Available online at www.sciencedirect.com
SCIENCE DIRECT®
Analytical Biochemistry 326 (2004) 190–199
ANALYTICAL BIOCHEMISTRY
www.elsevier.com/locate/yabio

Mutant rat trypsin selectively cleaves tyrosyl peptide bonds

Gábor Pál,[a] András Patthy,[b] József Antal,[c] and László Gráf[a,b,*]

[a] *Department of Biochemistry, Eötvös Loránd University, Pázmány St. 1/c, 1117 Budapest, Hungary*
[b] *Biotechnology Research Group of the Hungarian Academy of Sciences, Pázmány St. 1/c, 1117 Budapest, Hungary*
[c] *Agricultural Biotechnology Center, P.O. Box 170, 2100 Gödöllő, Hungary*

Received 22 September 2003

Abstract

A double mutant of rat trypsinogen (Asp189Ser, ΔAsp223) was constructed by site-directed mutagenesis. The recombinant protein was produced in *Escherichia coli* under the control of a periplasmic expression vector. The purified and enterokinase-activated enzyme was characterized by synthetic fluorogenic tetrapeptide and natural polypeptide substrates and by a recently developed method. In case of this latter method the specificity profile of the enzyme was examined by simultaneous digestion of a mixture of oligopeptide substrates each differing only at the P_1 site residue, and the results were analyzed by high-performance liquid chromatography. All these assays unanimously demonstrated that the recombinant proteinase lacks trypsin-like activity but acquired a rather unique selectivity: it preferentially hydrolyses peptide bonds C-terminal to tyrosyl residues. This narrow specificity should be useful in peptide-analytical applications such as sequence-specific fragmentation of large proteins prior to sequencing.

Keywords: Tyrosyl peptide bond cleavage; Peptide sequencing; Limited proteolysis

Although nowadays the complete primary structure of a protein is typically obtained by deducing it from the sequence of the encoding gene rather than sequencing the entire protein, direct sequencing of polypeptides remains an important tool. For example, it is still essential to provide partial sequence information for the design of oligonucleotide probes that can be used to screen genomic or cDNA libraries when searching for the encoding gene. Direct peptide sequencing is also a very important and powerful technique to localize post-translational modification sites, analyze partially digested products of proteins to localize potential domain borders, or determine connectivity patterns of cysteine residues in disulfide-containing proteins. However, the longer the polypeptide chain the less feasible is direct sequencing of the entire molecule.

There are at least four chemical procedures and about 10 proteinases commonly used for sequence-specific fragmentation of the polypeptide chain [1]. With chemical methods, cleavage of the peptide bond at Met-X (with CNBr), Trp-X (with BNPS-skatole), Asp-Pro (with mild acid treatment), and Asn-Gly (with hydroxylamine) can be achieved.

Several proteinases catalyze more or less selective cleavages of the following peptide bonds: Arg-X and Lys-X (trypsin); Glu-X and Asp-X (V8 protease); Lys-X (endoproteinase Lys-C); Arg-X (endoproteinase Arg-C); X-Asp and X-Cys (endoproteinase Asp-N); X-Phe (thermolysin); Ala-X and Gly-X (elastase); Pro-X (prolyl endopeptidase); and Trp-X, Phe-X, Tyr-X, and Leu-X (chymotrypsin). The letter "X" denotes any amino-acyl residues, usually other than Pro.

The above selection might seem to be satisfyingly large, but in reality it is rather limited and biased. For example, there are no known methods or enzymes that would cleave at small hydrophilic residues such as Ser or Thr or could cleave peptide bonds exclusively at certain hydrophobic residues such as Val, Ile, Leu, Tyr, or Phe. Efforts to discover or engineer enzymes with new specificities would promise not only theoretical but also practical benefits, since peptide chemistry could substantially gain from these additional types of selectivity traits.

* Corresponding author. Fax: +36-1-381-2172.
E-mail address: graf@ludens.elte.hu (L. Gráf).

0003-2697/$ - see front matter © 2003 Elsevier Inc. All rights reserved.
doi:10.1016/j.ab.2003.12.001

The majority of the enzymes mentioned above (trypsin, chymotrypsin, elastase, V8 protease, prolyl endopeptidase, and endoproteinase Lys-C) belong to the serine proteinase family. Since this group of proteinases is perhaps the most thoroughly characterized one by means of structural, functional, and mutagenesis studies, serine proteinases might represent the most promising targets to design more selective enzymes by protein engineering.

On the way to deciphering the structural backgrounds that ensure characteristically different substrate specificities for trypsin and chymotrypsin, large numbers of trypsin mutants have been constructed and analyzed in our laboratory. These variants are routinely tested in several assays to determine their respective catalytic parameters and substrate specificity traits. One of these mutants stood out with a remarkably narrow specificity toward the cleavage of tyrosyl peptide bonds. In this article we describe a number of experiments that illustrate this unique selectivity and discuss the potentials of the enzyme as an analytical tool for peptide and protein chemists.

Materials and methods

Bacterial strains, vectors, enzymes, and chemicals

Escherichia coli strains XL1 Blue and CJ236 (dut^-, ung^-) were purchased from Stratagene and Bio-Rad, respectively. Fractogel TSK-DEAE 650 (S) and TSK-CM 650 (S) were obtained from Merck. CNBr-activated Sepharose 4B, T7 sequencing Kit, and reagents for oligonucleotide synthesis were purchased from Pharmacia. DNA-modifying enzymes were obtained from Amersham, Boehringer, and Promega while GeneClean was from Bio101. The fluorescence turnover substrates with aminomethylcoumarine (AMC)[1] leaving group, *N*-succinyl-Ala-Ala-Pro-Tyr-AMC, *N*-succinyl-Ala-Ala-Pro-Phe-AMC, *N*-succinyl-Ala-Ala-Pro-Trp-AMC, *N*-succinyl-Ala-Ala-Pro-Leu-AMC, *N*-succinyl-Ala-Ala-Pro-Lys-AMC, and *N*-succinyl-Ala-Ala-Pro-Arg-AMC, were synthesized as described previously [2]. High-purity bovine enzymes, trypsin (Tosyl Phenylalanyl Chloromethyl Ketone (TPCK) treated) and α-chymotrypsin (Tosyl Lysyl Chloromethyl Ketone (TLCK) treated), horse cytochrome *c*, and 4-vinylpyridine were purchased from Sigma. Synthetic human β-endorphin was a gift from the late Professor Choh Hao Li (University of California, San Francisco) [3]. Recombinant desulfatohirudin was from the Institute for Drug Research (Budapest, Hungary). All other chemicals were reagent grade or the best commercially available.

[1] *Abbreviations used:* AMC, aminomethylcoumarine; Mes, 2-(N-Morpholino)ethanesulfonic acid; TFA, trifluoroacetic acid.

Site-directed mutagenesis

The standard techniques of molecular cloning were performed essentially as described [4,5]. The oligonucleotide used as primer for the mutagenesis reaction had the following sequence: 5′-GCCCTGCCAAACCCT GGTGTGTAC-3′. It was synthesized on an Applied Biosystems automated DNA synthesizer. The oligonucleotide was designed such that it deleted the triplet of Asp223 in the rat trypsinogen gene. The gene of the Asp189Ser mutant rat trypsinogen [2] was subcloned into the replicative form of the M13 mp18 vector between *Hin*dIII and *Sac*I restriction sites. Uracil containing single-stranded template DNA for in vitro mutagenesis was prepared by using the CJ236 (dut^-, ung^-) *E. coli* strain. The mutated strand was selected for in XL1 Blue (dut^+, ung^+) cells as described by Kunkel [6]. The mutant gene was then subcloned into the expression vector, pTrap [7], and sequenced before use. The clone is available on request.

Heterologous expression

In the expression vector, coding sequence of trypsinogen is fused to DNA sequences encoding the signal peptide [8] and regulatory regions of bacterial alkaline phosphatase as described [9]. The signal peptide of the alkaline phosphatase replaces the original signal peptide of pretrypsinogen and directs the recombinant protein into the periplasmic space. *E. coli* SM138 strain [7] was transformed with the expression vector. This strain ensures constitutive expression of the gene and secretes the zymogen into the periplasmic space.

Protein purification

The periplasmic protein fraction of *E. coli* was isolated by osmotic shock and dialyzed against 2.5 mM HCl. Precipitated proteins were removed by centrifugation (20,000*g*, 15 min) and the sample was dialyzed against 30 mM MeS, pH 6.0, 10 mM $CaCl_2$ buffer (buffer A). Precipitated proteins were removed as before and the supernatant was loaded onto a TSK-DEAE 650 (S) column (1 × 15 cm) equilibrated with buffer A. Proteins were eluted by linear salt gradient using 30 mM MeS, pH 6.0, 10 mM $CaCl_2$, 0.5 M NaCl solution (buffer B). Fractions were collected and checked on 15% SDS–PAGE [10] and Western blot.

The trypsinogen-containing fractions were pooled, dialyzed against 30 mM Na-citrate, 10 mM $CaC1_2$, pH 3.0, buffer, and loaded onto a TSK-CM 650 (S) column (1 × 15 cm) equilibrated with the same buffer. Proteins were eluted by NaCl gradient using 30 mM Na-citrate, 10 mM $CaC1_2$, 0.5 M NaCl, pH 3.0, solution (buffer B). The trypsinogen-containing fractions were identified using a combination of 15% SDS–PAGE and Western

blot analysis as above, pooled again, and dialyzed against 2.5 mM HCl. Trypsinogen was then activated by purified porcine enterokinase (1:50 wt/wt enterokinase/trypsinogen ratio) in 30 mM MeS, pH 6.0, 10 mM $CaCl_2$ buffer. The activated forms were separated by affinity chromatography on a column containing soybean trypsin inhibitor covalently immobilized to CNBr-activated Sepharose 4B matrix. Purity of the recombinant trypsin was verified on 15% SDS–PAGE. The concentration of the enzyme was determined by UV absorption ($\varepsilon_{280} = 38.000\ M^{-1}\ cm^{-1}$) and by the method of Bradford [11].

Characterization of the enzyme on synthetic fluorogenic substrates

Enzyme activity assays were carried out in 30 mM Tris–HCl, pH 8.0, 10 mM $CaCl_2$ buffer at 37 °C. The concentrations of fluorogenic substrates (Suc-Ala-Ala-Pro-Xaa-AMC) were 0.025, 0.05, 0.075, 0.1, 0.2, 0.4, and 0.6 mM. The excitation and emission wavelengths were 366 and 440 nm, respectively. Fluorescence intensities were converted to product concentrations by using standard solutions of AMC. The kinetic parameters K_m and k_{cat} were determined by nonlinear regression analysis of the data.

Characterization of the enzyme on natural polypeptide and protein substrates

A natural polypeptide (human β-endorphin) and two reduced and pyridylethylated [12] proteins (recombinant desulfatohirudin and horse cytochrome *c*) were used as substrates for the Asp189Ser, Δ223Asp mutant rat trypsin. From β-endorphin and desulfatohirudin 50 μg of the polypeptides was digested in 50-μl final volumes. From cytochrome *c* 500 μg was digested in a 500-μl final volume. The digestion reactions were carried out in 0.1 M ammonium bicarbonate, pH 8.0, 10 mM $CaCl_2$ buffer at room temperature. The enzyme/substrate ratio for the mutant enzyme was 1:10 (wt/wt). To get similar rates of substrate consumptions, 1:1000 (wt/wt) wild-type trypsin or chymotrypsin had to be added (data not shown). To search for the optimal digestion times, the reactions were terminated at various time points (5, 20, 60, and 240 min and 16 h overnight) by transferring 10 or 100-μl (for cytochrome *c*) aliquots into reaction tubes containing an equal volume of 5% TFA. Then 10-μl aliquots from β-endorphin and desulfatohirudin and 100-μl aliquots from cytochrome *c* of the TFA-treated digestion mixtures were loaded onto the Aquapore OD300 reverse-phase column (2.1 × 220 mm) of a 140B ABI HPLC system. For elution, acetonitrile gradient in 0.1% (v/v) aqueous TFA was applied. The detection wavelength was 220 nm. Digestion products were collected and identified from the best chromatographic separations corresponding to the optimal reaction times as indicated under Results.

Amino acid composition of the fragments was determined using a Waters workstation and a PicoTag HPLC system (Millipore Waters Chromatography Corp.). The peptides were hydrolyzed in 6 M HCl in the presence of crystalline phenol at 110 °C for 18 h, and their amino acid composition was analyzed after derivatization with phenylisothiocyanate [13]. Aliquots of the digestion mixtures and/or their HPLC-separated components were sequenced using a 471A pulsed-liquid-phase protein sequencer of ABI employing an Edman degradation program [14].

Characterization of Asp189Ser, Δ223Asp mutant rat trypsin by using competing oligopeptide substrates

Characterization of the mutant trypsin on competing oligopeptide substrates was done as described in detail [15]. Briefly, the enzyme reaction was monitored on a mixture of seven oligopeptide substrates that had the overall sequence His-Ala-Ala-Pro-Xxx-Ser-Ala-Asp-Ile-Gln-Ile-Asp-Ile. The individual peptides differed only at the potential P_1 site (according to the nomenclature of Schechter and Berger [16]) denoted as Xxx. In the individual peptides this position carried a residue Lys, Arg, Tyr, Leu, Phe, Trp, or Pro, respectively. An equimolar mixture of the seven components was made from 1 mg/ml stock solutions of the oligopeptides dissolved directly in the reaction buffer, 0.05 M Tris–HCl, pH 8.0, 18 mM $CaCl_2$. This mixture was digested in independent reactions with the Asp189Ser, Δ223Asp mutant rat trypsin, bovine trypsin, and bovine chymotrypsin. All the reactions were carried out in the above buffer at 25 °C. The enzyme concentrations were optimized to ensure comparable enzyme reaction rates with the individual proteinases. The concentration of the individual oligopeptide components was 40 μM and enzyme concentrations 0.0075 μM for bovine chymotrypsin, 0.0012 μM for bovine trypsin, and 0.15 μM for the Asp189Ser, Δ223Asp mutant rat trypsin were applied. The reaction volume was 700 μl. The digestion was terminated at different reaction times (0, 1, 4, 16, 32, 64 min) by transferring 100-μl aliquots to the injector tubes containing 20 μl of 5 M acetic acid. The aliquots were analyzed on reverse-phase HPLC according to the method described [15].

Results

Production of Asp189Ser, Δ223Asp mutant rat trypsinogen in E. coli

The codone for 223 Asp of the Asp189Ser mutant rat trypsinogen gene was deleted by site-directed

mutagenesis. The mutant trypsinogen was expressed in *E. coli* under the control of a bacterial alkaline phosphatase promoter. The recombinant protein was secreted into the periplasmic space of *E. coli* and purified to homogeneity. The purified proenzyme was activated by enterokinase and the activated form was isolated by affinity chromatography. Approximately 0.5 mg purified proteinase was isolated from a 1-liter culture of the bacteria.

Characterization of Asp189Ser, Δ223Asp mutant rat trypsin on synthetic substrates

Activity of the mutant enzyme was measured on synthetic fluorogenic tetrapeptide substrates Suc-Ala-Ala-Pro-Xaa-AMC, where the P_1 site Xaa residue was Lys, Arg, Trp, Phe, Leu, and Tyr, respectively (for details on the nomenclature of substrate (P) and enzyme (S) subsites see [16]). Kinetic parameters of the Asp189Ser, Δ223Asp trypsin measured on these substrates were compared to the corresponding values of three other related mutants (see details under Discussion). The k_{cat}, K_m, and catalytic efficiency (k_{cat}/K_m) values of these four mutants are summarized in Table 1. The catalytic efficiency of the Asp189Ser, Δ223Asp trypsin on the tetrapeptide substrate with P_1 Tyr is one or two orders of magnitude higher than on all the other substrates included in the test. This outlines a remarkable specificity profile that differs significantly from that of all the dozens of mutant serine proteinases that have so far been tested in our laboratory.

To highlight the striking tyrosyl selectivity of Asp189Ser, Δ223Asp trypsin we summarized its relative catalytic efficiency $(k_{cat}/K_m)_{Tyr}/(k_{cat}/K_m)_{Xaa}$ values in Table 2 and compared these values to the corresponding values of bovine α-chymotrypsin, wild-type rat trypsin, and three of its mutants.

The promising specificity profile of the Asp189Ser, Δ223Asp trypsin that we nicknamed "Tyrpsin" encouraged us to further investigate its apparent tyrosyl selectivity on larger substrates that offered more complete selection of the amino acid residues.

Characterization of Asp189Ser, Δ223Asp mutant rat trypsin on natural polypeptide substrates

Digestion of human β-endorphin. Human β-endorphin is a polypeptide composed of 31 amino acid residues [3,17] that represent 15 different types of side chains. Six positions are potential cleavage sites for chymotrypsin, while five positions should be cleaved with trypsin. The polypeptide contains 2 tyrosine residues; 1 is on the N terminus of the molecule: YGGFMTSEKSQTPLVTLFKNAIIKNAYKKGE. Potential cleavage sites for trypsin are denoted by white character and black background while cleavage sites for chymotrypsin are highlighted with a gray background. The tyrosyl residues are underlined.

Table 1
Amidolytic enzyme activities

Enzyme	Kinetic parameter	Tyr	Phe	Trp	Leu	Lys	Arg
Wild-type bovine chymotrypsin	k_{cat}	1.18×10^3	1.04×10^3	2.86×10^3	7.85×10^2	2.47	ND
	K_m	0.02	0.02	0.04	0.11	0.60	
	k_{cat}/K_m	5.13×10^4	4.72×10^4	7.74×10^4	7.14×10^3	4.122	
Wild-type rat trypsin[a]	k_{cat}	0.04	0.38	ND	0.01	7.27×10^3	5.46×10^3
	K_m	0.20	0.78		0.31	0.10	0.07
	k_{cat}/K_m	0.19	0.48		0.03	7.34×10^4	7.83×10^4
Asp189Ser rat trypsin[a]	k_{cat}	8.56	2.56	0.06	0.70	0.49	0.23
	K_m	0.82	0.69	0.26	1.70	0.50	0.17
	k_{cat}/K_m	10.45	3.71	0.23	0.41	0.97	1.36
Asp189Ser, TyrGly217-219Ser-GlyGly, Gln192Met rat trypsin[b]	k_{cat}	1.80	0.30	0.16	0.12	0.17	0.20
	K_m	0.19	0.23	0.14	0.23	0.18	0.18
	k_{cat}/K_m	9.73	1.28	1.15	0.53	0.96	1.14
Tr → Ch [S1 + L1 + L2] (loop mutant) rat trypsin[c]	k_{cat}	37.00	33.00	3.50	1.60	0.14	0.41
	K_m	0.67	1.52	0.42	0.77	1.00	0.34
	k_{cat}/K_m	55.56	21.78	8.41	2.08	0.14	1.22
Asp189Ser, Δ223Asp mutant rat trypsin, "Tyrpsin"	k_{cat}	4.80	0.24	0.01	0.13	0.24	0.10
	K_m	0.50	1.25	0.42	0.91	1.22	0.38
	k_{cat}/K_m	9.60	0.19	0.03	0.14	0.20	0.26

Kinetic parameters k_{cat} (min^{-1}), K_m (mM), and k_{cat}/K_m ($min^{-1}\,mM^{-1}$) were determined for four different mutants of recombinant rat trypsin using fluorogenic tetrapeptide substrates. The general structure of the substrate was Succinyl-Ala-Ala-Pro-Xaa-AMC, where Xaa denotes the P_1 site residue that was varied in the assay. The six different Xaa residues, Tyr, Phe, Trp, Leu, Lys, and Arg, are listed in the first row.

[a] Data from [2].

[b] The mutant was first reported in [30].

[c] The construct was published in [18]; data shown were determined in our laboratory. The mutants are described under Discussion.

Table 2
Relative catalytic activity (k_{cat}/K_m) values

Enzyme	$\frac{(k_{cat}/K_m)_{Tyr}}{(k_{cat}/K_m)_{Phe}}$	$\frac{(k_{cat}/K_m)_{Tyr}}{(k_{cat}/K_m)_{Trp}}$	$\frac{(k_{cat}/K_m)_{Tyr}}{(k_{cat}/K_m)_{Leu}}$	$\frac{(k_{cat}/K_m)_{Tyr}}{(k_{cat}/K_m)_{Lys}}$	$\frac{(k_{cat}/K_m)_{Tyr}}{(k_{cat}/K_m)_{Arg}}$
Bovine α-chymotrypsin	1.1	0.7	7.2	1.2×10^4	ND
Wild-type rat trypsin	0.4	ND	6.0	2.7×10^{-6}	2.5×10^{-6}
Asp189Ser rat trypsin	2.8	45.9	25.4	10.8	7.7
Asp189Ser,TyrGly217-219SerGlyGly, Gln192Met rat trypsin	7.6	8.5	18.4	10.2	8.5
Tr → Ch [S1 + L1 + L2] (loop mutant) rat trypsin	2.6	6.6	26.7	396.8	45.7
Asp189Ser, Δ223Asp mutant rat trypsin "Tyrpsin"	50.0	285.3	67.1	48.8	36.3

Relative catalytic activity (k_{cat}/K_m) values shown for two wild-type proteinases and four mutant rat trypsins. The data illustrate the relative efficiencies of the various enzymes on Tyr as a P_1 site residue compared to that of Phe, Trp, Leu, Lys, and Arg. The mutant enzymes are described in detail under Discussion.

According to the HPLC and sequencing data, after 5 min incubation almost 50% of the human β-endorphin is cleaved into two fragments by the mutant enzyme at 1:10 (wt/wt) enzyme substrate ratio (Fig. 1A). N-terminal sequencing revealed that the Tyr27-Lys28 bond was the only cleavage site for the mutant enzyme. No detectable cleavage occurred after the N-terminal tyrosyl residue. This was anticipated, since serine proteinases of the chymotrypsin family are usually quite ineffective for the removal of the N terminal amino acid residues. A longer, 3-h incubation resulted in an additional cleavage: the peptide bond was hydrolyzed after Phe 18. This cleavage, however, remained partial even after an overnight digestion (data not shown).

Digestion of reduced and pyridylethylated recombinant desulfatohirudin. Desulfatohirudin is a protein of 66 amino acid residues that, as a collection, represents 16 different types of potential P_1 site residues. Its sequence is labeled with highlighting as above for β-endorphin:
VVYTDCTESGQNLCLCEGSNVCGQGNKCILGSDGEKNQCV
TGEGTPKPQSHNDGDFEEIPEEYLQ.

Desulfatohirudin contains seven chymotryptic and two tryptic cleavage sites. Two of the chymotryptic sites are tyrosines at the 3rd and 63rd positions. Since desulfatohirudin is a very compact protein stabilized by three disulfide bridges, it is rather resistant to proteolytic attack in its native conformation by even the high-activity native enzymes, trypsin or chymotrypsin. Thus, as expected, the mutant rat trypsin failed to cleave it even during prolonged incubation (data not shown). However, reduction and alkylation transformed the hirudin molecule to a substrate. Overnight (16 h) incubation

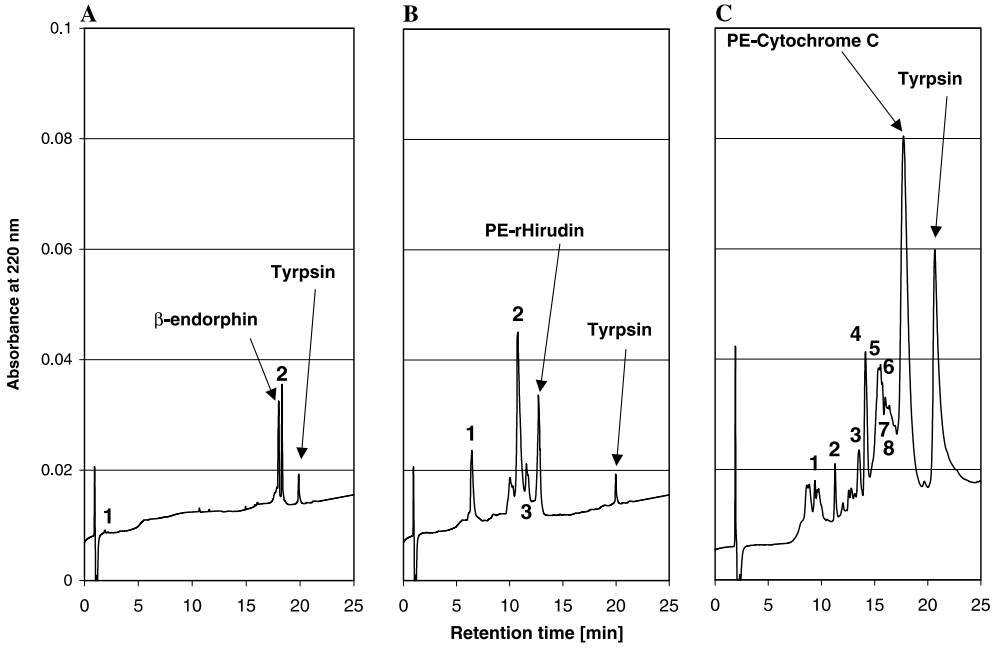

Fig. 1. Partial digestion of three polypeptide substrates was done by using the Asp189Ser, Δ223Asp double mutant of rat trypsin, "Tyrpsin." The substrate in (A) was β-endorphin, that in (B) was pyridylethylated (PE) recombinant desulfatohirudin, and that in (C) was PE horse cytochrome *c*. The sequence of the substrates is shown under Results. The proteolytic products were separated by HPLC and the cleavage products from the corresponding peaks were collected and analyzed by peptide sequencing. The peaks are labeled from A1 to C8 according to the substrate and the position of the peak in the chromatogram. The peptide content of the peaks is listed in Table 3. The label "1" in (A) points to the flow-through fraction. It contained a very hydrophilic cleavage product with an amino acid sequence of KKGE that was not retained by the column. Arrows point to the peak of the intact substrate and Tyrpsin in each inlet. Peaks without a label correspond to products that did not yield interpretable peptide sequence.

with the mutant enzyme resulted in 60% cleavage of hirudin, yielding specific proteolytic fragments. The reaction mixture and components from the separated peaks were sequenced and the cleavage sites were determined to be Tyr3-Thr4 and Tyr63-Leu64 (Fig. 1B and Table 3). The product of this latter cleavage, the Leu-Gln dipeptide, was identified only by pool sequencing of the reaction mixture, since it could not be detected by UV absorbance. No other peptide bonds were cleaved.

Digestion of horse cytochrome c. Horse cytochrome *c* is a protein of 104 amino acid residues with the sequence:
GDVEKGKKIFVQKCAQCHTVEKGGKHKTGPNLHGLFGRK
TGQAPGFTYTDANKNKGITWKEETLMEYLENPKKYIPGT
KMIFAGIKKKTEREDLIAYLKKATNE.

The cytochrome *c* sequence is shown with the same highlighting as used above for β-endorphin and desulfatohirudin. It contains four Tyr residues and altogether 21 tryptic and 15 chymotryptic cleavage sites. When the protein is digested with either chymotrypsin or trypsin, very large numbers of fragments that cannot be properly resolved by the reverse-phase HPLC system that we used are produced (data not shown). A 20 min partial digestion with the Asp189Ser, Δ223Asp trypsin, however, resulted in a few major products that were separated and their peptide content was determined by sequencing (Fig. 1C). The sequence analysis of the proteolysis products identified cleavages after each of the Tyr residues. However, one additional cleavage site at the Phe82-Ala83 peptide bond was also found (Table 3).

Characterization of Asp189Ser, Δ223Asp mutant rat trypsin using competing oligopeptide substrates

The Asp189Ser, Δ223Asp trypsin was analyzed using the method described [15]. An oligopeptide library of seven components each differing only at the P_1 cleavage site residue was digested with the mutant enzyme and the reaction was stopped at various times. The mixture containing the cleavage products and the remaining intact substrates was separated by reverse-phase HPLC. The activity of the enzyme on each individual substrate component was determined by measuring the decrease of the corresponding chromatographic peaks for each substrate. The results of such a kinetic measurement are illustrated in Fig. 2. Data derived from the assay are shown in Fig. 3 and compared to data obtained from control experiments using the same substrate mixture with bovine trypsin and chymotrypsin. For these wild-type proteinases the enzyme quantities were properly adjusted to achieve approximately the same degree of overall activity as that with the mutant enzyme (see Materials and methods). As readily visible in Fig. 2, the Asp189Ser, Δ223Asp trypsin selectively digests the Tyr-containing component. No cleavage of any of the other competing substrate components is visible even after a 1-h digestion period.

Table 3
Proteolytic fragments from three substrates

Substrate	Peak No.	Sequence of fragment
β-Endorphin	A/1	KKGE
	A/2	YGGFMTSEKSQTPLVTLFKNAIIKNAY
PE-rHirudin	B/1	VVY
	B/2	TDCTESGQNLCLCEGSNVCGQGNKCILGSDGEKNQCVTGEGTPKPQSHNDGDFEEIPEEY[a]
	B/3	TDCTESGQNLCLCEGSNVCGQGNKCILGSDGEKNQCVTGEGTPKPQSHNDGDFEEIPEEYLQ
		VVYTDCTESGQNLCLCEGSNVCGQGNKCILGSDGEKNQCVTGEGTPKPQSHNDGDFEEIPEEY
PE-cytochrome *c*	C/1	LKKATNE
	C/2	LENPKKY
	C/3	LENPKKYIPGTKMIF
	C/4	AGIKKKTEREDLIAYLKKATNE
	C/5	TDANKNKGITWKEETLMEY
	C/6	TDANKNKGITWKEETLMEYLENPKKY
	C/7	LENPKKYIPGTKMIFAGIKKKTEREDLIAY
		IPGTKMIFAGIKKKTEREDLIAYLKKATNE
	C/8	GDVEKGKKIFVQKCAQCHTVEKGGKHKTGPNLHGLFGRKTGQAPGFTY
		LENPKKYIPGTKMIFAGIKKKTEREDLIAYLKKATNE
		TDANKNKGITWKEETLMEYLENPKKYIPGTKMIF

Proteolytic fragments from three substrates: a peptide, β-endorphin (A), and two proteins, reduced and pyridylethylated recombinant desulfatohirudin (B) and horse cytochrome *c* (C). The fragments were produced by partial digestion of the polypeptide with the Asp189Ser, Δ223Asp mutant rat trypsin. The column labeled "Peak No." refers to HPLC peaks from Figs. 1A–C that contained the corresponding fragments. The complete sequence for he intact substrates is shown in the main text. The proteolytic fragments were identified by peptide sequencing. Since several peaks from C/5 to C/8 overlaps, these are cross-contaminated with each other. For clarity, each product is indicated only at the peak were it gave the highest N-terminal sequencing yield.

[a] The content of B2 is 20% contaminated with the two partially cleaved peptides from B/3.

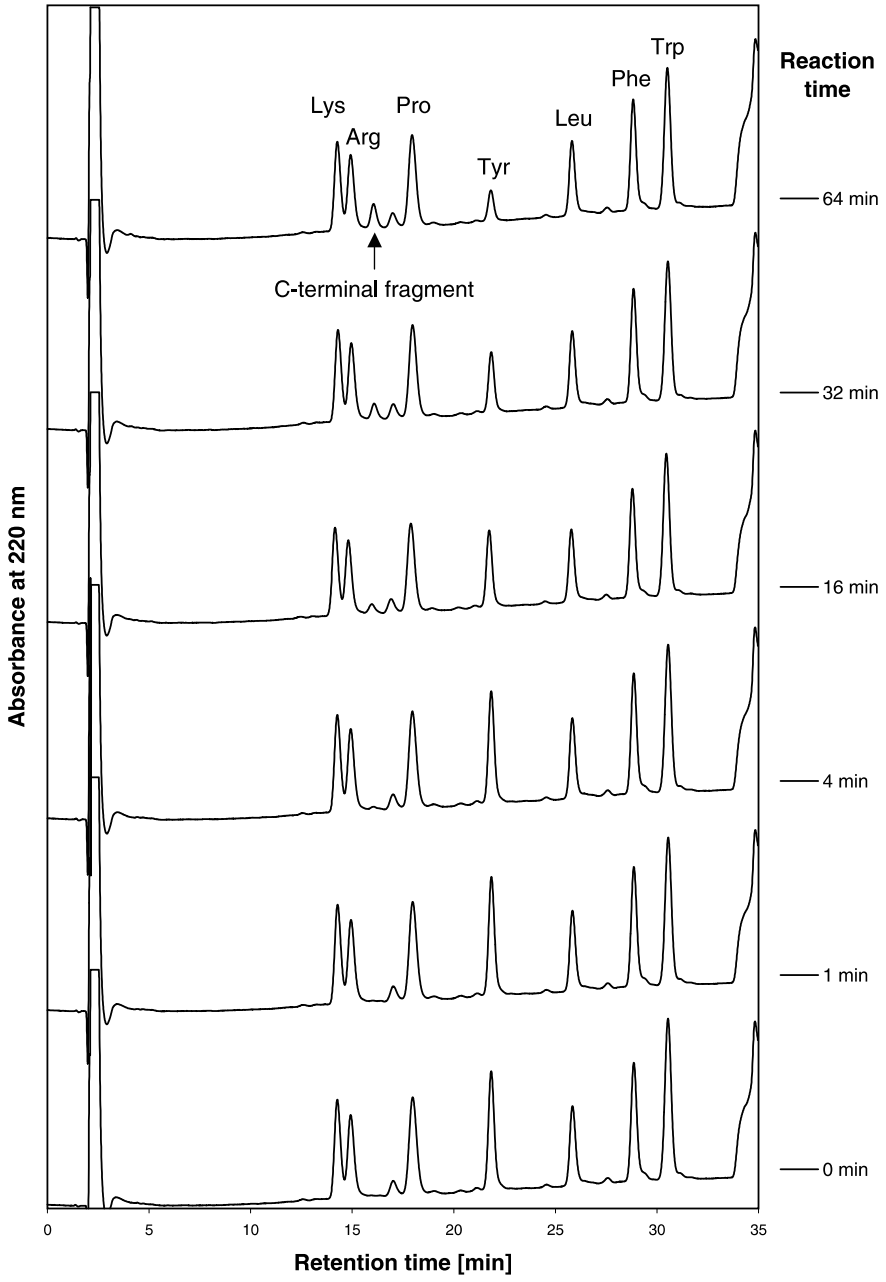

Fig. 2. RP-HPLC chromatogram set illustrating the time course of a digestion reaction. An oligopeptide mixture of seven peptides differing only at the P_1 cleavage site residue was digested with the Asp189Ser, Δ223Asp double mutant of rat trypsin. The overall sequence of the peptide library was His-Ala-Ala-Pro-Xxx-Ser-Ala-Asp-Ile-Gln-Ile-Asp-Ile, where Xxx denotes the potential P_1 cleavage site. In the individual peptides this position carried a Lys, Arg, Tyr, Leu, Phe, Trp, or Pro residue, respectively. The peaks corresponding to the individual substrate components are labeled at the top section according to the identity of the P_1 residue. At several time points (shown on the vertical axis) aliquots were taken out, their peptide components were separated on RP HPLC, and the corresponding chromatograms were compared. The relative activity of the enzyme on the individual substrates is determined by measuring the decrease of the area of the corresponding substrate component peaks. These values were normalized with data obtained on substrate with Pro as the P_1 residue, which cannot be cleaved by these enzymes. The normalized values for the 64-min time point are plotted in Fig. 3 along with analogous data obtained with two control enzymes, bovine trypsin and bovine chymotrypsin (chromatograms not shown). For a detailed description of the method see [15].

By that time, 75% of the Tyr substrate is already cleaved. In the same time frame bovine trypsin selectively cleaved 28% of the Lys- and 82% of the Arg-containing peptides, while bovine chymotrypsin cleaved 41% of the Tyr-, 5% of the Leu-, 29% of the Phe-, and 54% of the Trp-containing peptides (Fig. 3).

Discussion

Practical aspects

Routine kinetic studies using six different tetrapeptide substrates called our attention to Asp189Ser, Δ223Asp

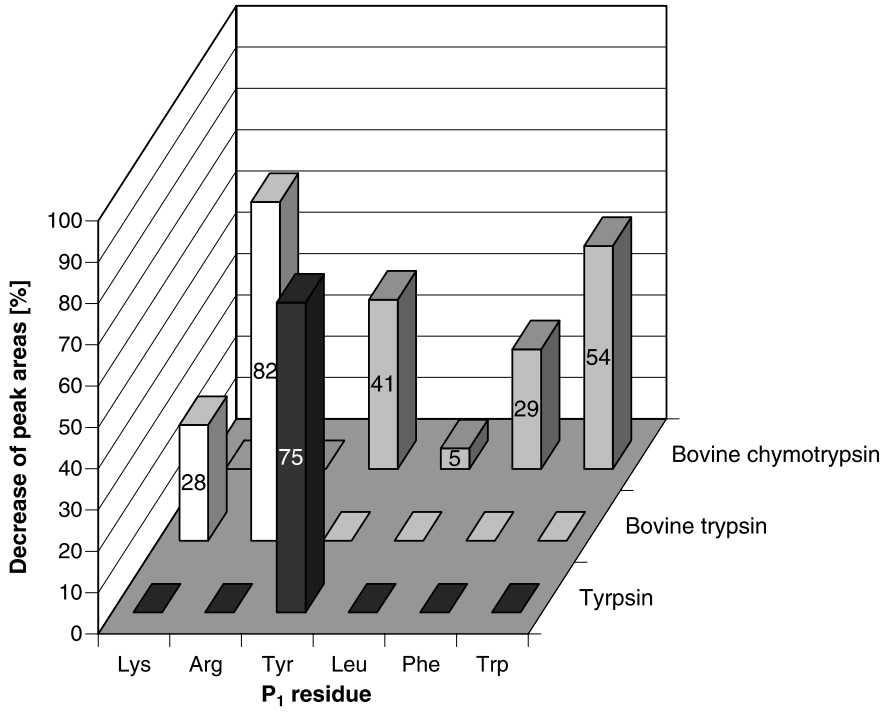

Fig. 3. Block diagram comparing the substrate specificities of three different proteinases, bovine chymotrypsin, bovine trypsin, and the Tyr-selective Asp189Ser, Δ223Asp double mutant of rat trypsin ("Tyrpsin"). Aliquots of the same oligopeptide mixture were digested in independent assays with the above three enzymes. Data sets corresponding to the 64-min time point of the cleavage reactions are arranged according to the three different enzymes as indicated on the right axis. The amount of the cleaved substrate is shown for each substrate–enzyme pair in percentage as seen on the left axis and indicated as numbers on the corresponding columns. For substrate–enzyme pairs where no number is shown, cleavage was not detected. The substrate mixture contained equimolar amounts of seven peptides differing only at the P_1 cleavage site residue. The overall sequence of the peptides is shown in the legend of Fig. 2. Results on the different substrates are organized according to the P_1 amino acid residues as labeled on the front axis. Data for Tyrpsin are derived from the chromatogram shown in Fig. 2. Similar chromatograms for the wild-type enzymes are shown in [15].

double mutant of rat trypsin. Although its catalytic efficiency was several orders of magnitude lower than that of the ancestor enzyme, the specificity profile was remarkable. Of the six different amino acid residues (Arg, Lys, Trp, Phe, Tyr, and Leu) at the P_1 site, tyrosine was preferred by the enzyme. As readily apparent from this comparison, none of the other enzymes has a similar high selectivity toward any of the residues.

As seen in Table 2, trypsin has a well-known and strict preference for Lys and Arg P_1 residues and has many orders of magnitude lower activity on substrates with hydrophobic P_1 sites. Nevertheless, this low activity is measurable and in this regard trypsin hardly distinguishes between Phe, Tyr, or Leu P_1 sites. Bovine α-chymotrypsin is selective for hydrophobic P_1 sites but does not distinguish between Tyr, Phe, or Trp and is only slightly better on these P_1 residues than on Leu.

The preference of Tyrpsin for the tyrosyl side chain was so significant that it urged us to start further investigation. The most important question was whether the mutant trypsin presents the same phenomena when a more complete selection of residues is supplied, as in the case of a polypeptide substrate.

Results on three polypeptides convinced us that, indeed, the double mutant trypsin is a quite selective proteinase that could be used as a tool for peptide chemical approaches such as sequence-specific proteolysis prior to peptide sequencing. The enzyme cleaved after all the Tyr residues of the polypeptides tested in our assays except in one case, when the residue was located at the N terminus of the peptide chain. Combining the results obtained for the three peptides, seven cleavages after Tyr residues were identified and only two other cleavages were observed, both after Phe residues.

Although a high relative amount of the mutant enzyme is needed to obtain efficient cleavage, the results are always easy to interpret since no autolysis of the proteinase was detected. Even under those conditions where considerable self-digestion might be expected (such as in the presence of denaturants) complete sequence information on the enzyme should prevent misinterpretation of the data. The relatively high enzyme/substrate ratio does not necessarily require large quantities of the recombinant proteinase. Today's advanced analytical and separation techniques allow for peptide sequencing in the picomolar range; therefore a couple of micrograms of the enzyme can be enough for a complete analysis. Hence, enzyme isolated from only several liters of easy to handle *E. coli* culture can supply proteinase that is enough for hundreds of experiments.

Tyr is not as common a residue as Ser, Leu, Lys, or Asp nor as rare as Met or Trp. Thus, a Tyr-selective enzyme might provide an optimal choice when the target peptide or protein does not contain Met and Trp residues but is digested to too-small fragments by enzymes that are less selective or that recognize more common amino acid residues.

An unexplored application of the enzyme could be to map modified Tyr residues. Since the modified Tyr is not expected to be a substrate for the enzyme, comprehensive analysis of the peptide or protein with and without modification might be able to locate the position of phosphorylation or iodination. This potential of the enzyme will be investigated in our lab.

Structural considerations

All the serine proteinases catalyze peptide or ester bond hydrolysis by essentially the same molecular mechanism. The catalysis requires proper positioning of the scissile bond of the substrate relative to the catalytic machinery. It is the substrate binding apparatus that provides this accurate positioning of the substrate and therefore it is the source of the selectivity. The structural basis of the substrate specificity of the serine proteinase families was discussed in details in two review papers [18,19] and more recently in the review article of Czapinska and Otlewski [20]. The four serine proteinases best-characterized by means of substrate specificity are perhaps BPN' subtilisin, α-lytic proteinase (a small chymotrypsin-like bacterial enzyme), trypsin, and chymotrypsin.

The originally broad P_1 specificity of BPN' subtilisin with a bias for large hydrophobic side chains was successfully altered by rational design to prefer smaller nonpolar residues [21] and to recognize dibasic [22] and tribasic [23] substrates. The changes were limited to the S_1 substrate binding pocket and altered only those residues in direct contact with the substrate. Importing the substrate specificity profile of SCARL subtilisin into the BPN' framework clearly demonstrated that residues outside the subtilisin S_1 site have minor influence on specificity [24]. Random mutagenesis of α-lytic proteinase also demonstrated that introducing changes exclusively into the substrate binding site can generate new forms with altered specificity profile [25].

Quite surprisingly, alteration of the specificity of trypsin turned out to be more complicated. Trypsin and chymotrypsin are favorite model proteins of enzymology and structural biology. The particular interest toward the chymotrypsin-like serine proteinases can partly be attributed to their fundamental biological functions. Another reason for continuous interest in these proteinases is that even several decades after the structure and possible mechanism of action of the catalytic triad has been revealed [26–28], the structural basis of their substrate-specific catalysis is still not fully understood.

Trypsin and chymotrypsin have specificity pockets of essentially the same geometry, yet trypsin is specific for positively charged residues and chymotrypsin for bulky hydrophobic residues at the P_1 site of the substrate. Thus, it has long been thought that the negative charge at site 189 is the major specificity determinant: its presence results in tryptic specificity while its lack results in chymotryptic specificity [29].

However, as was shown in the study of Graf et al. [2] replacement of the negatively charged Asp189 residue trypsin with the corresponding Ser189 residue of chymotrypsin resulted in a weak nonspecific enzyme. As illustrated in Table 2, the Asp189Ser mutant has broad substrate specificity with only a 10-fold preference for Tyr compared to the traditional trypsin substrate P_1 sites, Arg and Lys. Also, this enzyme hardly discriminates between Tyr and Phe P_1 sites. This observation initiated a long investigation to reveal the structural requirements for trypsin- and chymotrypsin-like specificities and provided unexpected results [2,30–34].

Even exchanging all those residues of the substrate binding pocket of trypsin with the chymotrypsin residues that directly contact the substrate did not import chymotrypsin-like activity into trypsin [30]. The corresponding Asp189Ser, TyrGly217-219SerGlyGly, Gln192Met trypsin mutant, which has a chymotrypsin-like substrate binding pocket, is rather nonspecific. While it is about 10-fold more efficient on Tyr than on any of the other P_1 residues included in this assay, it is equally good on the other three chymotrypsin and the two trypsin substrates (see Tables 1 and 2).

More distal elements, namely two surface loops, had to be exchanged to produce a hybrid trypsin with chymotrypsin-like specificity pattern and considerable activity [31]. The corresponding $Tr \rightarrow Ch$ $[S_1 + L_1 + L_2]$ (loop mutant) rat trypsin has sequences for the entire substrate binding pocket (S_1) and two surface loops (L_1 and L_2) replaced with homologous sequences from chymotrypsin. As illustrated in Table 2 this mutant trypsin has chymotrypsin-like specificity as it prefers hydrophobic P_1 residues over the positively charged trypsin substrates, Lys and Arg. The enzyme has comparable activities on Tyr, Phe, and Trp.

Loop 1 extends from residues 185 through 188 and Loop 2 from residues 221 through 225. While group-specific Loop 1 sequences are conserved in all the chymotrypsin and trypsin sequences, Loop 2 sequence is conserved only in chymotrypsins.

The rat trypsin sequence at the Loop 2 region is one residue longer than the corresponding chymotrypsin sequence. Deletion of Asp223 in the Asp189Ser mutant was meant to diminish this difference and thought to represent an alternative route of converting trypsin to a chymotrypsin-like proteinase. This mutation turned out to be successful by means of providing tyrosyl peptide bond selectivity for the otherwise nonspecific parent

molecule. In this regard it is interesting to note that the function of Loop 2 has been proposed to maintain the appropriate main chain conformation of the selectivity determinant G1y216 residue [34]. This residue ensures productive binding of the specific substrate at its P_3 site through an antiparallel β sheet. As we observed in the case of the β-endorphin, the Asp189Ser, Δ223Asp trypsin is unable to remove the N-terminal Tyr residue. In contrast, the enzyme efficiently cleaves internal tyrosyl bonds. This observation suggests that for the catalysis the mutant enzyme can productively use binding energy through the extended S_3–S_1 sites.

The effect of the Asp223 deletion on the conformation of Loop 2 and indirectly on the primary binding site is far from trivial. Analysis of the three-dimensional structure of the mutant enzyme in complex with a Tyr P_1 site inhibitor will be the next step on the way to elucidating the structural background of this unique selectivity.

References

[1] T.E. Creighton, Proteins: Structures and Molecular Properties, second ed., Freeman and Co, New York, NY, 1993.

[2] L. Graf, A. Jancso, L. Szilagyi, Gy. Hegyi, K. Pinter, G. Naray-Szabo, J. Hepp, K. Medzihradszky, W.J. Rutter, Electrostatic complementarity within the substrate-binding pocket of trypsin, Proc. Natl. Acad. Sci. USA 85 (1988) 4961–4965.

[3] C.H. Li, D. Yamashiro, L.-F. Tseng, H.H. Loh, Solid-phase synthesis of human beta-endorphin, J. Med. Chem. 20 (1979) 325–328.

[4] T. Maniatis, E.F. Fritsch, J. Sambrook, Molocular Cloning: A Laboratory Manual, Cold Spring Harbor Press, Cold Spring Harbor, NY, 1982.

[5] F.M. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.A. Smith, J.G. Seidman, K. Struhl, Current Protocols in Molecular Biology, Wiley, New York, NY, 1988.

[6] T.A. Kunkel, Rapid and efficient site-specific mutagenesis without phenotypic selection, Proc. Natl. Acad. Sci. USA 82 (1985) 488–492.

[7] L. Graf, C.S. Craik, A. Patthy, S. Roczniak, R.J. Fletterick, W.J. Rutter, Selective alteration of substrate specificity by replacement of aspartic acid-189 with lysine in the binding pocket of trypsin, Biochemistry 26 (9) (1987) 2616–2623.

[8] C.S. Craik, Q.L. Choo, G.H. Swift, C. Quinto, R.J. MacDonald, W.J. Rutter, Structure of two related rat pancreatic trypsin genes, J. Biol. Chem. 259 (1984) 14255–14264.

[9] H. Inouye, S. Michaelis, A. Wright, J. Beckwith, Cloning and restriction mapping of the alkaline phosphatase structural gene (phoA) of *Escherichia coli* and generation of deletion mutants in vitro, J. Bacteriol. 146 (1981) 668–675.

[10] U.K. Laemmli, Cleavage of structural proteins during the assembly of the head of Bacteriophage T4, Nature 227 (1970) 680–685.

[11] M.M. Bradford, A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein dye binding, Anal. Biochem. 72 (1976) 248–254.

[12] B.S. Hampton, D.R. Marshak, W.H. Burgess, Structural and functional characterization of full-length heparin-binding growth associated molecule, Mol. Biol. Cell. 3 (1992) 85–93.

[13] B.A. Bidlingmeyer, S.A. Cohen, T.L. Tarvin, PITC derivatives in amino acid analysis, J. Chromatogr. 336 (1984) 93–104.

[14] M.W. Hunkapiller, R.M. Hewick, W.J. Dreyer, L.E. Hood, High-sensitivity sequencing with a gas-phase sequenator, Methods Enzymol. 91 (1983) 399–413.

[15] J. Antal, G. Pal, B. Asboth, Z. Buzas, A. Patthy, L. Graf, Specificity assay of serine proteinases by reverse-phase high-performance liquid chromatography analysis of competing oligopeptide substrate library, Anal. Biochem. 288 (2) (2001) 156–167.

[16] I. Schechter, A. Berger, On the size of the active site in proteases. I. Papain, Biochem. Biophys. Res. Commun. 27 (2) (1967) 157–162.

[17] G. Cseh, E. Barat, A. Patthy, L. Graf, Studies on the primary structure of human beta-lipotropic hormone, FEBS Lett. 21 (1972) 344–346.

[18] J.J. Perona, C.S. Craik, Structural basis of substrate specificity in the serine proteases, Protein Sci. 4 (3) (1995) 337–360.

[19] L. Graf, L. Szilagyi, I. Venekei, Chymotrypsin, in: A.J. Barrett, F. Woessner, N. Rawlings (Eds.), Handbook of Proteolytic Enzymes, Academic Press, London, 1998, pp. 30–38.

[20] H. Czapinska, J. Otlewski, Structural and energetic determinants of the S_1-site specificity in serine proteases, Eur. J. Biochem. 260 (3) (1999) 571–595.

[21] J.A. Wells, D.B. Powers, R.R. Bott, T.P. Graycar, D.A. Estell, Designing substrate specificity by protein engineering of electrostatic interactions, Proc. Natl. Acad. Sci. USA 84 (5) (1987) 1219–1223.

[22] M.D. Ballinger, J. Tom, J.A. Wells, Designing subtilisin BPN' to cleave substrates containing dibasic residues, Biochemistry 34 (41) (1995) 13312–13319.

[23] M.D. Ballinger, J. Tom, J.A. Wells, Furilisin: a variant of subtilisin BPN' engineered for cleaving tribasic substrates, Biochemistry 35 (42) (1996) 13579–13585.

[24] J.A. Wells, B.C. Cunningham, T.P. Graycar, D.A. Estell, Recruitment of substrate-specificity properties from one enzyme into a related one by protein engineering, Proc. Natl. Acad. Sci. USA 84 (15) (1987) 5167–5171.

[25] L.D. Graham, K.D. Haggett, P.A. Jennings, D.S. Le Brocque, R.G. Whittaker, P.A. Schober, Random mutagenesis of the substrate-binding site of a serine protease can generate enzymes with increased activities and altered primary specificities, Biochemistry 32 (24) (1993) 6250–6258.

[26] A.A. Kossiakoff,, S.A. Spencer, 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 (1981) 6462–6474.

[27] L. Polgar, Mechanisms of Protease Action, CRC Press, Boca Raton, FL, 1989.

[28] S. Sprang, T. Standing, R.J. Fletterick, R.M. Stroud, J. Finer-Moore, N.H. Xuong, R. Hamlin, W.J. Rutter, C.S. Craik, The three-dimensional structure of Asn102 mutant of trypsin: role of Asp102 in serine protease catalysis, Science 237 (1987) 905–909.

[29] R.M. Stroud, A family of protein-cutting proteins, Sci. Am. 231 (1974) 74–88.

[30] L. Graf, L. Boldogh, L. Szilagyi, W.J. Rutter, Redesigning the substrate specificity of trypsin: can trypsin be converted to a chymotrypsin-like protease?, in: Z.H. Zaidi, A. Abbasi, D.L. Smith (Eds.), Protein Structure-Function, TWEL Publishers, Karachi-Doncaster-Gaithersburg, 1990, pp. 49–55.

[31] L. Hedstrom, L. Szilagyi, W.J. Rutter, Converting trypsin to chymotrypsin: the role of surface loops, Science 255 (1992) 1249–1253.

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

[33] L. Hedstrom, S. Farr-Jones, C.A. Kettner, W.J. Rutter, Converting trypsin to chymotrypsin: ground-state binding does not determine substrate specificity, Biochemistry 33 (1994) 8764–8769.

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