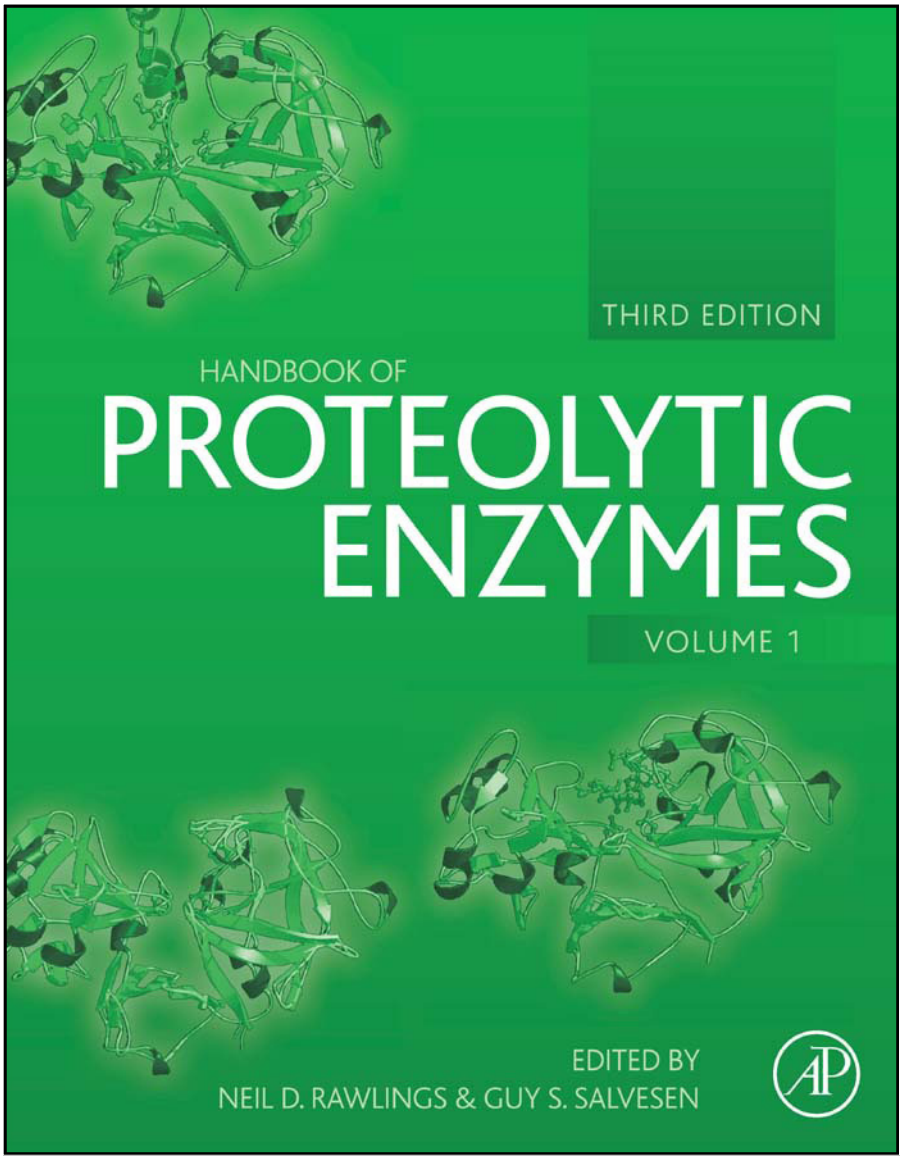

From Charles S. Craik, Trypsin. In: Neil D. Rawlings and Guy S. Salvesen, editors, Handbook of Proteolytic Enzymes. Oxford: Academic Press, 2013, pp. 2594-2600.
ISBN: 978-0-12-382219-2

Chapter 575

Trypsin

Databanks

MEROPS name: trypsin 1 (*Rattus*-type)
MEROPS classification: clan PA, subclan PA(S), family S1, subfamily S1A, peptidase S01.094
Tertiary structure: Available
Species distribution: known only from *Rattus norvegicus*
Reference sequence from: *Rattus norvegicus* (UniProt: P00762)

MEROPS name: trypsin-2 type C
MEROPS classification: clan PA, subclan PA(S), family S1, subfamily S1A, peptidase S01.119
Tertiary structure: Available
Species distribution: superorder Euarchontoglires
Reference sequence from: *Rattus norvegicus* (UniProt: P00763)

MEROPS name: trypsin-2 type B
MEROPS classification: clan PA, subclan PA(S), family S1, subfamily S1A, peptidase S01.120
Species distribution: class Mammalia
Reference sequence from: *Canis familiaris* (UniProt: P06872)

MEROPS name: trypsin 1
MEROPS classification: clan PA, subclan PA(S), family S1, subfamily S1A, peptidase S01.151
IUBMB: EC 3.4.21.4 (BRENDA)
Tertiary structure: Available
Species distribution: superclass Tetrapoda
Reference sequence from: *Bos taurus* (UniProt: P00760)

Name and History

Trypsin was first described and named in 1876 by Kühne [1] as the proteolytic activity in pancreatic secretions [2]. 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 that cleaved peptide bonds after Arg or Lys. The ready availability of trypsin from bovine pancreas allowed the enzyme to be purified by crystallization [3].

Activity and Specificity

Trypsin can be considered a prototype of the serine endopeptidases of family S1, as much of the fundamental understanding of the family has been derived from the study of this enzyme [4]. 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_{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 for different peptidic substrates [5]. However, discrimination between Arg and Lys for ester substrates is much less strict because the reactivity of the ester bond reduces the selectivity of the enzyme. Secondary binding sites on both sides of the scissile bond play only a minor role in the determination of substrate specificity, although occupancy of these sites does contribute to catalytic efficiency [6,7]. 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 [8] 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 [9]. 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 [10], and is one reason why trypsin is more promiscuous with ester substrates than with peptides. For protein substrates, the substrate binding may be the rate-limiting step.

Typical colorimetric assay substrates include Bz-Arg-OEt, Tos-Arg-OMe, Bz-Arg-*p*NA and Suc-Ala-Ala-*p*NA [11]. Pro-Arg Activity is also easily monitored fluorometrically with substrates containing an aminomethylcoumarin group or one of its derivatives, *e.g.* Z-Gly-Pro-Arg-AMC [11]. The Pro commonly used 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 [6], substrate cleavage being monitored by HPLC. A gel overlay assay has also been developed that can detect sub-nanogram amounts of trypsin [12]. Active-site titrations are usually accomplished with *p*-nitrophenyl-*p*′-guanidinobenzoate [13] or 4-methylumbelliferyl-*p*-guanidinobenzoate [14].

The pH optimum of trypsin is approximately 8, although this varies slightly with trypsin from different species. The reaction buffer requires moderate amounts (20 mM) of calcium ($CaCl_2$) 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^{-1} μM^{-1}, whereas the value for the smaller substrate, Z-Gly-Pro-Arg-AMC, is 210 min^{-1} μM^{-1} [6].

Trypsin is stable for extended periods of time as a lyophilized powder, or in solution at pH 3 at which pH the enzyme is largely inactive. The enzyme can be reversibly denatured by urea indicating the enzyme can spontaneously refold without assistance of chaperones [15]. Trypsin from natural sources is available from many commercial suppliers at various degrees of purity. Tos-Phe-CH_2Cl-treatment of the enzyme minimizes contamination by active chymotrypsin. It is also available from Sigma bound to agarose and to polyacrylamide for easy separation from tryptic digests. Recombinant forms of trypsin have also been made to reagent quantities and purity (see below).

Many general serine protease inhibitors (*i.e.* phenylmethanesulfonyl fluoride, diisopropylfluoro phosphate, and dichloroisocoumarin) inhibit trypsin, but a greater specificity for enzymes with trypsin-like specificity is shown by leupeptin, benzamidine, Tos-Lys-CH2Cl, and APMSF. Protein inhibitors include ecotin, soybean trypsin inhibitor, aprotinin, α_2-macroglobulin, and α_1-proteinase inhibitor.

Structural Chemistry

The three-dimensional structure of bovine trypsin was determined in 1974 independently by two groups [16,17] 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 [18,19].

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 bovine cationic and rat anionic trypsin backbones have a root-mean-square deviation (r.m.s.d.) of only 0.29 Å [20] and there is an r.m.s.d. 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. 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

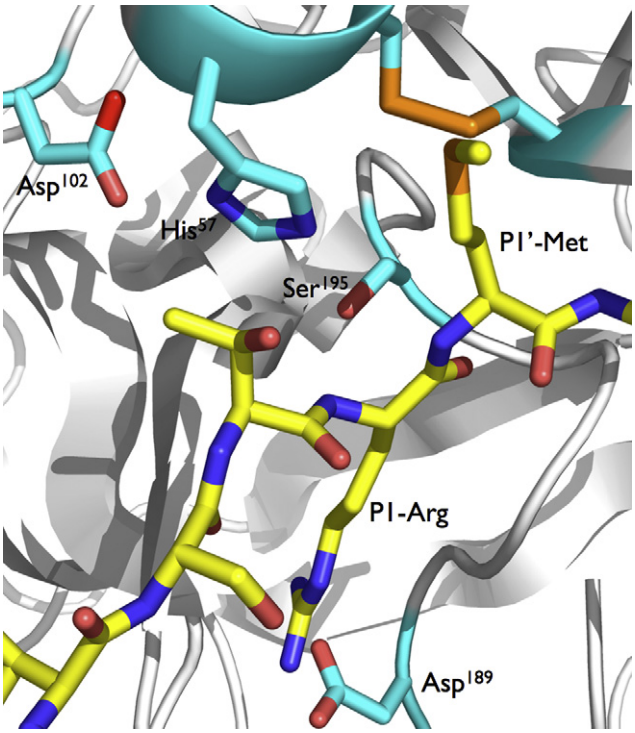

FIGURE 575.1 Active site of trypsin complexed with ecotin. The amino acids of the enzyme that are shown (cyan) are the residues of the catalytic triad (Asp102, His57, Ser195), Asp189 at the base of the S1 sub-site which confers primary specificity, and Cys42 and Cys58 in the S1′ sub-site. In the macromolecular inhibitor, ecotin (yellow), position P1, the primary determinant of specificity, is occupied by an Arg residue, which interacts electrostatically with Asp189. The P1′ position is occupied by a Met residue and associates with the S1′ sub-site. This figure was created using MacPyMOL [67] and adapted from PDB ID 1EZS.

β-barrel domains. Other forms (termed α, γ and π), in which the polypeptide backbone has been clipped, also possess varying degrees of activity [15]. As in other serine proteases of the family, the most important catalytic residues are those of the Asp/His/Ser triad (Figure 575.1). Ser195 acts as a nucleophile in the cleavage reaction, producing an acyl enzyme intermediate. Based on ^{15}N-NMR and crystallography the catalytic triad is no longer thought to act as a 'charge-relay system', but instead His57 is thought to act as a general base [21,22]. Asp102 is believed to stabilize the correct tautomer of His57, and to neutralize the developing positive charge during the catalytic reaction. Replacement of this aspartic acid with Asn results in an enzyme that is approximately 10^4-fold less active than the wild-type enzyme [23], while relocation of Asp102 to position 214 yields a protease that retains approximately 0.5% of the wild-type activity on peptide substrates. Replacement of either His57 or Ser195 with Ala results in a protease that is 10^5- to 10^6-fold less active than the wild-type enzyme [24]. The residual activity in the His57Ala and Ser195Ala variants is believed to be due to other structural features in the protease that help to stabilize the tetrahedral intermediate.

The role of key amino acid residues in catalysis and substrate specificity has been addressed through site-directed mutagenesis studies [6,20,24,25]. While replacement of the active-site His57 with Ala reduced the activity of the enzyme 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 [26]. In addition, to this obvious determinant of catalysis and engineered specificity, numerous sites were shown to contribute to inherent substrate specific recognition [5,27–30]. Subsequent studies sought to alter the activity and substrate specificity of the enzyme using metal ions as cofactors. For example, His substrate specificity was engineered into the sub-sites of trypsin by creating metal-binding sites that bridge the substrate and enzyme [31]. Engineered metal-binding sites were also shown to be useful in the reversible regulation of trypsin activity: several variants that involve the active site His in metal binding were effective at allowing the reversible inhibition of trypsin with sub-micromolar concentrations of transition metal ions [32–34].

All naturally occurring trypsins are synthesized as pro-enzymes. The mammalian propeptide (usually a hexapeptide) contains the consensus sequence for cleavage by enteropeptidase, -(Asp)$_4$-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 [19,35]. Stabilization of these new conformations in the so-called activation domain is due principally to hydrophobic interactions of the Ile16 side chain [36]. 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 β-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_{cat}/K_m for Arg/Lys substrates (with most of the decrease coming from a lowered k_{cat}), and no significant activity towards substrates with hydrophobic side chains [27,37] indicating that this residue is critical for recognition and proper positioning of the substrate. Removal of the side chain at position 189 in the Asp189Gly variant permits binding of a well-ordered acetate ion in a similar position of the crystallized enzyme [38]. 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 non-covalently 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 in rat trypsin by introducing Ala at positions 216 and 226 (normally Gly). This in turn dramatically alters the Arg/Lys specificity of the resulting enzymes and represented the first example of an enzyme with a genetically engineered function [5]. 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 [39]. 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 [4,30].

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 [40]. The protein has six completely conserved disulfide bonds, at positions 15–145, 33–49, 117–118, 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 α-trypsin form, which retains some catalytic activity. This and other clipped forms are present in most preparations of trypsin [15]. 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

Multiple forms of trypsin are readily available from many sources. However, trypsin purified from animal sources is often contaminated with other pancreatic enzymes. Bovine, porcine and murine and pancreatic cationic and anionic trypsin is commercially available from Sigma, Boehringer Mannheim, Worthington, Fluka and others. Recombinant human and rat anionic trypsin has been expressed in many different systems, including *Escherichia coli* [12,41,42], *Saccharomyces cerevisiae* [28], and *Pichia pastoris* [34]. The mature form of the enzyme has been successfully expressed in *Escherichia coli* [12]. Purification from these systems often involves affinity chromatography on immobilized benzamidine, ecotin 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 [29,42]. Trypsin has also been displayed on bacteriophage, permitting an *in vitro* selection of trypsin activity [43].

Biological Aspects

Trypsin is one of several digestive enzymes secreted into the intestine of animals. It is found in all insects and animals. In bovine pancreatic secretions, it represents approximately 15% of the digestive enzymes [44]. Trypsin is synthesized as a pre-proenzyme 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, 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 and elastase, 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 [45], cow [46], dog [47] and rat [48]. 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 [49]. 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 that encode trypsinogen [50]. Approximately 2–5% of the total adult rat pancreatic mRNA encodes anionic trypsin [51]. There are eight human genes, of which five are transcribed to RNA, two are pseudogenes, and one is a relic gene [52]. The human trypsinogen genes are intercalated in two pieces within the human β-T cell receptor locus on chromosome 7q35 [52,53]. At the 5′-end of the locus lie the three non-functional trypsinogen genes, while other genes, including those coding for the known trypsinogen isoenzymes, are 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 [54] in the mouse on chromosome 6 [55] and have been identified in the chicken [56]. The rat genes corresponding to the predominant trypsin forms have been shown to contain four introns [50]. Although these introns are positioned at the same sites within the genes, the introns have no sequence similarity among the different family members. 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' [57,58].

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 [54]. 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 non-pancreatic tissues on the basis of PCR analysis [56,59].

Distinguishing Features

Trypsin present in higher animals has very little sequence identity with microbial trypsins, but all trypsins are similar structurally. In particular, *Streptomyces griseus* trypsin has a very similar fold to these proteases [60]. Other microbial proteases are more distantly related, having shorter amino acid sequences and corresponding surface loops. For instance, α-lytic protease has the same fold as trypsin, but differs greatly in many structural aspects [61]. Many other serine proteases such as the kallikreins, elastase and chymotrypsin comprise the trypsin family and these are very similar structurally and mechanistically to trypsin, but differ in other key aspects such as substrate preferences and overall biological function.

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. This active site conservation permits 'consensus sequence cloning' where degenerate oligonucleotide probes are designed for the sequences flanking the active site Ser and His residues, and used in in a polymerase chain reaction of genetic material to isolate and identify trypsins and related serine proteases from organisms of diverse phyla [62].

Related Enzymes

Trypsin-like enzymes are present in many different organisms. Bacterial trypsins, such as those from *Streptomyces* species are dealt with elsewhere in this volume. The trypsin from the fungus *Fusarium oxysporum* is approximately 45% identical to both the bacterial and mammalian trypsins, and all trypsins are thought to have evolved from a common ancestor [63]. While functional parts of the proteins are strongly conserved, other regions are less so [64]. 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.

Acknowledgments

This work was supported by grants from the NSF MCB-0643988 to TTB and the NIH CA128765 to CSC.

Further Reading

The reader is referred to the articles of Perona & Craik [4] and Hedstrom [65,66].

References

[1] Kühne, W. (1876). Über die Verdauung der Eiwesstoffe durch den Pankeassaft. *Virchows Arch.* 39, 130.
[2] Neurath, H. (1986). Willy Kühne und die Anfänge der Enzymologie. Semper Apertus, Berlin: Springer-Verlag.
[3] Northrop, J.H., Kunitz, M. (1931). Isolation of protein crystals possessing tryptic activity. *Science* 73, 262–263.
[4] Perona, J.J., Craik, C.S. (1995). Structural basis of substrate specificity in the serine proteases. *Protein Sci.* 4, 337–360.
[5] Craik, C.S., Largman, C., Fletcher, T., Roczniak, S., Barr, P.J., Fletterick, R., Rutter, W.J. (1985). Redesigning trypsin: alteration of substrate specificity. *Science* 228, 291–297.
[6] Corey, D.R., McGrath, M.E., Vásquez, 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.
[7] 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.
[8] 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. *P. Natl. Acad. Sci. USA* 97, 7754–7759.
[9] Baird, T., Wang, B., Lodder, M., Hecht, S.M. (2000). Generation of active trypsin by chemical cleavage. *Tetrahedron* 56, 9477–9485.
[10] Hedstrom, L., Farr-Jones, S., Kettner, C.A., Rutter, W.J. (1994). Converting trypsin to chymotrypsin: ground-state binding does not determine substrate specificity. *Biochemistry* 33, 8764–8769.
[11] 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.
[12] Vásquez, J., Evnin, L., Higaki, J. (1989). An expression system for trypsin. *J. Cell Biochem.* 39, 265–276.
[13] Chase, T., Shaw, E. (1967). *p*-Nitrophenyl-*p*′-guanidinobenzoate HCl: a new active site titrant for trypsin. *Biochem. Biophys. Res. Commun.* 29, 508–514.

[14] Jameson, G.W., Roberts, D.V., Adams, R.W., Kyle, W.S., 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.

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

[16] Huber, R., Kukla, D., Bode, W., Schwager, P. (1974). Structure of the complex formed by bovine trypsin and bovine pancreatic trypsin inhibitor* 1: II. Crystallographic refinement at 1.9 Å resolution. *J. Mol. Biol.* 89, 73–101.

[17] Stroud, R., Kay, L. (1974). The structure of bovine trypsin: Electron density maps of the inhibited enzyme at 5 Å and at 2.7 Å resolution. *J. Mol. Biol.* 83, 185–208.

[18] 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.

[19] 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 A resolution. *J. Mol. Biol.* 118, 99–112.

[20] 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.

[21] 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.

[22] Tsilikounas, E., Tao, T., Gutheil, W.G., Bachovchin, W.W. (1996). ^{15}N and ^{1}H NMR spectroscopy of the catalytic histidine in chloromethyl ketone-inhibited complexes of serine proteases. *Biochemistry* 35, 2437–2444.

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

[24] 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.

[25] 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.

[26] Corey, D.R., Willett, W., Coombs, G., Craik, C.S. (1995). Trypsin specificity increased through substrate-assisted catalysis. *Biochemistry* 34, 11521–11527.

[27] Gráf, L., Craik, C., Patthy, A., Roczniak, S., Fletterick, R.J., Rutter, W.J. (1987). Selective alteration of substrate specificity by replacement of aspartic acid-189 with lysine in the binding pocket of trypsin. *Biochemistry* 26, 2616–2623.

[28] Hedstrom, L., Szilágyi, L., Rutter, W.J. (1992). Converting trypsin to chymotrypsin: The role of surface loops. *Science* 255, 1249–1253.

[29] 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.

[30] Perona, J., Hedstrom, L., Rutter, W.J., Fletterick, R.J. (1995). Structural origins of substrate discrimination in trypsin and chymotrypsin. *Biochemistry* 34, 1489–1499.

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

[32] Higaki, J., 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.

[33] 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.

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

[35] Fehlhammer, H., Bode, W., Huber, R. (1977). Crystal structure of bovine trypsinogen at 1-8 Å resolution. II. Crystallographic refinement, refined crystal structure and comparison with bovine trypsin. *J. Mol. Biol.* 111, 415–438.

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

[37] Gráf, L., Jancsó, A., Szilágyi, L., Hegyi, G., Pintér, K., Náray-Szabó, G., Hepp, J., Medzihradszky, K., Rutter, W.J. (1988). Electrostatic complementarity within the substrate-binding pocket of trypsin. *Proc. Natl. Acad. Sci. USA* 85, 4961–4965.

[38] Perona, J.J., Hedstrom, L., Wagner, R.L., Rutter, W.J., Craik, C.S., Fletterick, R.J. (1994). Exogenous acetate reconstitutes the enzymatic activity of trypsin Asp189Ser. *Biochemistry* 33, 3252–3259.

[39] Wang, E.C., 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, 405–411.

[40] Bode, W., Schwager, P. (1975). The refined crystal structure of bovine [beta]-trypsin at 1.8 Å resolution. *J. Mol. Biol.* 98, 693–717.

[41] Higaki, J.N., Evnin, L.B., Craik, C.S. (1989). Introduction of a cysteine protease active site into trypsin. *Biochemistry* 28, 9256–9263.

[42] Evnin, L.B., Vásquez, J.R., Craik, C.S. (1990). Substrate specificity of trypsin investigated by using a genetic selection. *P. Natl Acad Sci USA* 87, 6659–6663.

[43] 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.

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

[45] 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.

[46] Louvard, M., Puigserver, A. (1974). On bovine and porcine anionic trypsinogens. *Biochimica et Biophysica Acta (BBA) – Protein Structure* 371, 177–185.

[47] 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.

[48] 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.

[49] 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.

[50] 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.
[51] MacDonald, R., 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.
[52] 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.
[53] Honey, N.K., Sakaguchi, A.Y., Quinto, C., MacDonald, R.J., Bell, G.I., Craik, C., Rutter, W.J., Naylor, S.L. (1984). Chromosomal assignments of human genes for serine proteases trypsin, chymotrypsin B, and elastase. *Somat. Cell Mol. Genet.* 10, 369–376.
[54] Whitcomb, D.C., Gorry, M.C., Preston, R.A., Furey, W., Sossenheimer, M.J., Ulrich, C.D., Martin, S.P., Gates, L.K. Jr., 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.
[55] Honey, N.K., Sakaguchi, A.Y., Lalley, P.A., Quinto, C., MacDonald, R.J., Craik, C., Bell, G.I., Rutter, W.J., Naylor, S.L. (1984). Chromosomal assignments of genes for trypsin, chymotrypsin B, and elastase in mouse. *Somat. Cell Mol. Genet.* 10, 377–383.
[56] Wang, K., Gan, L., Lee, I., Hood, L. (1995). Isolation and characterization of the chicken trypsinogen gene family. *Biochem. J.* 307, 471–479.
[57] Craik, C.S., Rutter, W.J., Fletterick, R. (1983). Splice junctions: association with variation in protein structure. *Science* 220, 1125–1129.
[58] Craik, C., Sprang, S., Fletterick, R., Rutter, W.J. (1982). Intron–exon splice junctions map at protein surfaces. *Nature* 299, 180–182.
[59] Wiegand, U., Corbach, S., Minn, A., Kang, J., Müller-Hill, B. (1993). Cloning of the cDNA encoding human brain trypsinogen and characterization of its product. *Gene* 136, 167–175.
[60] Read, R.J., James, M.N. (1988). Refined crystal structure of Streptomyces griseus trypsin at 1.7 Å resolution. *J. Mol. Biol.* 200, 523–551.
[61] 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, J., eds., New York: Academic Press, pp. 125–142.
[62] Sakanari, J.A., Staunton, C.E., Eakin, A.E., Craik, C.S., McKerrow, J.H. (1989). Serine proteases from nematode and protozoan parasites: isolation of sequence homologs using generic molecular probes. *P. Natl. Acad. Sci. USA* 86, 4863–4867.
[63] Hewett-Emmett, D., Czelusniak, J., Goodman, M. (1981). The evolutionary relationships of the enzymes involved in blood coagulation and hemostasis. *Ann. N.Y. Acad. Sci.* 370, 511–527.
[64] 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.
[65] Hedstrom, L. (1996). Trypsin: a case study in the structural determinants of enzyme specificity. *Biol. Chem.* 377, 465–470.
[66] Hedstrom, L. (2002). Serine protease mechanism and specificity. *Chem. Rev.* 102, 4501–4524.
[67] The PyMOL Molecular Graphics System, Version 1.4, Schrödinger, LLC.

Teaster T. Baird, Jr.

Department of Chemistry and Biochemistry, San Francisco State University, 1600 Holloway Avenue, San Francisco, CA 94132, USA.
Email: tbaird@sfsu.edu

Charles S. Craik

Department of Pharmaceutical Chemistry, University of California, San Francisco, 600 16th Street, S512C, Box 2280, San Francisco, CA 94158, USA. Email: craik@cgl.ucsf.edu

Handbook of Proteolytic Enzymes, 3rd Edn
ISBN: 978-0-12-382219-2

DOI: http://dx.doi.org/10.1016/B978-0-12-382219-2.00575-5