



## 452. Trypsin

---

### Databanks

MEROPS name: Trypsin 1 (tetrapod, non-human)

MEROPS classification: clan [PA\(S\)](#), family [S1A](#), peptidase [S01.151](#)

IUBMB: EC 3.4.21.4

CAS registry: 9002-07-7

Species distribution: *Chordata*

Sequence known from: *Balaenoptera acutorostrata*, *Bos taurus*, *Bothrops jararaca*, *Canis familiaris*, *Gallus gallus*, *Mus musculus*, *Rana esculenta*, *Rattus norvegicus*, *Sus scrofa*

Tertiary structure: Available

### Name and History

Trypsin was first described and named in 1876 by Kühne as the proteolytic activity in pancreatic secretions (Kühne, 1876; Neurath & Zwilling, 1986). Kühne differentiated this activity from that of pepsin by the higher optimal pH of trypsin. As separation and characterization of the individual pancreatic proteases was achieved, the name **trypsin** became associated with the proteolytic activity which cleaved peptide bonds C-terminal to Arg or Lys. The ready availability of trypsin from the pancreas of cattle allowed the enzyme to be purified by crystallization in 1931 (Northrop & Kunitz, 1931).

### Activity and Specificity

Trypsin can be seen as a prototype of the serine endopeptidases of family S1, and much of the fundamental knowledge about the family has been derived from the study of this enzyme (Perona & Craik, 1995). Trypsin strongly prefers to cleave amide substrates following P1 Arg or Lys residues. The preference for these basic side chains is reflected by relative values for catalytic efficiency ( $k_{\text{cat}} / K_m$ ) at least  $10^5$  greater than for other natural amino acids. The preference for Arg over Lys is 2- to 10-fold (Craik *et al.*, 1985). However, discrimination among ester substrates is much less strict. Secondary binding sites on both sides of the scissile bond play only a very minor role in the determination of substrate specificity, although occupancy of these sites does contribute to catalytic efficiency (Corey *et al.*, 1992; Schellenberger *et al.*, 1994). The specific identities of the amino acids occupying positions P2 through P4 do not enhance the activity or specificity of trypsin. However, tetrapeptide substrate libraries (Harris *et al.*, 2000) illustrate that certain amino acids in specific positions may inhibit activity. Relative to other amino acids, an Arg, Ile, Leu, Lys or Phe at P2 decreases the activity 2- to 16-fold while a Pro residue at position P3 decreases activity 3- to 9-fold. The residue occupying P4 does not affect activity (Baird *et al.*, 2001). In small synthetic substrates, formation of an acyl enzyme intermediate is usually the rate-

determining step for the cleavage of amide bonds by trypsin, whereas hydrolysis of this intermediate is the rate-determining step for ester cleavage. The binding of the substrate influences not only  $K_m$  but also  $k_{cat}$ . In fact, the acylation rate with a substrate is a major specificity determinant (Hedstrom *et al.*, 1994b), and this is one reason why trypsin is much more promiscuous with ester substrates than with peptides. With protein substrates, the binding event may be the rate-limiting step.

Typical spectroscopic assay substrates include Bz-Arg-OEt, Tos-Arg-OMe, Bz-Arg-NHPhNO<sub>2</sub> and Suc-Ala-Ala-Pro-Arg+NHPhNO<sub>2</sub> (Zimmerman *et al.*, 1977). Activity may also be easily monitored fluorimetrically with substrates containing an aminomethylcoumarin group or one of its derivatives, e.g. Z-Gly-Pro-Arg+NHMeC (Zimmerman *et al.*, 1977). The Pro commonly utilized in the P2 position of these substrates helps to align the substrate for productive binding, since positioning of this residue in either the P3 or P1 positions is strongly disfavored. A commonly used peptide substrate is the insulin B chain (Corey *et al.*, 1992), substrate cleavage being monitored by HPLC. A gel overlay assay has also been developed that can detect subnanogram amounts of trypsin (Vasquez *et al.*, 1989). Active-site titrations are usually accomplished with *p*-nitrophenyl-*p'*-guanidinobenzoate (Chase & Shaw, 1967) or 4-methylumbelliferyl-*p*-guanidinobenzoate (Jameson *et al.*, 1973).

The pH optimum of trypsin is approximately 8, although this varies slightly with species. The reaction buffer is required to contain moderate amounts (20 mM) of CaCl<sub>2</sub> for maximal activity and stability of the protease. Under these conditions, the catalytic efficiency ( $k_{cat} / K_m$ ) of insulin B chain cleavage by rat trypsin is 18 min<sup>-1</sup> μM<sup>-1</sup>, whereas the value for the smaller substrate, Z-Gly-Pro-Arg-NHMeC, is 210 min<sup>-1</sup> μM<sup>-1</sup> (Corey *et al.*, 1992).





Trypsin is stable for extended periods of time as a lyophilized powder, or in solution at pH 3 (in which it is largely inactive). The enzyme from natural sources is available from many commercial suppliers at various degrees of purity. Tos-Phe-CH<sub>2</sub>Cl-treatment of the enzyme minimizes contamination by active chymotrypsin. Trypsin is also available from Sigma (see Appendix 2 for full names and addresses of suppliers) bound to agarose and to polyacrylamide for easy separation from tryptic digests. Recombinant forms of trypsin have also been made (see below).

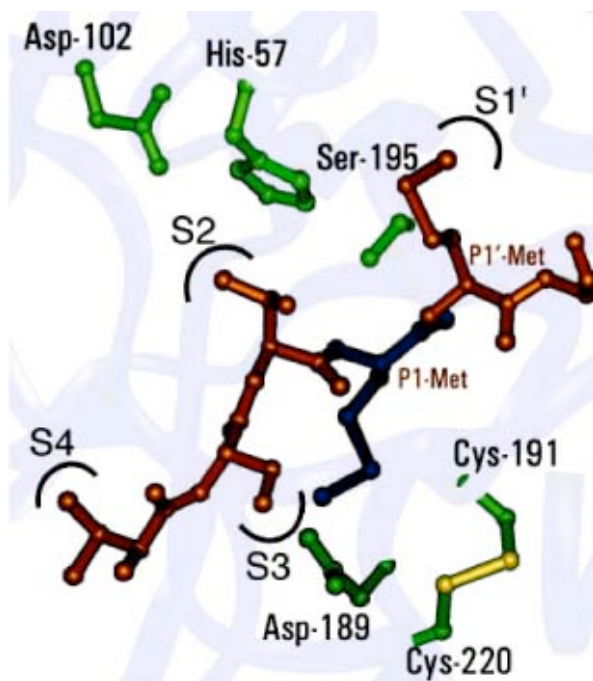
Many general serine protease inhibitors (i.e. PMSF, DFP, DCI) inhibit trypsin, but a greater specificity for enzymes with trypsin-like specificity is shown by leupeptin, benzamidine, Tos-Lys-CH<sub>2</sub>Cl, and APMSF. Protein inhibitors include ecotin, soybean trypsin inhibitor, aprotinin,  $\alpha_2$ -macroglobulin, and  $\alpha_1$ -proteinase inhibitor. Trypsin can also be reversibly denatured by urea (Higaki & Light, 1985).

### **Structural Chemistry**

The three-dimensional structure of cattle trypsin was determined in 1974 independently by two groups (Huber *et al.*, 1974; Stroud *et al.*, 1974) and this structure has become the prototype for the S1 family of proteases. Structural analyses of eukaryotic and prokaryotic members of this family have revealed a common three-dimensional structure (Delbaere *et al.*, 1975; Bode *et al.*, 1983).

The tertiary structures of the enzymes belonging to the S1 family are strongly conserved, and this is seen very clearly for trypsin. Although the primary structures of trypsins can vary substantially, the folds are closely similar. For example, the cattle cationic and rat anionic trypsin backbones have a root-mean-square (r.m.s.) deviation of only 0.29 Å (Sprang *et al.*, 1987) and there is an r.m.s. deviation of only about 1 Å between the hydrophobic cores of the prokaryotic trypsin of *S. griseus* and rat anionic trypsin. The usual numbering system of the residues follows that of the homologous protease chymotrypsin (Chapter 455). Chymotrypsin has a nearly identical tertiary structure, but is less than 50% identical in primary structure, and differs from trypsin in substrate specificity. The positions of key residues, such as those of the catalytic triad, are identical in the two proteases. Active trypsin consists of a single polypeptide chain, in which the catalytic residues bridge two  $\beta$ -barrel domains. Other forms (termed  $\alpha$ ,  $\gamma$  and  $\omega$ ), in which the polypeptide backbone has been clipped, also possess varying degrees of activity (Higaki & Light, 1985). As in other serine proteases of the family, the most important catalytic residues are those of the Asp/His/Ser triad (Figure 452.1). Ser195 acts as a nucleophile in the cleavage reaction, producing an acyl enzyme intermediate. The catalytic triad is no longer thought to act as a 'charge-relay system' as was once widely assumed, but instead His57 is thought to act as a general base (Kossiakoff & Spencer, 1981). Asp102 is believed to stabilize the correct tautomer of His57, and to provide compensation for the developing positive charge during the catalytic reaction. Replacement of this aspartic acid with asparagine results in an enzyme that is approximately 10<sup>4</sup>-fold less active than the wild-type enzyme on synthetic substrates with an aminomethylcoumarin leaving group (Craik *et al.*, 1987),

while relocation of Asp102 to position 214 yields a protease which retains approximately 0.5% of the wild-type activity on peptide substrates. Replacement of either His57 or Ser195 with alanine results in a protease which is  $10^5$ - to  $10^6$ -fold less active than the wild-type enzyme on synthetic substrates with an aminomethylcoumarin leaving group (Corey & Craik, 1992). The residual activity in the His57Ala and Ser195Ala mutants is believed to be due to other structural features in the protease which help to stabilize the tetrahedral intermediate.



*Figure 452.1* Active site of trypsin complexed with ecotin. The Asp/His/Ser triad, Asp of the specificity pocket and Cys residues 191 and 220 are colored green. The Cys191-Cys220 disulfide bridge is colored yellow. In the macromolecular inhibitor ecotin (brown), position P1, the primary determinant of specificity, is occupied by Met (blue). S4, S3 and S2 denote the corresponding binding pockets for the P4 (Val), P3 (Ser) and P2 (Thr) residues N-terminal to the scissile bond. S1' denotes the binding pocket for P1' (Met). The preferred substrates for trypsin are P3 (Gly), P2 (Pro), P1 (Arg/Lys) (see text)

Work with recombinant trypsin has addressed the role of particular amino acid residues through site-directed mutagenesis (Sprang *et al.*, 1987; McGrath *et al.*, 1989; Corey & Craik, 1992; Corey *et al.*, 1992). While replacement of the active-site His57 by Ala reduced the activity of the enzyme toward peptide substrates by four orders of magnitude, significant 'substrate-assisted catalysis' could be observed with peptide substrates in which the function of the catalytic His could be fulfilled by the substrate (Corey *et al.*, 1995). In addition, it could be shown that many parts of the trypsin molecule contribute to the specific recognition of substrate (Craik *et al.*, 1985; Gráf *et al.*, 1987; Hedstrom *et al.*, 1992, 1994a; Perona *et al.*, 1993, 1995). Subsequent studies sought to alter the activity and substrate specificity of the enzyme. Trypsin has been engineered to cleave the Arg+Arg bond when occupying positions P1' and P1. Additionally, His<sup>+</sup> substrate specificity has been engineered into the subsites of trypsin by creating metal-binding sites that bridge the substrate and enzyme (Willett *et al.*,

1996). Engineered metal-binding sites were also shown to be useful in the reversible regulation of trypsin activity: several variants that involve the active site histidine in metal binding were effective at allowing the reversible inhibition of trypsin with submicromolar concentrations of transition metal ions (Higaki *et al.*, 1990; Brinen *et al.*, 1996; Halfon & Craik, 1996).





All naturally occurring trypsins are synthesized as proenzymes. The mammalian propeptide (usually a hexapeptide) contains the consensus sequence for cleavage by enteropeptidase (Chapter 461),  $-(\text{Asp})_4\text{-Lys}$ -preceding the mature N-terminal sequence Ile16-Val17-Gly18-Gly19- (chymotrypsinogen numbering). Cleavage of the propeptide results in disruption of a His40 to Asp194 hydrogen bond, and this is followed by rotation of Asp194 so that it can interact with the new N-terminus at Ile16. This conformational change completes formation of the oxyanion hole (comprising backbone amides of Gly193 and Ser195) and the binding pockets (Fehlhammer *et al.*, 1977; Bode *et al.*, 1978). Stabilization of these new conformations in the so-called activation domain is due principally to hydrophobic interactions of the Ile16 side chain (Hedstrom *et al.*, 1996). Only small adjustments in position are made in other regions (including the catalytic residues) of the protein during the activation process.

The substrate forms an antiparallel  $\beta$  sheet with the protein-binding site. The substrate specificity is primarily determined by the Asp189 side chain which lies at the bottom of the S1 binding pocket. Substitution of this residue with Ser results in a  $10^5$ -fold decrease in  $k_{\text{cat}} / K_{\text{m}}$  for Arg/Lys substrates (with most of the decrease coming from a lowered  $k_{\text{cat}}$ ), without providing significant activity towards substrates with hydrophobic side chains (Gráf *et al.*, 1987, 1988). Removal of the side chain at position 189 in the Asp189Gly variant permits binding of a well-ordered acetate ion in a similar position (Perona *et al.*, 1994). High concentrations of acetate increase the catalytic efficiency of the variant enzyme by 300-fold, demonstrating that the negative charge at the base of the trypsin specificity pocket may be provided by a noncovalently bound ligand.

In the case of Arg-containing substrates, a direct interaction occurs between the substrate guanidinium group and the carboxyl group of Asp189, but for Lys side chains, the contact is mediated by a water molecule. These specific interactions can be selectively altered by introducing Ala at positions 216 and 226 (normally Gly) in rat and cattle trypsin. This in turn dramatically alters the Arg/Lys specificity of the resulting enzymes (Craik *et al.*, 1985). A disulfide bridge between Cys191 and Cys220, and a loop comprising residues 214–220 also contribute to the structure of the specificity pocket. However, mutagenesis studies indicate that the absence of this disulfide bridge does not significantly affect its enzymatic activity or substrate specificity (Wang *et al.*, 1997). Numerous experiments have shown that trypsin surface loops comprising amino acids 185–193 (loop 1) and amino acids 217–224 (loop 2) strongly influence the specificity of the enzyme, even though they do not directly contact the substrate (Perona *et al.*, 1995; Perona & Craik, 1995).

Other structural features of trypsin include a high-affinity calcium-binding site, which is required for stability of the enzyme; autodegradation quickly occurs in its absence. This site is formed by the loop Glu70–Glu80 (Bode & Schwager, 1975). The protein has six completely conserved disulfide bonds, at positions 15–145, 33–49, 117–218, 124–191, 156–170, and 181–205. The 'autolysis' loop, comprising residues 143–151, is very flexible in both trypsin and trypsinogen. Cleavage of this loop at Lys145 yields the  $\alpha$ -trypsin form which retains some catalytic activity. This and other clipped forms are present in most preparations of trypsin (Higaki & Light, 1985). Molecular mass values

are approximately 25 kDa, while the pI values can vary widely for forms of trypsin, both cationic and anionic forms existing in many species. Trypsin and its zymogen form, trypsinogen, contain no post-translational modifications aside from the proteolytic processing required for activity.

### **Preparation**

Trypsin is readily available from many sources. Cattle pancreatic trypsin is commercially available from Sigma, Boehringer Mannheim, Worthington and Fluka. However, it should be noted that trypsin from commercial sources is often contaminated with other pancreatic enzymes. Recombinant trypsin has been expressed in many different systems, including *Escherichia coli* (Higaki *et al.*, 1989; Vasquez *et al.*, 1989; Evnin *et al.*, 1990), *Saccharomyces cerevisiae* (Hedstrom *et al.*, 1992), and *Pichia pastoris* (Halfon & Craik, 1996). The mature form has been successfully expressed in *E. coli* (Vasquez *et al.*, 1989). Purification from these systems often involves affinity chromatography on immobilized benzamidine or aprotinin.

Genetic selections have been established for isolating trypsins from libraries of variants. An *in vivo* selection in bacteria with a dynamic range of five orders of magnitude has been used to isolate trypsins with altered substrate specificities (Evnin *et al.*, 1990; Perona *et al.*, 1993). Recently, trypsin was displayed on bacteriophage, permitting an *in vitro* selection of trypsin activity (Corey *et al.*, 1993).

### **Biological Aspects**

Trypsin is one of several digestive enzymes secreted into the intestine of animals. It is found in all animals, including insects, fish and mammals. In bovine pancreatic secretions, it represents approximately 15% of the digestive enzymes (Keller *et al.*, 1958). Trypsin is synthesized as a preproenzyme by the acinar cells of the pancreas and is stored as the proenzyme trypsinogen in secretory granules. Following release into the gut, trypsinogen is activated by enteropeptidase (Chapter 461), or by trypsin itself. Once activated, the enzyme is responsible for the activation of the proenzymes of all the other digestive enzymes such as chymotrypsin (see Chapter 455) and elastase (see Chapters 457 and 458), and contributes to the digestion of consumed protein.





Numerous genes encoding both anionic and cationic forms of trypsin are present in most animal species, together with the corresponding proteins. The individual proteins are expressed at differing levels. For instance, anionic and cationic trypsins have been isolated from human (Guy *et al.*, 1978) (Chapter 453), cattle (Louvard & Puigserver, 1974), dog (Ohlsson & Tegner, 1973) and rat (Brodrick *et al.*, 1980). In general, the percentage sequence identity of cationic trypsins of different organisms is closer than to the anionic forms in the same species. In most species, one cationic form predominates, while two anionic trypsins are present at much lower levels (less than 10% of the total). In rats, however, the anionic variant is the major form (Fletcher *et al.*, 1987). The reason for the presence of these multiple forms is not known. Since the various forms have distinct amino acid sequences and the differences are distributed throughout the protein, they are clearly products of separate genes. In the adult rat, there are at least ten genes which encode trypsinogen (Craik *et al.*, 1984). Approximately 2–5% of the total adult rat pancreatic mRNA encodes anionic trypsin (MacDonald *et al.*, 1982). There are eight human genes, of which five are transcribed to RNA, two are pseudogenes, and one is a relic gene (Rowen *et al.*, 1996). The human trypsinogen genes are intercalated in two pieces within the human  $\beta$ -T cell receptor locus on chromosome 7q35 (Honey *et al.*, 1984a; Rowen *et al.*, 1996). At the 5'-end of the locus lie the three nonfunctional trypsinogen genes, while other genes, including those coding for the known trypsinogen isoenzymes, lie approximately 500 kb 3'-terminal to these, near the opposite end of the locus. These corresponding genes have also been found in the TCR locus (Whitcomb *et al.*, 1996) in the mouse on chromosome 6 (Honey *et al.*, 1984b) and have been identified in the chicken (Wang *et al.*, 1995). The rat genes corresponding to the predominant trypsin forms have been shown to contain four introns (Craik *et al.*, 1984). Although these introns are positioned at the same sites within the genes, the introns have no homology across variants. The intron–exon junctional amino acids of trypsin and related serine proteases map to the surface of the enzyme and provide a model for evolution of the structure and function of the enzymes through 'junctional sliding' (Craik *et al.*, 1982, 1983).

It has been suggested that hereditary pancreatitis may be associated with an Arg117His mutation in the human cationic trypsin which prevents autolysis at this site (Whitcomb *et al.*, 1996) (see also Chapter 453). The consequent stability of the mutated enzyme could result in increased proteolytic activity that could damage the pancreas. Some reports have indicated expression of very low levels of trypsin in nonpancreatic tissues on the basis of PCR analysis (Wiegand *et al.*, 1993; Wang *et al.*, 1995).

### ***Distinguishing Features***

The form of trypsin present in higher animals has very little sequence identity with microbial trypsins, but is similar structurally. In particular, *Streptomyces griseus* trypsin has a very similar fold to these proteases (Read & James, 1988). Other microbial proteases are more distantly related, having shorter amino acid sequences and corresponding surface loops. For instance,  $\alpha$ -lytic protease (Chapter 446) has the same fold as trypsin, but differs greatly in many structural aspects (James, 1976). Many other serine proteases such as the kallikreins, elastase and chymotrypsin comprise the trypsin

family and these are very similar structurally and catalytically to trypsin, but differ in other key aspects such as substrate preferences. Therefore, while the fold of this class of enzymes is similar among all its members, the activities and functions can vary greatly.

In addition to the conservation of the three-dimensional fold in all known trypsins, the catalytic triad is entirely conserved, and amino acids flanking these residues are also conserved. Degenerate oligonucleotide probes have been designed for the sequences flanking the active site Ser and His residues, and used in PCR to isolate and identify trypsins and related serine proteases from organisms of diverse phyla (Sakanari *et al.*, 1989).

### ***Related Enzymes***

Trypsin-like enzymes are present in many different organisms. Bacterial trypsins, such as those from *Streptomyces* species are dealt with elsewhere (Chapters 450 and 451). The trypsin from the fungus *Fusarium oxysporum* is equally (approximately 45%) identical to the bacterial and mammalian trypsins, and all trypsins are thought to have evolved from a common ancestor (Hewett-Emmett *et al.*, 1981). While functional parts of the proteins are strongly conserved, other regions are less so (Rypniewski *et al.*, 1993). For instance, the *Fusarium oxysporum* trypsin lacks a calcium-binding site, and has a propeptide that lacks homology to the mammalian trypsins. The optimal activity of this enzyme is at 40°C and pH 9.5.

### ***Further Reading***

The reader is referred to the articles of Hedstrom (1996) and Perona & Craik (1995).

### ***Acknowledgments***

This work was supported by a grant from the National Science Foundation MCB-9604379 (to C.S.C.) and an NSF post-doctoral fellowship (to T.T.B.).





## References

- Baird, T.T., Jr., Wang, B., Lodder, M., Hecht, S.M. & Craik, C.S. (2000) Generation of active trypsin with chemical cleavage. *Tetrahedron Lett.* **56**, [PMID:94779485](#).
- Bode, W. & Schwager, P. (1975) The refined crystal structure of bovine  $\beta$ -trypsin at 1.8 Å resolution. *J. Mol. Biol.* **98**, 693–717. [PMID:512](#)
- 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. [PMID:6551452](#)
- Bode, W., Schwager, P. & Huber, R. (1978) The transition of bovine trypsinogen to a trypsin-like state upon strong ligand binding. The refined crystal structures of the bovine trypsinogen-pancreatic trypsin inhibitor complex and of its ternary complex with Ile-Val at 1.9 Å resolution. *J. Mol. Biol.* **118**, 99–112. [PMID:625059](#)
- Brinen, L.S., Willett, W.S., Craik, C.S. & Fletterick, R.J. (1996) X-ray structures of a designed binding site in trypsin show metal-dependent geometry. *Biochemistry* **35**, 5999–6009. [PMID:8634241](#)
- Brodrick, J.W., Largman, C., Geokas, M.C., O'Rourke, M. & Ray, S.B. (1980) Clearance of circulating anionic and cationic pancreatic trypsinogens in the rat. *Am. J. Physiol.* **239**, G511–515.
- Chase, T., Jr. & Shaw, E. (1967) *p*-Nitrophenyl-*p*'-guanidinobenzoate HCl: a new active site titrant for trypsin. *Biochem. Biophys. Res. Commun.* **29**, 508–514.
- Corey, D.R. & Craik, C.S. (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, D.R., McGrath, M.E., Vasquez, J.R., Fletterick, R.J. & Craik, C.S. (1992) An alternate geometry for the catalytic triad of serine proteases. *J. Am. Chem. Soc.* **114**, 4905–4907.
- Corey, D.R., Shiau, A.K., Yang, Q., Janowski, B.A. & Craik, C.S. (1993) Trypsin display on the surface of bacteriophage. *Gene* **128**, 129–134. [PMID:8508953](#)
- Corey, D.R., Willett, W.S., Coombs, G.S. & Craik, C.S. (1995) Trypsin specificity increased through substrate-assisted catalysis. *Biochemistry* **34**, 11521–11527.
- Craik, C.S., Sprang, S., Fletterick, R. & Rutter, W.J. (1982) Intron-exon splices junctions map at protein surfaces. *Nature* **299**, 180–182. [PMID:7110339](#)
- Craik, C.S., Rutter, W.J. & Fletterick, R. (1983) Splice junctions: association with variation in protein structure. *Science* **220**, 1125–1129. [PMID:6344214](#)
- Craik, C.S., Choo, Q.L., Swift, G.H., Quinto, C., MacDonald, R.J. & Rutter, W.J. (1984)

Structure of two related rat pancreatic trypsin genes. *J. Biol. Chem.* **259**, 14255–14264. [PMID:6094547](#)

Craik, C.S., Largman, C., Fletcher, T., Barr, P., Fletterick, R. & Rutter, W.J. (1985) Redesigning trypsin: alteration of substrate specificity, catalytic activity and protein conformation. *Science* **228**, 291–297. [PMID:3838593](#)

Craik, C.S., Rocznik, S., Largman, C. & Rutter, W.J. (1987) The catalytic role of the active site aspartic acid in serine proteases. *Science* **237**, 909–913. [PMID:3303334](#)

Delbaere, L.T., Hutcheon, W.L., James, M.N. & Thiessen, W.E. (1975) Tertiary structural differences between microbial serine proteases and pancreatic serine enzymes. *Nature* **257**, 758–763. [PMID:1186854](#)

Evnin, L.B., Vasquez, J.R. & Craik, C.S. (1990) Substrate specificity of trypsin investigated by using a genetic selection. *Proc. Natl. Acad. Sci. USA* **87**, 6659–6663. [PMID:2204062](#)

Fehlhammer, H., Bode, W. & Huber, R. (1977) Crystal structure of bovine trypsinogen at 1.8 Å resolution. *J. Mol. Biol.* **111**, 415–438. [PMID:864704](#)

Fletcher, T.S., Alhadeff, M., Craik, C.S. & Largman, C. (1987) Isolation and characterization of a cDNA encoding rat cationic trypsinogen. *Biochemistry* **26**, 3081–3086. [PMID:3607011](#)

Gráf, L., Craik, C.S., Patthy, A., Rocznik, S., Fletterick, R.J. & Rutter, W.J. (1987) Selective alteration of substrate specificity by replacement of aspartic acid-89 with lysine in the binding pocket of trypsin. *Biochemistry* **26**, 2616–2623. [PMID:3111531](#)

Gráf, L., Jancso, A., Szilágyi, L., Hegyi, G., Pinter, K., Naray-Szabó, G., Hepp, J., Medzihradzsky, K. & Rutter, W.J. (1988) Electrostatic complementarity within the substrate-binding pocket of trypsin. *Proc. Natl. Acad. Sci. USA* **85**, 4961–4965. [PMID:3134655](#)

Guy, O., Lombardo, D., Bartelt, D.C., Amic, J. & Figarella, C. (1978) Two human trypsinogens. Purification, molecular properties, and N-terminal sequences. *Biochemistry* **17**, 1669–1675. [PMID:656395](#)

Halfon, S. & Craik, C.S. (1996) Regulation of proteolytic activity by engineered metal binding loops. *J. Am. Chem. Soc.* **118**, 1227–1228.

Harris, J.L., Backes, B.J., Leonetti, F., Mahrus, S., Ellman, J.A. & Craik, C.S. (2000) Rapid and general profiling of protease specificity by using combinatorial fluorogenic substrate libraries. *Proc. Natl. Acad. Sci. USA* **97**, 7754–7759. [PMID:10869434](#)

Hedstrom, L. (1996) Trypsin: a case study in the structural determinants of enzyme specificity. *Biol. Chem.* **377**, 465–470. [PMID:8922280](#)

Hedstrom, L., Szilagy, L. & Rutter, W. (1992) Converting trypsin to chymotrypsin. *Science* **255**, 1249–1253. [PMID:1546324](#)

Hedstrom, L., Perona, J.J. & Rutter, W.J. (1994a) Converting trypsin to chymotrypsin: residue 172 is a substrate specificity determinant. *Biochemistry* **33**, 8757–8763. [PMID:8038165](#)

Hedstrom, L., Farr-Jones, S., Kettner, C.A. & Rutter, W.J. (1994b) Converting trypsin to chymotrypsin: ground-state binding does not determine substrate specificity. *Biochemistry*

**33**, 8764–8769. [PMID:8038166](#)

Hedstrom, L., Lin, T.Y. & Fast, W. (1996) Hydrophobic interactions control zymogen activation in the trypsin family of serine proteases. *Biochemistry* **35**, 4515–4523. [PMID:8605201](#)

Hewett-Emmett, D., Czelusniak, J. & Goodman, M. (1981) The evolutionary relationship of the enzymes involved in blood coagulation and hemostasis. *Ann. N.Y. Acad. Sci.* **370**, 511–527. [PMID:6455952](#)

Higaki, J.N. & Light, A. (1985) The identification of neotrypsinogens in samples of bovine trypsinogen. *Anal. Biochem.* **148**, 111–120. [PMID:4037294](#)

Higaki, J.N., Evnin, L.B. & Craik, C.S. (1989) Introduction of a cysteine protease active site into trypsin. *Biochemistry* **29**, 8582–8586.

Higaki, J.N., Haymore, B.L., Chen, S., Fletterick, R. & Craik, C.S. (1990) Regulation of serine protease activity by an engineered metal switch. *Biochemistry* **29**, 8582–8586. [PMID:2125468](#)

Honey, N.K., Sakaguchi, A.Y., Quinto, C., MacDonald, R.J., Bell, G.I., Craik, C.S., Rutter, W.J. & Naylor, S.L. (1984a) Chromosomal assignments of human genes for serine proteases trypsin, chymotrypsin B, and elastase. *Somat. Cell Mol. Genet.* **10**, 369–376. [PMID:6589790](#)

Honey, N.K., Sakaguchi, A.Y., Lalley, P.A., Quinto, C., MacDonald, R.J., Craik, C.S., Bell, G.I., Rutter, W.J. & Naylor, S.L. (1984b) Chromosomal assignments of genes for trypsin, chymotrypsin B, and elastase in mouse. *Somat. Cell Mol. Genet.* **10**, 77–83.

Huber, R., Kukla, D., Bode, W., Schwager, P., Bartels, K., Deisenhofer, J. & Steigemann, W. (1974) Structure of the complex formed by bovine trypsin and bovine pancreatic trypsin inhibitor. II. Crystallographic refinement at 1.9 Å resolution. *J. Mol. Biol.* **89**, 73–101. [PMID:4475115](#)

James, M.N.G. (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: *Proteolysis and Physiological Regulation* (Ribbons, D.W. & Brew, K., eds). New York: Academic Press, pp. 125–142.

Jameson, G.W., Roberts, D.V., Adams, R.W., Kyle, W.S.A. & Elmore, D.T. (1973) Determination of the operational molarity of solutions of bovine  $\alpha$ -chymotrypsin, trypsin, thrombin and factor Xa by spectrofluorimetric titration. *Biochem. J.* **131**, 107–117. [PMID:4737291](#)

Keller, P., Cohen, E. & Neurath, H. (1958) The proteins of bovine pancreatic juice. *J. Biol. Chem.* **233**, 344–349.

Kossiakoff, A.A. & Spencer, S.A. (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**, 6462–6474. [PMID:7030393](#)

Kühne, W.F. (1876) Über die Verdauung der Eiweisstoffe durch den Pankreassaft. *Virchows Arch.* **39**, 130 (reproduced by Gutfreund, H. (1976) *FEBS Lett.* **62**(suppl.), E1–E12).

- Louvard, M.N. & Puigserver, A. (1974) On bovine and porcine anionic trypsinogens. *Biochim. Biophys. Acta* **371**, 177–185. [PMID:4429716](#)
- MacDonald, R.J., Stary, S.J. & Swift, G.H. (1982) Two similar but nonallelic rat pancreatic trypsinogens. Nucleotide sequences of the cloned cDNAs. *J. Biol. Chem.* **257**, 9724–9732. [PMID:6896710](#)
- McGrath, M.E., Wilke, M.E., Higaki, J.N., Craik, C.S. & Fletterick, R.J. (1989) Crystal structures of two engineered thiol trypsins. *Biochemistry* **28**, 9264–9270. [PMID:2611228](#)
- Neurath, H. & Zwillig, R. (1986) Willy Kühne und die Anfänge der Enzymologie. In: *Semper Apertus*. Berlin: Springer-Verlag, pp. 361–374.
- Northrop, J.H. & Kunitz, M. (1931) Isolation of protein crystals possessing tryptic activity. *Science* **73**, 262–263.
- Ohlsson, K. & Tegner, H. (1973) Experimental pancreatitis in the dog. Demonstration of trypsin in ascitic fluid, lymph and plasma. *Scand. J. Gastroenterol.* **8**, 129–133. [PMID:4697072](#)
- Perona, J.J. & Craik, C.S. (1995) Structural basis of substrate specificity in the serine proteases. *Protein Sci.* **4**, 337–360. [PMID:7795518](#)
- Perona, J.J., Tsu, C.A., Craik, C.S. & Fletterick, R.J. (1993) Crystal structures of rat anionic trypsin complexed with the protein inhibitors APPI and BPTI. *J. Mol. Biol.* **230**, 919–933. [PMID:7683059](#)
- Perona, J.J., Hedstrom, L., Wagner, R., Rutter, W., Craik, C.S. & Fletterick, R. (1994) Exogenous acetate reconstitutes the enzymatic activity of trypsin Asp189Ser. *Biochemistry* **33**, 3252–3259. [PMID:8136360](#)
- Perona, J.J., Hedstrom, L., Rutter, W.J. & Fletterick, R.J. (1995) Structural origins of substrate discrimination in trypsin and chymotrypsin. *Biochemistry* **34**, 1489–1499. [PMID:7849008](#)
- Read, R.J. & James, M.N.G. (1988) Refined crystal structure of *Streptomyces griseus* trypsin at 1.7 Å resolution. *J. Mol. Biol.* **200**, 523–551. [PMID:3135412](#)
- Rowen, L., Koop, B.F. & Hood, L. (1996) The complete 685-kilobase DNA sequence of the human  $\beta$ -T cell receptor locus. *Science* **272**, 1755–1762. [PMID:8650574](#)
- Rypniewski, W.R., Hastrup, S., Betzel, C., Dauter, M., Dauter, Z., Papendorf, G., Branner, S. & Wilson, K.S. (1993) The sequence and X-ray structure of the trypsin from *Fusarium oxysporum*. *Protein Eng.* **6**, 341–348. [PMID:8332590](#)
- Sakanari, J., Staunton, C.E., Eakin, A.E., Craik, C.S. & McKerrow, J.H. (1989) Serine proteases from nematode and protozoan parasites: Isolation of sequence homologous using generic molecular probes. *Proc. Natl. Acad. Sci. USA* **86**, 4863–4867. [PMID:2662185](#)
- Schellenberger, V., Turck, C.W. & Rutter, W.J. (1994) Role of the S' subsites in serine protease catalysis. Active-site mapping of rat chymotrypsin, rat trypsin,  $\alpha$ -lytic protease, and cercarial protease from *Schistosoma mansoni*. *Biochemistry* **33**, 4251–4257. [PMID:8155642](#)
- Sprang, S., Standing, T., Fletterick, R.J., Stroud, R.M., Finer-Moore, J., Xuong N.H.,

Hamlin, R., Rutter, W.J. & Craik, C.S. (1987) The three-dimensional structure of Asn102 mutant of trypsin: role of Asp102 in serine protease catalysis. *Science* **237**, 905–909. [PMID:3112942](#)

Stroud, R.M., Kay, L.M. & Dickerson, R.E. (1974) The structure of bovine trypsin: electron density maps of the inhibited enzyme at 5Å and at 2.7Å resolution. *J. Mol. Biol.* **83**, 185–208. [PMID:4821870](#)

Vasquez, J.R., Evin, L.B., Higaki, J.N. & Craik, C.S. (1989) An expression system for trypsin. *J. Cell Biochem.* **39**, 265–276. [PMID:2651464](#)

Wang, K., Gan, L. & Hood, L. (1995) Isolation and characterization of the chicken trypsinogen gene family. *Biochem. J.* **307**, 471–479.

Wang, E.C.W., Hung, S.-H., Cahoon, M. & Hedstrom, L. (1997) The role of the Cys191-Cys220 disulfide bond in trypsin: new targets for engineering substrate specificity. *Protein Eng.* **10**, 405411. [PMID:9194165](#)

Whitcomb, D.C., Gorry, M.C., Preston, R.A., Furey, W., Sossenheimer, M.J., Ulrich, C.D., Martin, S.P., Gates, Jr., L.K., Amann, S.T., Toskes, P.P., Liddle, R., McGrath, K., Uomo, G., Post, J.C. & Ehrlich, G.D. (1996) Hereditary pancreatitis is caused by a mutation in the cationic trypsinogen gene. *Nature Genet.* **14**, 141–145. [PMID:8841182](#)

Wiegand, U., Corbach, S., Minn, A., Kang, J. & Muller-Hill, B. (1993) Cloning of the cDNA encoding human brain trypsinogen and characterization of its product. *Gene* **136**, 167–175. [PMID:8294000](#)

Willett, W.S., Brinen, L.S., Fletterick, R.J. & Craik, C.S. (1996) Delocalizing trypsin specificity with metal activation. *Biochemistry* **35**, 5992–5998. [PMID:8634240](#)

Zimmerman, M., Ashe, B., Yurewicz, E.C. & Patel, G. (1977) Sensitive assays for trypsin, elastase, and chymotrypsin using new fluorogenic substrates. *Anal. Biochem.* **78**, 47–51. [PMID:848756](#)

---

**Sherin Halfon**

Millennium Pharmaceuticals, Inc., 256 East Grand Avenue, South San Francisco, CA 94080, USA

**Teaster T. Baird**

Department of Pharmaceutical Chemistry, University of California, San Francisco, San Francisco, CA 94143-0446, USA

**Charles S. Craik**

Department of Pharmaceutical Chemistry, University of California, San Francisco, San Francisco, CA 94143-2280, USA

Email: [craik@cgl.ucsf.edu](mailto:craik@cgl.ucsf.edu)

---

