kossiakoff AA, Spencer SA. 1981. Direct determination of the protonation states of aspartic acid-102 and histidine-57 in the tetrahedral intermediate of the serine proteases: Neutron structure of trypsin. *Biochemistry 20*:6462–6473.

Kraut J. 1977. Serine proteases: Structure and mechanism of catalysis. *Annu Rev Biochem 46*:331–358.

LaVallie ER, Rehemtulla A, Racie LA, DiBlasio EA, Ferentz C, Grant KL, Light A, McCoy JM. 1993. Cloning and functional expression of a cDNA encoding the catalytic subunit of bovine enterokinase. *J Biol Chem 268*: 23311–23317.

Lazure C, Scidah NG, Pelaprat D, Chretien M. 1983. Proteases and post-translational processing of prohormones: A review. *Can J Biochem Cell Biol 61*:501–515.

LeTrong H, Neurath H, Woodbury RG. 1987a. Substrate specificity of the chymotrypsin-like protease in secretory granules isolated from rat mast cells. *Proc Natl Acad Sci USA 84*:364–367.

LeTrong H, Parmelee DD, Walsh KA, Neurath H, Woodbury RG. 1987b. Amino acid sequence of rat mast cell protease I (chymase). *Biochemistry 26*:6988–6994.

Liao D, Breddam K, Sweet RM, Bullock T, Remington SJ. 1992. Refined atomic model of wheat serine carboxypeptidase II at 2.2 Å resolution. *Biochemistry 31*:9796–9812.

Liao D, Remington SJ. 1990. Structure of wheat serine carboxypeptidase II at 3.5 Å resolution. A new class of serine protease. *J Biol Chem 265*:6528–6531.

Light A, Fonseca P. 1984. The preparation and properties of the catalytic subunit of bovine enterokinase. *J Biol Chem 259*:13195–13198.

Lobe CG, Finlay BB, Paranchych W, Paetkau VH, Bleackley RC. 1986. Novel serine proteases encoded by two cytotoxic T lymphocyte-specific genes. *Science 232*:858–861.

Loewenthal R, Sancho J, Reinikainen T, Fersht AR. 1993. Long-range surface charge–charge interactions in proteins. Comparison of experimental results with calculations from a theoretical method. *J Mol Biol 232*: 574–583.

Madison EL, Goldsmith EJ, Gerard RD, Gething MJH, Sambrook JF, Bassel-Duby RS. 1990. Amino acid residues that affect interaction of tissue-type plasminogen activator with plasminogen activator inhibitor 1. *Proc Natl Acad Sci USA 87*:3530–3533.

Magee AI, Grant DAW, Hermon-Taylor J. 1977. The apparent molecular weights of human intestinal aminopeptidase, enterokinase and maltase in native duodenal fluid. *Biochem J 165*:583–585.

Markland FS, Smith EL. 1971. Subtilisins: Primary structure, chemical and physical properties. In: Boyer PD, ed. *The enzymes, vol 3*. New York: Academic Press. pp 516–608.

Markley JL. 1979. Catalytic groups of serine proteases. NMR investigations. In: Shulman RG, ed. *Biological applications of magnetic resonance*. New York: Academic Press. pp 397–461.

Maroux S, Barratti J, Desnuelle P. 1971. Purification and specificity of porcine enterokinase. *J Biol Chem 246*:5031–5039.

Matthews BW. 1977. X-ray structure of proteins. In: Neurath H, Hill RL, eds. *The proteins, vol 3*. New York: Academic Press. pp 404–590.

Matthews BW, Sigler PB, Henderson R, Blow DM. 1967. Three-dimensional structure of tosyl-α-chymotrypsin. *Nature 214*:652–656.

Matthews DJ, Wells JA. 1993. Substrate phage: Selection of protease substrates by monovalent phage display. *Science 260*:1113–1117.

Matthews G, Shennan KI, Seal AJ, Taylor NA, Colman A, Docherty K. 1994. Autocatalytic maturation of the prohormone convertase PC2. *J Biol Chem 269*:588–592.

McGrath ME, Haymore BL, Summers NL, Craik CS, Fletterick RJ. 1993. Structure of an engineered, metal-actuated switch in trypsin. *Biochemistry 32*:1914–1919.

McGrath ME, Vasquez JR, Craik CS, Yang AS, Honig B, Fletterick RJ. 1992. Perturbing the polar environment of Asp 102 in trypsin: Consequences of replacing conserved Ser 214. *Biochemistry 31*:3059–3064.

McGrath M, Wilke ME, Higaki JN, Craik CS, Fletterick R. 1989. Crystal structures of two engineered thiol trypsins. *Biochemistry 28*:9264–9270.

McPhalen CA, James MNG. 1988. Structural comparison of two serine proteinase–protein inhibitor complexes: Eglin C–subtilisin Carlsberg and CI-2–subtilisin Novo. *Biochemistry 27*:6582–6598.

Mizushima N, Spellmeyer D, Hirono S, Pearlman D, Kollman P. 1991. Free energy perturbation calculations on binding and catalysis after mutating threonine 220 in subtilisin. *J Biol Chem 266*:11801–11809.

Mortenson UH, Remington SJ, Breddam K. 1994. Site-directed mutagenesis on (serine) carboxypeptidase Y: A hydrogen-bond network stabilizes the transition state by interaction with the C-terminal carboxylate of the substrate. *Biochemistry 33*:508–517.

Moult J, Sussman F, James MNG. 1985. Electron density calculations as an extension of protein structure refinement. *Streptomyces griseus* protease A at 1.5 Å resolution. *J Mol Biol 182*:555–566.

Murphy MEP, Moult J, Bleackley RC, Gershenfeld H, Weissman IL, James MNG. 1988. Comparative molecular model building of two serine proteinases from cytotoxic T lymphocytes. *Proteins Struct Funct Genet 4*: 190–204.

Nakatsuka T, Sasaki T, Kaiser ET. 1987. Peptide segment coupling catalyzed by the semisynthetic enzyme thiolsubtilisin. *J Am Chem Soc 109*:3808–3810.

Narajana SV, Carson M, el-Kabbiani O, Kilpatrick JM, Moore D, Chen X, Bugg CE, DeLucas LJ. 1994. Structure of human factor D. A complement system protein at 2.0 Å resolution. *J Mol Biol 235*:695–708.

Navia MA, McKeever BM, Springer JP, Lin TY, Williams HR, Fluder EM, Dorn CP, Hoogsteen K. 1989. Structure of human neutrophil elastase in complex with a peptide chloromethyl ketone inhibitor at 1.84 Å resolution. *Proc Natl Acad Sci USA 86*:7–11.

Neurath H. 1984. Evolution of proteolytic enzymes. *Science 224*:350–357.

Neurath H. 1985. Proteolytic enzymes, past and present. *Fed Proc 44*: 2907–2913.

Nienaber VL, Breddam K, Birktoft JJ. 1993. A glutamic acid specific serine protease utilizes a novel histidine triad in substrate binding. *Biochemistry 32*:11469–11475.

Ny T, Sawdey M, Lawrence D, Millan JL, Lotsukoff DJ. 1986. Cloning and sequence of a cDNA coding for the human beta-migrating endothelial-cell-type plasminogen activator inhibitor. *Proc Natl Acad Sci USA 83*:6776–6780.

Odake S, Kam CM, Narasimhan L, Poe M, Blake JT, Krahenbuhl O, Tschopp J, Powers JC. 1991. Human and murine cytotoxic T-lymphocyte serine proteases: Subsite mapping with peptide thioester substrates and inhibition of enzyme activity and cytolysis by isocoumarins. *Biochemistry 30*:2217–2227.

Ollis DL, Cheah E, Cygler M, Dykstra B, Frolow F, Franken SM, Harel M, Remington SJ, Silman I, Schrag J, Sussman JL, Verschueren KHG, Goldman A. 1992. The alpha/beta hydrolase fold. *Protein Eng 5*:197–211.

Otwinowski Z, Schevitz RW, Zhang RG, Lawson CL, Joachimiak A, Mar-

morstein RQ, Luisi BF, Sigler PB. 1988. Crystal structure of *trp* repressor/operator at atomic resolution. *Nature 335*:321–327.

Padmanabhan K, Padmanabhan KP, Tulinsky A, Park CH, Bode W, Huber R, Blankenship DT, Cardin AD, Kisiel W. 1993. Structure of human ds(1–45) factor Xa at 2.2 Å resolution. *J Mol Biol 232*:947–966.

Perona JJ, Evnin LB, Craik CS. 1993a. A genetic selection elucidates structural determinants of arginine versus lysine specificity in trypsin. *Gene 137*:121–126.

Perona JJ, Hedstrom L, Rutter WJ, Fletterick RJ. 1995. Structural origins of substrate discrimination in trypsin and chymotrypsin. *Biochemistry 34*:1489–1499.

Perona JJ, Hedstrom L, Wagner R, Rutter WJ, Craik CS, Fletterick RJ. 1994. Exogenous acetate reconstitutes the enzymatic activity of Asp 189 Ser trypsin. *Biochemistry 33*:3252–3259.

Perona JJ, Tsu CA, Craik CS, Fletterick RJ. 1993b. Crystal structures of rat anionic trypsin complexed with the protein inhibitors APPI and BPTI. *J Mol Biol 230*:919–933.

Perona JJ, Tsu CA, McGrath ME, Craik CS, Fletterick RJ. 1993c. Relocating a negative charge in the binding pocket of trypsin. *J Mol Biol 230*:934–949.

Polgar L. 1989. Structure and function of serine proteases. In: *Mechanisms of protease action*. Boca Raton, Florida: CRC Press. Chapter 3.

Polgar L. 1991. pH-dependent mechanism in the catalysis of prolyl endopeptidase from pig muscle. *Eur J Biochem 197*:441–447.

Poulos TL, Alden RA, Freer ST, Birktoft JJ, Kraut J. 1976. Polypeptide halomethyl ketones bind to serine proteases as analogs of the tetrahedral intermediate. *J Biol Chem 251*:1097–1103.

Powers JC, Tanaka T, Harper JW, Minematsu Y, Barker L, Lincoln D, Crumley KV, Fraki JE, Schechter NM, Lazarus GG, Nakajima K, Nakashino K, Neurath H, Woodbury RG. 1985. Mammalian chymotrypsin-like enzymes. Comparative reactivities of rat mast cell proteases, human and dog skin chymases, and human cathepsin G with peptide 4-nitroanilide substrates and with peptide chloromethyl ketone and sulfonyl fluoride inhibitors. *Biochemistry 24*:2048–2058.

Rao SN, Singh UC, Bash PA, Kollman PA. 1987. Free energy perturbation calculations on binding and catalysis after mutating Asn 155 in subtilisin. *Nature 328*:551–554.

Read RJ, James MNG. 1988. Refined crystal structure of *Streptomyces griseus* trypsin at 1.7 Å resolution. *J Mol Biol 200*:523–551.

Rehemtulla A, Barr PJ, Rhodes CJ, Kaufman RJ. 1993. PACE4 is a member of the mammalian propeptidase family that has overlapping but not identical substrate specificity to PACE. *Biochemistry 32*:11586–11590.

Remington SJ, Woodbury RG, Reynolds RA, Matthews BW, Neurath H. 1988. The structure of rat mast cell protease at 1.9-Å resolution. *Biochemistry 27*:8097–8105.

Rheinnecker M, Baker G, Eder J, Fersht AR. 1993. Engineering a novel specificity in subtilisin BPN′. *Biochemistry 32*:1199–1203.

Rheinnecker M, Eder J, Pandey PS, Fersht AR. 1994. Variants of subtilisin BPN′ with altered specificity profiles. *Biochemistry 33*:221–225.

Robertus JD, Alden RA, Birktoft JJ, Kraut J, Powers JC, Wilcox PE. 1972a. An X-ray crystallographic study of the binding of peptide chloromethyl ketone inhibitors to subtilisin BPN′. *Biochemistry 11*:2439–2449.

Robertus JD, Kraut J, Alden RA, Birktoft J. 1972b. Subtilisin: A stereochemical mechanism involving transition-state stabilization. *Biochemistry 11*:4293–4303.

Rose GD, Creamer TP. 1994. Protein folding: Predicting predicting. *Proteins Struct Funct Genet 19*:1–3.

Ruhlmann A, Kukla D, Schwager P, Bartels K, Huber R. 1973. Structure of the complex formed by bovine trypsin and bovine pancreatic trypsin inhibitor. *J Mol Biol 77*:417–436.

Russell AJ, Thomas PG, Fersht AR. 1987. Electrostatic effects on modification of charged groups in the active site cleft of subtilisin by protein engineering. *J Mol Biol 193*:803–813.

Salvesen G, Farley D, Shuman J, Przybyla A, Reilly C, Travis J. 1987. Molecular cloning of human cathepsin G: Structural similarity to mast cell and cytotoxic T lymphocyte proteinases. *Biochemistry 26*:2289–2293.

Schechter I, Berger A. 1968. On the size of the active site in proteases. I. Papain. *Biochem Biophys Res Commun 27*:157–162.

Schellenberger V, Turck CW, Hedstrom L, Rutter WJ. 1993. Mapping the S′ subsites of serine proteases using acyl transfer to mixtures of peptide nucleophiles. *Biochemistry 32*:4349–4353.

Schellenberger V, Turck CW, Rutter WJ. 1994. Role of the S′ subsites in serine protease catalysis. Active-site mapping of rat chymotrypsin, rat trypsin, α-lytic protease and cercarial protease from *Schistosoma mansoni*. *Biochemistry 33*:4251–4257.

Seidah NG, Day R, Marcinkiewicz M, Benjannet S, Chretien M. 1991. Mammalian neural and endocrine pro-protein and pro-hormone convertases belonging to the subtilisin family of serine proteases. *Enzyme 45*:271–284.

Sellos D, Van Wormhoudt A. 1992. Molecular cloning of a cDNA that encodes a serine protease with chymotryptic and collagenolytic activities in the hepatopancreas of the shrimp *Penaeus vanameii* (Crustacea, Decapoda). *FEBS Lett 309*:219–224.

Shakked Z, Guzikevich-Guerstein G, Frolow F, Rabinovich D, Joachimiak A, Sigler PB. 1994. Determinants of repressor–operator recognition from the structure of the *trp* operator binding site. *Nature 368*:469–473.

Shaw E, Mares-Guia M, Cohen W. 1965. Evidence for an active site histidine in trypsin through use of a specific reagent, 1-chloro-3-tosylamido-7-amino-2-heptanone, the chloromethyl ketone derived from N^{α}-tosyl-L-lysine. *Biochemistry 4*:2219–2226.

Siezen RJ, Bruinenberg PG, Vos P, van Alen-Boerrigter I, Nijhuis M, Alting AC, Exterkate FA, de Vos WM. 1993. Engineering of the substrate-binding region of the subtilisin-like, cell-envelope proteinase of *Lactococcus lactis*. *Protein Eng 6*:927–937.

Siezen RJ, de Vos WM, Leunissen JAM, Dijkstra BW. 1991. Homology modelling and protein engineering strategy of subtilases, the family of subtilisin-like serine proteases. *Protein Eng 4*:717–719.

Sinha S, Watorek W, Karr S, Giles J, Bode W, Travis J. 1987. Primary structure of human neutrophil elastase. *Proc Natl Acad Sci USA 84*:2228–2232.

Smeekens SP, Avruch AS, LaMendola J, Chan SJ, Steiner DF. 1991. Identification of a cDNA encoding a second putative prohormone convertase related to PC2 in AtT20 cells and islets of Langerhans. *Proc Natl Acad Sci USA 88*:340–344.

Smith CL, DeLotto R. 1994. Ventralizing signal determined by protease activation in *Drosophila* embryogenesis. *Nature 368*:548–551.

Sørensen SB, Bech LM, Meldal M, Breddam K. 1993. Mutational replacements of the amino acid residues forming the hydrophobic S4 binding pocket of subtilisin 309 from *Bacillus lentus*. *Biochemistry 32*:8994–8999.

Sprang S, Standing T, Fletterick RJ, Stroud RM, Finer-Moore J, Xuong NH, Hamlin R, Rutter WJ, Craik CS. 1987. The three-dimensional structure of Asn 102 mutant of trypsin: Role of Asp 102 in serine protease catalysis. *Science 237*:905–909.

Stein RL, Strimpler AM, Hori H, Powers JC. 1987. Catalysis by human leukocyte elastase: Mechanistic insights into specificity requirements. *Biochemistry 26*:1301–1305.

Steitz TA, Shulman RG. 1982. Crystallographic and NMR studies of the serine proteases. *Annu Rev Biophys Bioeng 11*:419–444.

Stroud RM. 1974. A family of protein-cutting proteins. *Sci Am 23*:74–88.

Svendsen I, Jensen MR, Breddam K. 1991. The primary structure of the glutamic acid-specific protease of *Streftomyces griseus*. *FEBS Lett 292*:165–167.

Takeuchi Y, Noguchi S, Satow Y, Kojima S, Kumagai I, Miura K, Nakamura KT, Mitsui Y. 1991a. Molecular recognition at the active site of subtilisin BPN′: Crystallographic studies using genetically engineered proteinaceous inhibitor SSI (*Streptomyces* subtilisin inhibitor). *Protein Eng 4*:501–508.

Takeuchi Y, Satow Y, Nakamura KT, Mitsui Y. 1991b. Refined crystal structure of the complex of subtilisin BPN′ and *Streptomyces* subtilisin inhibitor at 1.8 Å resolution. *J Mol Biol 221*:309–325.

Teplyakov AV, van der Laan JM, Lammers AA, Kelders H, Kalk KH, Misset O, Mulleners LSJM, Dijkstra BW. 1992. Protein engineering of the high-alkaline serine protease PB92 from *Bacillus alcalophilus*: Functional and structural consequences of mutation at the S4 substrate binding pocket. *Protein Eng 5*:413–420.

Thompson RC, Blout ER. 1970. Dependence of the kinetic parameters for elastase-catalyzed amide hydrolysis on the length of peptide substrates. *Proc Natl Acad Sci USA 67*:1734–1743.

Tsu CA, Perona JJ, Schellenberger V, Turck CW, Craik CS. 1994. The substrate specificity of *Uca pugilator* collagenolytic serine protease 1 correlates with the bovine type I collagen cleavage sites. *J Biol Chem 269*:19565–19572.

Van den Ouweland AMW, Van Duijnhoven HLP, Keizer GD, Dorssers LCJ, Van de Ven WJM. 1990. Structural homology between the human *fur* gene product and the subtilisin-like protease encoded by yeast KEX2. *Nucleic Acids Res 18*:664–674.

van der Laan JM, Teplyakov AV, Kelders H, Kalk KH, Misset O, Mulleners LJSM, Dijkstra BW. 1992. Crystal structure of the high-alkaline serine protease PB92 from *Bacillus alcalophilus*. *Protein Eng 5*:405–411.

Van de Ven WJ, Roebroek AJ, Van Duijnhoven HL. 1993. Structure and function of eukaryotic proprotein processing enzymes of the subtilisin family of serine proteases. *Crit Rev Oncogen 4*:115–136.

Warshel A, Naray-Szabo G, Sussman F, Hwang JK. 1989. How do serine proteases really work? *Biochemistry 28*:3629–3637.

Watson HC, Shotton DM, Cox JC, Muirhead H. 1970. Three-dimensional Fourier synthesis of tosyl elastase at 3.5 Å resolution. *Nature 225*:806–811.

Wei AZ, Mayr I, Bode W. 1988. The refined 2.3 Å crystal structure of human leukocyte elastase in a complex with a valine chloromethyl ketone inhibitor. *FEBS Lett 234*:367–373.

Wells JA, Cunningham BC, Graycar TP, Estell DA. 1986. Importance of hydrogen bond formation in stabilizing the transition state of subtilisin. *Philos Trans R Soc Lond A 317*:415–423.

Wells JA, Cunningham BC, Graycar TP, Estell DA. 1987a. Recruitment of substrate-specificity properties from one enzyme into a related one by protein engineering. *Proc Natl Acad Sci USA 84*:5167–5171.

Wells JA, Cunningham BC, Graycar TP, Estell DA, Carter P. 1987b. On the evolution of specificity and catalysis in subtilisin. *Cold Spring Harbor Symp Quant Biol 52*:647–652.

Wells JA, Estell DA. 1988. Subtilisin—An enzyme designed to be engineered. *Trends Biochem Sci 13*:291–297.

Wells JA, Powers DB, Bott RR, Graycar TP, Estell DA. 1987c. Designing substrate specificity by protein engineering of electrostatic interactions. *Proc Natl Acad Sci USA 84*:1219–1223.

Wilke ME, Higaki JN, Craik CS, Fletterick RJ. 1991. Crystallographic analysis of trypsin G226A. A specificity pocket mutant of rat trypsin with altered binding and catalysis. *J Mol Biol 219*:525–532.

Willett WS, Gillmor S, Perona JJ, Fletterick RJ, Craik CS. 1995. Engineered metal regulation of trypsin substrate specificity. *Biochemistry*. Forthcoming.

Wilson C, Mace J, Agard DA. 1991. Computational method for the design of enzymes with altered substrate specificity. *J Mol Biol 220*:495–506.

Woodbury RG, Everitt MT, Sanada Y, Katunuma N, Lagunoff D, Neurath H. 1978a. A major serine protease in rat skeletal muscle: Evidence for its mast cell origin. *Proc Natl Acad Sci USA 75*:5311–5313.

Woodbury RG, Gruzenski GM, Lagunoff D. 1978b. Immunofluorescent localization of a serine protease in rat small intestine. *Proc Natl Acad Sci USA 75*:2785–2789.

Wright CS, Alden RA, Kraut J. 1969. Structure of subtilisin BPN′ at 2.5 Å resolution. *Nature 221*:235–242.

Yoshida N, Everitt MT, Neurath H, Woodbury RG, Powers JC. 1980. Substrate specificity of two chymotrypsin-like proteases from rat mast cells. Studies with peptide 4-nitroanilides and comparison with cathepsin G. *Biochemistry 19*:5799–5804.

Zerner B, Bender ML. 1964. The kinetic consequences of the acyl-enzyme mechanism for the reactions of specific substrates with chymotrypsin. *J Am Chem Soc 86*:3669–3674.

Zhou GW, Guo J, Huang W, Fletterick RJ, Scanlan TS. 1994. The three-dimensional structure of a catalytic antibody with active site similarity to serine proteases. *Science 265*:1059–1064.