Cloning and Characterization of Complementary DNA for Human Tryptase

Jeffrey S. Miller, Eric H. Westin, and Lawrence B. Schwartz
Department of Medicine, Medical College of Virginia, Virginia Commonwealth University, Richmond, Virginia 23298

Abstract

The amino acid sequence of human mast cell tryptase was determined from corresponding cDNA cloned from a lambda ZAP library made with mRNA derived from a human mast cell preparation. Tryptase is the major neutral protease present in human mast cells and serves as a specific marker of mast cells by immunohistologic techniques and as a specific indicator of mast cell activation when detected in biologic fluids. Based on nucleic acid sequence, human tryptase consists of a 244-amino acid catalytic portion of 27,423 D with two putative *N*-linked carbohydrate binding sites and a 30-amino acid leader sequence of 3,048 D. A His_{74}, Asp_{120}, Ser_{223} catalytic triad and four cystine groups were identified by analogy to other serine proteases. Regions of amino acid sequence that are highly conserved in serine proteases, in general, were conserved in tryptase. The catalytic portion of human tryptase had an 84% amino acid sequence similarity with that of dog tryptase; their leader sequences had a 67% similarity. Asp_{217} in the substrate binding pocket of human tryptase is consistent with a specificity for Arg and Lys residues at the site of cleavage (P1), whereas Glu_{245} is consistent with the known preference of human tryptase for substrates with Arg or Lys also at P3, analagous residues also being present in dog tryptase. Asp_{244}, which is substituted for the Gly found in dog tryptase and in most serine proteases, is present in the putative substrate binding pocket and may confer additional substrate specificity on human tryptase for basic residues. Further studies now can be designed to elucidate these structure-function relationships.

Introduction

Human tryptase is a serine protease that is selectively concentrated in the secretory granules of human mast cells and is secreted upon the coupled activation-degranulation response of this cell type (1–3). Its selective presence in mast cells permitted its use as a specific clinical indicator of mast cell activation by measurements of its level in biologic fluids (4–7) and as a selective marker of intact mast cells using immunohistochemical techniques with antitryptase antibodies (8). It is selectively concentrated in both the MC_T[1] (10 pg/cell) (tryptase$^+$, chymase$^-$) and MC_{TC} (35 pg/cell) (tryptase$^+$, chymase$^+$) cell types (9), composing the major protein component of these mast cells (3). Human tryptase, like other mast cell proteases, is both stored and released in its catalytically active form; thus far no inactive proenzyme form has been detected. The active form of the enzyme purified from dispersed human mast cells is a tetramer, with four subunits of 31,000–33,000 D (2, 10, 11). Each subunit binds [^{3}H]difluorophosphate [DFP] and is recognized by the same murine antitryptase MAbs. NH_2-terminal amino acid sequence analysis obtained from a holoenzyme preparation containing all subunits revealed a single sequence from the NH_2-terminal end (12). Each subunit thus appears to be derived from the same or a highly homologous gene and size differences may be due to posttranslational modifications. The presence of carbohydrate on tryptase from human pituitary tissue has been inferred by a reduction in subunit size after treatment with *N*-endoglycosidase F (12).

Human tryptase is stabilized as a catalytically active tetramer by heparin, which also resides in secretory granules (13). This stabilizing activity occurs with heparin glycosaminoglycan chains of 6,000 D or greater. Stabilization is based upon negative charge density rather than carbohydrate composition and occurs with dextran sulfate, heparin, chondroitin sulfate E, and chondroitin sulfate A to successively lesser extents (14). Salt concentrations of 0.75 M NaCl are needed to dissociate tryptase from heparin. Incubation of tryptase alone in physiologic buffer leads to dissociation of its subunits and the formation of inactive monomers without any apparent autodegradation. Tryptase is not inhibited by a variety of biologic protease inhibitors; including human serpins, alpha 2 macroglobulin, and soy bean and lima bean trypsin inhibitors, as well as by plasma and urine (1, 11, 12, 14, 15). Its regulation thus may occur primarily through modulation of its interaction with heparin.

Although the biologic function of human tryptase is unknown, in vitro, the activation of C3 to C3a (16) and prostromelysin to stromelysin (16a) and the inactivation of fibrinogen as a clottable substrate for thrombin (17) are tryptase-catalyzed events of possible interest. Using synthetic substrates with tryptase preparations from lung and skin, tryptase was shown to have a particular affinity for dipeptide substrates with adjacent basic amino acid residues in the P_1 and P_2 positions and for tripeptide substrates with basic residues at P_1 and P_3 (18). In addition, dog tryptase increases the response to histamine of isolated guinea pig pulmonary parenchymal strips (19). Thus, the abundance and specificity of human tryptase, along with its prolonged presence after release, suggest an important functional role in both immediate-type and late mast cell–mediated reactions. This manuscript details the cloning and sequencing of human tryptase cDNA, correlates the deduced amino acid sequence with selective properties of human tryptase, and compares this sequence with that of tryptase cloned from dog mast cells (20) and with other serine proteases.

Address reprint requests to Dr. Lawrence B. Schwartz, Medical College of Virginia, Box 263, Richmond, VA 23298.

Received for publication 18 April 1989 and in revised form 23 June 1989.

1. *Abbreviations used in this paper:* MC_T, tryptase$^+$, chymase$^-$ mast cell; MC_{TC}, tryptase$^+$, chymase$^+$ mast cell.

J. Clin. Invest.

0021-9738/89/10/1188/08 $2.00
Volume 84, October 1989, 1188–1195

Methods

Materials. DNA restriction endonucleases were obtained from Bethesda Research Laboratories (Bethesda, MD). [^{32}P]Gamma dATP used for 5′ end labeling of synthetic oligonucleotides and [^{35}S]αdATP used for DNA sequencing were obtained from New England Nuclear (Boston, MA). Sequenase and corresponding nonradioactive nucleotides used for DNA sequencing were obtained from U.S. Biochemical Corp. (Cleveland, OH) and Klenow fragment from *Escherichia coli* polymerase was obtained from Amersham Corp. (Arlington Heights, IL). Acrylamide was purchased from Research Organics, Inc. and treated with activated charcoal (Sigma Chemical Co., St. Louis, MO) and mixed bed resin (AG501XB; Bio-Rad Laboratories, Richmond, CA) to remove impurities before use in DNA sequencing gels. Alkaline phosphatase–conjugated goat anti–mouse IgG antibodies (affinity purified) from Bio-Rad Laboratories, nitroblue tetrazolium and 5-bromo-4-chloro-3-indolyl phosphate from Sigma Chemical Co., and nitroplus 2000 nitrocellulose filters from Micron Separations, Inc. (Westboro, MA) were obtained as indicated. Purified tryptase and antitryptase MAbs were obtained as previously reported (2, 13). The NH_2-terminal portion of tryptase protein was sequenced on a protein sequencer (Model 470A; Applied Biosystems, Foster City, CA) and DNA sequencing was performed on a vertical sequencing apparatus (American Bionetics, Hayward, CA). Guanidine thiocyanate from Fluka AG (Buchs, Switzerland), Sarcosyl from Sigma Chemical Co., cesium chloride from Bio-Rad Laboratories, and oligo dT-cellulose from Collaborative Research Corp. (Lexington, MA) were obtained as indicated.

Mast cell and mRNA purification. A surgical specimen of human lung weighing 200 *g* was minced and digested with proteases to disperse cells as described (1), containing $\sim 100 \times 10^6$ mast cells of 12% purity. Approximately 45×10^6 mast cells were enriched to 30% by density-dependent centrifugation (room temperature, 300 *g*, 15 min) of cells on a 50% (vol/vol)/70%(vol/vol)/80%(vol/vol) discontinuous Percoll gradient in HBSS (minus calcium). MC_T cells accounted for $\sim$ 60% of the mast cells. Resultant washed cells were resuspended in 4.0 M guanidine thiocyanate, 10 mM Tris-HCl, pH 7.4 (8×10^6 cells/ml) and Dounce homogenized 10–15 times with a loose-fitting pestle to completely lyse cells and preserve mRNA (21, 22). Sarcosyl (20% wt/vol) was then added; the mixture was Dounce homogenized again and passed four times through a 23-gauge needle to shear nuclear DNA. The homogenate was then placed in 15 ml heat-sealed ultracentrifuge tubes, underlayered with 4.5 ml of 5.7 M cesium chloride (0.97 g/ml), and subjected to centrifugation at 41,000 rpm (Ti 70.1 fixed rotor) at 22°C for 20 h. RNA in the pellet was dissolved in RNase-free water, precipitated with 0.3 M sodium acetate, and 2 vol of ethanol at −20°C, collected by centrifugation, and resuspended in RNase-free water. In order to enrich for polyadenylated RNA, total RNA (560 μg) in 0.01 M Tris-HCl, pH 7.5, 1 mM EDTA, 0.5 M NaCl was gently mixed with oligo dT-cellulose (after fine particles of resin were removed by gravity sedimentation) for 15 min at room temperature (23). The resin was then gently washed twice with 0.5 M NaCl and then twice again with 0.1 M NaCl. Approximately 11 μg of bound RNA was then eluted with two washes of RNase-free water, precipitated with 2 vol of ethanol and 0.3 M sodium acetate at 20°C and shipped to Stratagene Corp. (La Jolla, CA) for preparation of the cDNA and insertion into the lambda ZAP vector at a unique Eco RI restriction site. The lambda ZAP vector is a modified lambda phage expression vector that contains unique cloning sites within the lac operon, which in turn is within a pBluescript plasmid. The pBluescript plasmid vector autoexcises in the presence of helper phage, which allows for simplified subcloning.

Immunoscreening of lambda ZAP cDNA library for expression of tryptase fusion protein. Approximately 100,000 plaques were screened after infection of Y1090 *E. coli* host cells with an amplified portion of the lambda ZAP library (24). Bacterial lawns were first grown on petri plates at 42°C for 3 h, which suppressed lysis by a temperature-sensitive suppressor. Nitrocellulose filters soaked in 10 mM isopropythiogalactoside (IPTG) were overlaid and the plates were placed at 37°C for 5 h to induce expression of the β-galactosidase fusion proteins. The filters were then washed and blocked in Tris buffer (0.05 M Tris-HCl, pH 8.0, 0.15 M NaCl) with 1% Tween 20 and 10% nonfat dried milk, and then incubated with a mixture of three murine IgG antitryptase MAbs, G3 (25), G5 (2), and H4 (2) (20 μg of each antibody/ml Tris buffer at room temperature overnight).

On day 2, filters were washed in Tris buffer (as above) containing 1% Triton X-100 and 0.1% SDS four times, then incubated with commercial alkaline phosphatase–conjugated goat anti–mouse IgG (1:2,500 dilution) in Tris buffer for 1 h at room temperature, and then washed again four times. Filters were developed with nitroblue tetrazolium and 5-bromo-4-chloro-3-indolyl phosphate in veronal acetate and dimethylformamide solution; the reaction was stopped by washing in water. Positive plaques typically appeared within 3–5 min.

cDNA sequence determination. Plaque-derived phage corresponding to antitryptase positive material were purified and used to infect *E. coli* (XL1 Blue) in the presence of R408 helper phage to cause the autoexcision of the plasmid, pBluescript, containing the cDNA inserts of interest. Plasmid DNA from miniplasmid preparations of the rescued plasmids were cut with Eco RI restriction endonuclease and subjected to Southern blot analysis. A 29-mer oligonucleotide probe was synthesized corresponding to residue numbers 10–20 of the NH_2-terminal amino acid sequence of purified tryptase determined as above. Seven redundant nucleotide positions were substituted with inosine. Hybridizing inserts should contain the 5′ sequence coding for the NH_2-terminal region of tryptase. The only hybridizing clone producing a strong signal (GRA-2) was then selected for sequencing (see results). Four other clones generated a weak hybridization signal. Four clones in addition to GRA-2 have been sequenced, overlapping regions corresponding exactly with GRA-2. GRA-6 contains the entire 3′ nontranslated region as shown below.

Plasmid DNA from each of the clones that were sequenced was extracted from batch cultures of infected XL1 Blue *E. coli* and purified by ultracentrifugation in CsCl density gradients as described (26). Double-stranded DNA dideoxy sequencing using α[^{35}S]dATP was then performed after alkali denaturation/neutralization and primer annealing followed by addition of Klenow fragment (27) or, in some cases, Sequenase 2.0 (28) by standard protocols. Where necessary, dITP was substituted for dGTP in the Sequenase 2.0 system to resolve compression patterns.

DNA sequence data was stored and analyzed by a VAX computer using the Genetics Computer Group Sequence Analysis Software Package from the University of Wisconsin (version 5.3; July, 1988) (29) and the GenBank data bank (release 58.0; December, 1988).

Results

Of $\sim$ 100,000 plaques produced with phage obtained from the mast cell-enriched cDNA lambda ZAP library, 10 plaques yielded antitryptase positive material. Slot blot analysis revealed that in each case, expressed protein reacted strongly with the H4 and G3 antitryptase antibodies and weakly with the G5 antibody. Southern blot analysis using the 29-mer synthetic oligonucleotide made from the known NH_2-terminal region of the tryptase amino acid sequence and containing seven inosines at sites of redundancy revealed that 1 of the 10 positive cDNA inserts produced a strong hybridization signal with the probe. Inserts ranged in size from $\sim$ 800–900 bases, regardless of hybridization. Each was subcloned into pBluescript, as described, and the clone that did strongly hybridize with the inosine containing oligonucleotide, GRA-2, was subjected to the sequencing strategy shown in Fig. 1. The sequencing strategy for GRA-6, from which the nontranslated 3′

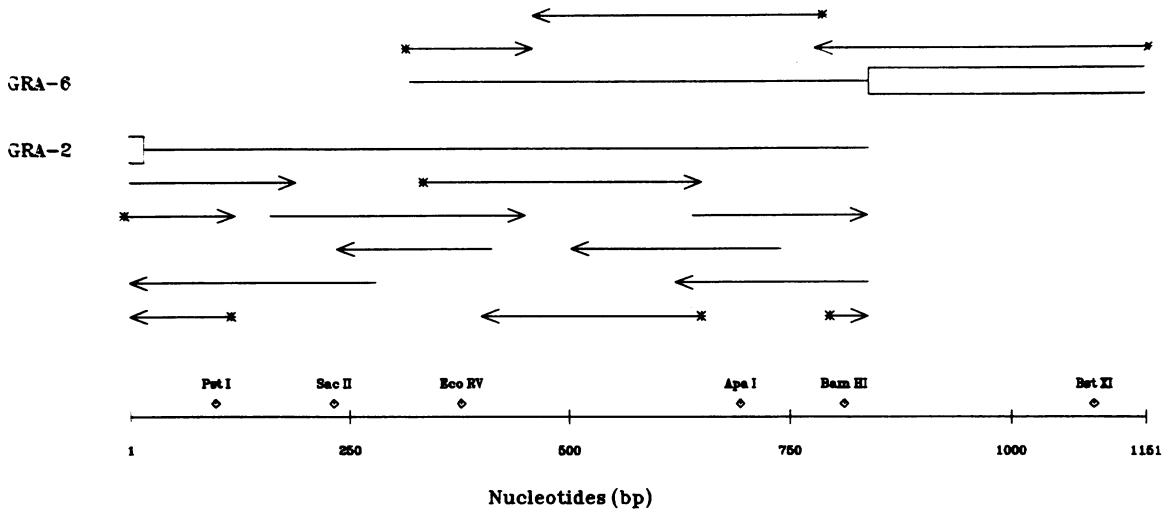

Figure 1. Sequencing strategy for tryptase cDNA. Hexanucleotide restriction enzymes with one recognition site in the tryptase cDNA are shown at their sites of cleavage. Regions of GRA-2 sequenced with Sequenase 2.0 as well as with Klenow fragment are labeled with an asterisk. GRA-6 was sequenced with Sequenase 2.0 alone. Boxed areas of GRA-2 and GRA-6 represent nontranslated regions (see Fig. 2).

region was obtained, is shown in Fig. 1. The entire sequence corresponding to the translated region of human tryptase mRNA was obtained in addition to nontranslated regions of 9 nucleotides 5′ to the ATG initiation site and 308 nucleotides 3′ from the TGA termination codon to the polyA stretch (Fig. 2). The 3′ nucleotide sequence of the nontranslated region of tryptase in GRA-6 was confirmed by finding identical sequences in GRA-4, GRA-7, and GRA-9, all of which lacked variable portions of the 5′ end. The ATG codon at nucleotide positions 18–20 was assigned as the initiator Met because it was the only one detected 5′ to the known NH_2 terminus of the catalytic sequence and because the nucleotide sequence surrounding this codon, TGGCCAGGATGC, is comparable to the (CC)ACCATGG sequence suggested as being optimal for translation initiation by Kozak (30) and to the TGG(0–1 bp)CA(2 bp)ATG sequence preceding the translation initiation site for intracellular proteoglycan core and certain rodent mast cell serine proteases (31). A polyA consensus signal sequence, ATTAAA, was present at positions 1,133–1,138, 13 nucleotides before the polyA region. Overlapping sequences between the two sequenced clones shown in Fig. 1 were identical, suggesting that the identified clones represent the same gene. The composite cDNA contained 1,143 nucleotides up to the polyA region.

The nucleotide and corresponding amino acid sequence in Fig. 2 shows the translated region to contain 274 amino acids, including a 30-amino acid leader sequence of 3048 D and a 244-amino acid catalytic sequence of 27,423 D. The NH_2-terminal amino acid sequence of the catalytic chain, as deduced from the nucleic acid sequence, corresponds exactly with the 20-amino acid NH_2-terminal sequence determined from naturally occurring tryptase purified from human lung as well as with the 8-amino acid NH_2-terminal sequence of purified human pituitary tryptase (12). This along with the antibody recognition and amino acid composition data (see Table I) confirms the identity of this cDNA as representing human tryptase.

The amino acid sequence of human tryptase contained a number of regions that are highly conserved in the trypsin superfamily of serine proteases (Fig. 3). These include an IVGG activation site region (residues 31–34) preceded by a leader sequence, a WVLTAAHC active site histidine region (residues 68–75), a DIALL active-site aspartic acid region (residues 120–124), M-CAG (residues 208–212) downstream from the active site serine, and the GDSGGP active-site serine region (residues 221–226). Putative cystine linkages were present between positions 59 and 75, 154 and 229, 187 and 210, and 219 and 247. Potential *N*-linked carbohydrate binding regions were present at positions 131–133 (NIS) and 232–234 (NGT). In addition, the aspartic acid residue at position 217 in the substrate binding pocket confers a trypsin-like substrate specificity (32), consistent with the substrate specificity of tryptase.

The nucleotide composition of the human tryptase coding region of the cDNA was rich in C and G, the percentage of each base being 34 for C, 30 for G, 19 for A and 17 for T. The amino acid composition of tryptase calculated from the cDNA is shown in Table I, where it is compared with previous determinations performed directly on purified human tryptase. The cDNA-derived amino acid sequence revealed Cys and Trp residues that for technical reasons were not measured in the original report and Asn and Gln residues that were detected as Asp and Glu residues in both reports. This permitted calculation of the net charge of the human tryptase monomer at neutral pH as −4 and the pI as 6.65.

Discussion

The cDNA for human tryptase has been cloned from a library prepared in lambda ZAP with poly A-enriched mRNA isolated from dispersed and partially purified human lung mast cells. Verification of the cDNA as tryptase was based upon complete identity between the NH_2-terminal 20-amino acid sequence obtained with purified, catalytically active tryptase and the corresponding region of the cloned cDNA. In addition, recognition of cDNA-derived protein by three murine MAbs, each recognizing distinct epitopes on human tryptase,

```
        10                  30                  50
GGAATTCCGTGGCCAGGATGCTGAGCCTGCTGCTGCTGGCGCTGCCCGTCCTGGCGAGCC
                  M  L  S  L  L  L  L  A  L  P  V  L  A  S  R   15

        70                  90                  110
GCGCCTACGCGGCCCCTGCCCCAGTCCAGGCCCTGCAGCAAGCGGGTATCGTCGGGGGTC
 A  Y  A  A  P  A  P  V  Q  A  L  Q  Q  A  G |I  V  G  G  Q     35

        130                 150                 170
AGGAGGCCCCCAGGAGCAAGTGGCCCTGGCAGGTGAGCCTGAGAGTCCGCGACCGATACT
 E  A  P  R  S  K  W  P  W  Q  V  S  L  R  V  R  D  R  Y  W     55

        190                 210                 230
GGATGCACTTCTGCGGGGGCTCCCTCATCCACCCCCAGTGGGTGCTGACCGCGGCGCACT
 M  H  F  C  G  G  S  L  I  H  P  Q  W  V  L  T  A  A [H] C     75

        250                 270                 290
GCCTGGGACCGGACGTCAAGGATCTGGCCACCCTCAGGGTCAACTCCGGGACACACCTCT
 L  G  P  D  V  K  D  L  A  T  L  R  V  N  S  G  T  H  L  Y     95

        310                 330                 350
ACTACCAGGACCAGCTGCTGCCAGTCAGCAGGATCATGGTGCACCCACAGTTCTACATCA
 Y  Q  D  Q  L  L  P  V  S  R  I  M  V  H  P  Q  F  Y  I  I    115

        370                 390                 410
TCCAGACTGGAGCGGATATCGCCCTGCTGGAGCTGGAGGAGCCCGTGAACATCTCCAGCC
 Q  T  G  A [D] I  A  L  L  E  L  E  E  P  V  N  I  S  S  R    135

        430                 450                 470
GCGTCCACACGGTCATGCTGCCCCCTGCCTCGGAGACCTTCCCCCCGGGGATGCCGTGCT
 V  H  T  V  M  L  P  P  A  S  E  T  F  P  P  G  M  P  C  W    155

        490                 510                 530
GGGTCACTGGCTGGGGCGATGTGGACAATGATGAGCCCCTCCCACCGCCATTTCCCCTGA
 V  T  G  W  G  D  V  D  N  D  E  P  L  P  P  P  F  P  L  K    175

        550                 570                 590
AGCAGGTGAAGGTCCCCATAATGGAAAACCACATTTGTGACGCAAAATACCACCTTGGCG
 Q  V  K  V  P  I  M  E  N  H  I  C  D  A  K  Y  H  L  G  A    195

        610                 630                 650
CCTACACGGGAGACGACGTCCGCATCATCCGTGACGACATGCTGTGTGCCGGGAACAGCC
 Y  T  G  D  D  V  R  I  I  R  D  D  M  L  C  A  G  N  S  Q    215

        670                 690                 710
AGAGGGACTCCTGCAAGGGCGACTCTGGAGGGCCCCTGGTGTGCAAGGTGAATGGCACCT
 R |D  S  C  K| G  D [S] G  G  P  L  V  C  K  V  N  G  T  W    235

        730                 750                 770
GGCTACAGGCGGGCGTGGTCAGCTGGGACGAGGGCTGTGCCCAGCCCAACCGGCCTGGCA
 L  Q  A  G  V  V |S  W  D| E  G  C  A  Q  P  N  R  P  G  I    255

        790                 810                 830
TCTACACCCGTGTCACCTACTACTTGGACTGGATCCACCACTATGTCCCCAAAAAGCCGT
 Y  T  R  V  T  Y  Y  L  D  W  I  H  H  Y  V  P  K  K  P       274

        850                 870                 890
GAGTCAGGCCTGGGTGTGCCACCTGGGTCACTGGAGGACCAACCCCTGCTGTCCAAAACA

        910                 930                 950
CCACTGCTTCCTACCCAGGTGGCGACTGCCCCCCACACCTTCCCTGCCCCGTCCTGAGTG

        970                 990                 1010
CCCCTTCCTGTCCTAAGCCCCCTGCTCTCTTCTGAGCCCCTTCCCCTGTCCTGAGGACCC

        1030                1050                1070
TTCCCCATCCTGAGCCCCCTTCCCTGTCCTAAGCCTGACGCCTGCACTGCTCCGGCCCTC

        1090                1110                1130
CCCTGCCCAGGCAGCTGGTGGTGGGCGCTAATCCTCCTGAGTGCTGGACCTCATTAAAGT

        1150
GCATGGAAATCA_n
```

Figure 2. Nucleotide and amino acid sequence of tryptase cDNA. Catalytic triad residues ([__]), leader sequence (___), substrate binding pocket regions (| ‾‾‾ |), and putative NH_2-terminal carbohydrate binding sites (____) are shown as indicated.

further validates the tryptase identity of the cloned cDNA molecule, GRA-2, as well as GRA-6. The fact that three antibodies recognize GRA-2 (containing amino acids 1–274) as well as GRA-6 (containing amino acids 102–274) indicates that the antigenic epitopes reside in the COOH-terminal two-thirds of tryptase. Tryptase, like most other serine proteases, contains a hydrophobic NH_2-terminal leader sequence as now recognized from the cDNA, but no apparent activation peptide between the leader and catalytic sequences. Tryptase is stored in mast cell secretory granules fully active without an NH_2-terminal leader peptide. Presumably, the leader directs tryptase to the secretory granules and is removed before or soon after the enzyme is packaged. Whether the leader sequence might play a role in the association of tryptase monomers to form the active tetramer also should be considered.

Enzymatically active tryptase migrates in gel filtration matrices with an apparent size of 120,000–140,000 D, and by SDS PAGE has subunits of 31,000–34,000 D. The evidence accumulated thus far indicates that these subunits are products of a single gene and vary in size due to posttranslational events. A single NH_2-terminal amino acid sequence in preparations of purified tryptase containing a range of subunit sizes has been reported previously (19) and in the current study. A blocked NH_2-terminus in one of the major subunits was unlikely because the amount of protein calculated with data from the amino acid sequencing procedure was similar to the amount calculated with data from the amino acid composition for the same tryptase preparation (Schwartz, L. B., unpublished data). In addition, each subunit contains one active site (10), regardless of size, and shares common epitopes. However, the possibility that two different but highly homologous tryptase genes do exist has not been ruled out and is currently under investigation. The presence of two putative *N*-linked glycosylation sites and the possibility of COOH-terminal processing need to be considered as potential explanations for size heterogeneity. The smaller subunit size of 27,423 D calculated from the cDNA than estimated by SDS-PAGE suggests that the addition of carbohydrate residues may be more likely than cleavage of COOH-terminal peptides in the production of subunits with higher apparent molecular weights. Pituitary, but not lung-derived human tryptase, reportedly is sensitive to *N*-glycosidase F, subunits being reduced from 36,300 and 34,600 D to 33,400 and 32,400 D (12). This reduction in size does not account for the difference in predicted verses measured molecular weights of tryptase monomers.

The net charge of each catalytic tryptase monomer is −4, yet active tryptase binds more tightly at neutral and slightly acidic pH to negatively charged species such as heparin (elution at 0.8 M NaCl) than positively charged species such as DEAE-Sephacyl (elution at 0.18 M NaCl). This suggests that regions of positive charge, particularly those in the nonconserved positions between CRI and CRII (net +2) and after CRVII (net +2) may be optimally oriented in the tetramer for binding to negatively charged matrices.

The amino acid sequence of the catalytic portion of tryptase exhibits a high degree of homology with the catalytic portions of thirteen other members of the trypsin superfamily of serine proteases (20, 33–43) (Fig. 3). Percent similarities with tryptase ranged from 53 to 84. In contrast, a computer search failed to identify proteins containing regions with a high degree

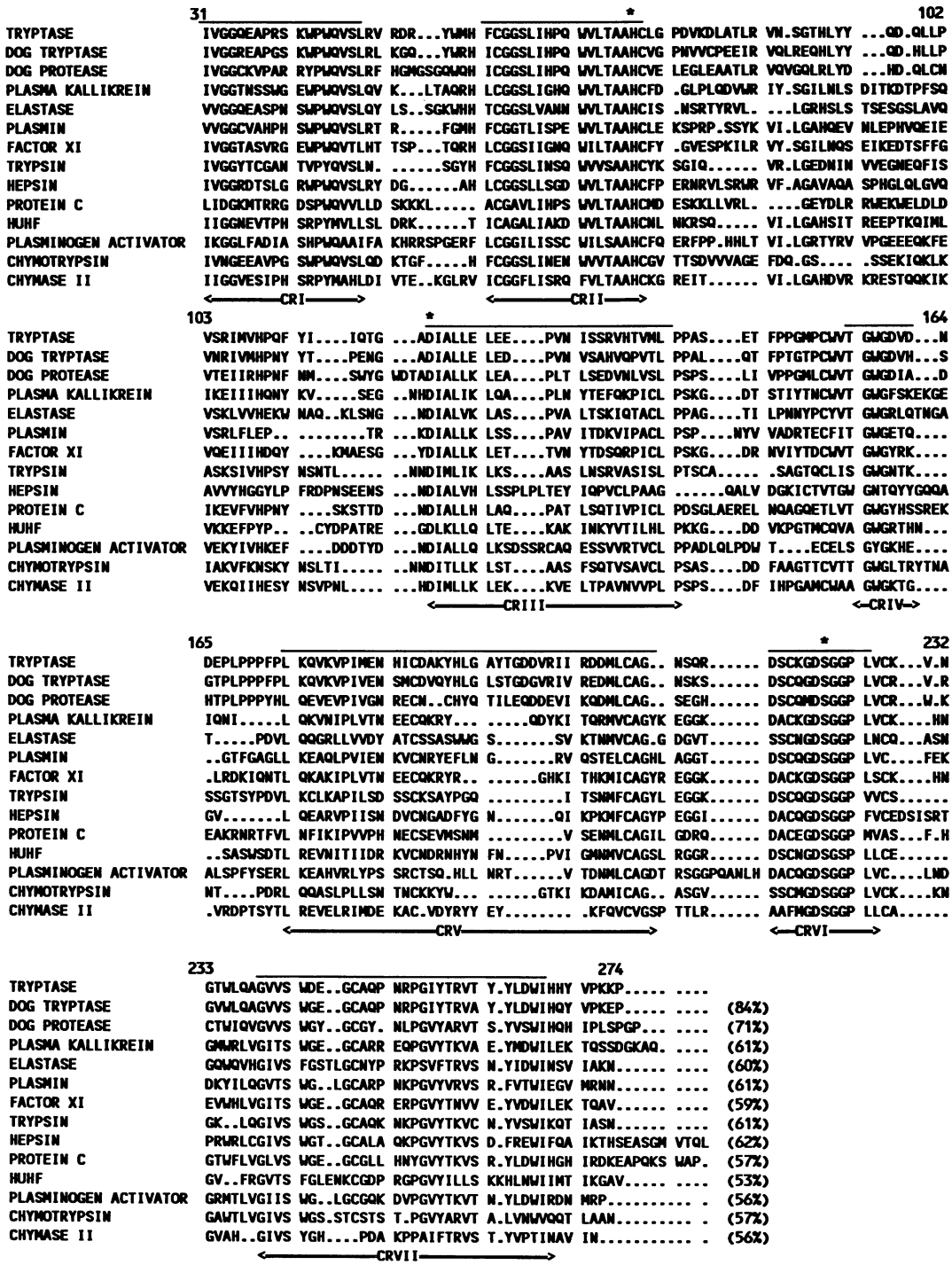

Figure 3. Comparisons of the amino acid sequence of the catalytic portion of tryptase with the catalytic chains of other serine proteases within the trypsin superfamily. These proteases together with the percent similarity (in brackets) of each to tryptase, calculated according to Needleman and Wunsch (gap weight = 5, length weight = 0.3) (53), include dog tryptase (20), dog protease (20), human plasma kallikrein (33), rat elastase II (34), human plasmin (35), human Factor XI (36), bovine trypsin (37), human liver hepsin (38), human protein C (39), human Hanukah factor (40), tissue plasminogen activator (41), bovine chymotrypsin (42) and rat chymase II (rat mast cell proteinase II) (43). Residues are numbered according to the amino acid sequence of the catalytic portion of tryptase. The seven conserved regions (CRI-VII) of the trypsin superfamily are designated according to Furie et al. (44) and contain the major regions of similarity. Gaps were inserted to bring these conserved regions into optimal alignment. Asterisks have been placed above the active site His, Asp and Ser residues.

of homology to the leader sequence of human tryptase other than dog tryptase (Fig. 4). Internal regions of the catalytic chains of serine proteases are known to be conserved, whereas variable regions tend to be on the surface. Seven such conserved regions, CRI–CRVII, separated by six variable regions and a seventh COOH-terminal variable region were identified by Furie et al. (44). Human tryptase clearly follows this same pattern.

The similarity of the catalytic portions of human tryptase with those of dog tryptase (84%) and dog protease (71%) relative to other proteases is striking. Dog tryptase has one additional amino acid, shown between Asn_{89} and Ser_{90} of human tryptase in Fig. 2, whereas five gaps are needed to optimally align human tryptase with dog protease. Dog tryptase also has a net negative charge, −3, with regions of positive charge between CRI and CRII (net +2) and between CRVI and CRVII (net +2). The single putative *N*-linked glycosylation site in dog tryptase appears to contain carbohydrate in the native enzyme (20). Dog tryptase also has a 30-amino acid leader sequence, which when compared to that of human tryptase, has 67%

Table I. Amino Acid Composition of Human Tryptase

Amino acid	1981 (1)	1985 (11)	1989
Basic			
Arg	5.2	4.9	4.9
Lys	3.7	4.2	3.7
Acidic			
Asp	9.7	8.8	6.9
Glu	11.9	9.1	3.3
Nonpolar			
Ala	6.7	5.6	4.9
Val	7.9	8.8	9.0
Ile	3.3	4.2	5.3
Leu	8.4	9.1	8.2
Met	1.3	1.4	2.5
Pro	9.4	9.5	9.4
Gly	11.9	9.5	8.2
Polar, nonionic			
Ser	7.5	4.6	4.9
Asn	—	—	2.9
Gln	—	—	4.5
Thr	4.5	4.6	4.5
Polar, ionic			
Cys	0.3	3.1	3.3
His	4.1	3.5	4.1
Aromatic			
Trp	—	3.2	3.7
Phe	1.9	1.8	1.6
Tyr	2.5	3.9	4.1

similarity and 53% identity with no gaps (Fig. 4). The terminal Gly residue of the leader in both enzymes is uncommon in other proteases and has been considered to suggest a novel mode of activation (20). Dog and human tryptase also have 3′ nontranslated regions of 48% similarity that contain a 20-nucleotide region of identity before the polyA region; the length

```
Leaders

HUMAN     1  MLSLLLLALPVLASRAYAAPAPVQALQQAG 30
             | | | ||| | |    ||| |||| |
DOG       1  MPSPLVLALALLGSLVPVSPAPGQALQRVG 30

3′ Untranslated Regions

Human   990 TTCTGAGCCCCTTCCCCTGTCCTGAGGACCCTTCCCCATCCTGAGCCCCC 1039
            |||||||   |||| | ||    |                 ||| ||
Dog     831 ...TGAGCCCAGCCCCCGGGCCGCCCGGGTTGGTGGGGAAGCCGGCCTCC  877

Human  1040 TTCCCTGTCCTAAGCCTGACGCCTGCACTGCTCCGGCCCTCCCCTGCCCA 1089
            |    |    |    | || || | |   | |||||   |  |||
Dog     878 ACGGCGCCTCACCCCTGCCCCCCGGCCCGGGGGCCGCCCTTTTC.CCCCT  926

Human  1090 GGCAGCTGGTGGTGGGCGCTAATCCTCCTGAGTGCTGGACCTCATTAAAG 1139
            | |  |  |   ||  | | | ||| | | | ||    ||||||||||
Dog     927 GTCCCCGTGGCCTGTCCCCGGAGCCCGCCGTGGGCCCACCCTCATTAAAG  976

Human  1140 TGCATGGAAATC 1151
            |||||||||
Dog     977 TGCATGGAAGC.  987
```

Figure 4. Comparisons of the leader amino acid and 3′ nontranslated nucleotide sequences of human tryptase and dog tryptase. Identical residues are noted by a hatch mark.

of this region in the human is 162 bp and in the dog it is only 57-bp. The extensive homologies across the cDNAs of mast cell tryptase in dog and human species clearly indicate conserved characteristics.

Crystallographic studies of chymotrypsin, trypsin, chymase II, and elastase I (45–48) have indicated regions of conserved function for serine proteases that can be applied to tryptase. Human tryptase residues 219 and 221–225 are highly conserved in most serine proteases, and provide a rigid backbone to properly orient the active-site Ser_{223}. Hydrogen bonding between Cys_{219}/Asp_{222} and Asp_{222}/Gly_{225} and flexibility due to Gly_{221} and Gly_{224} are critical (49) even though Phe can replace Cys_{219} in cytotoxic T lymphocyte chymotryptic proteinase I (50) and rat chymase II, which lack the fourth disulfide linkage present in most serine proteases. Residues Met in chymotrypsin and Gln in trypsin corresponding to Lys_{220} in tryptase may provide a flexible cover over the entrance to the substrate binding pocket (32, 51). Lys_{220} in human tryptase and the corresponding Lys in plasma kallikrein and Factor XI may provide a polar environment for substrate side chains and only allow entry into the active site of similarly charged residues that repel its side chain. In addition, the sequence Gly_{158}, Trp_{159}, and Gly_{160} in CRIV is highly conserved in serine proteases, including human tryptase, and reportedly is involved in the activation process (52).

Human tryptase residues 217–220 (DSCK) probably form one side of the substrate binding pocket, the Asp residue conferring specificity for internal Arg and Lys amino acids. The other side of the binding pocket in human tryptase contains residues 242–244 (SWD). Although these Ser and Trp residues are consistent with comparable residues in other serine proteases the presence of Asp_{244}, rather than Gly, is rather unique to human tryptase, dog tryptase having SWG. Asp rather than Gly results from an adenosine rather than guanosine in position 2 of that codon. This adenosine was found in each of five different tryptase cDNA clones and therefore is not likely to represent a random error of transcription during cDNA formation or a polymorphism. These SWD(G) residues presumably interact with substrate in an antiparallel β-structure and thereby properly orient the bond to be cleaved (46). In particular the amido nitrogen and carbonyl oxygen of P3 on the substrate are predicted to form hydrogen bonds with the carbonyl oxygen and amido nitrogen, respectively, of Gly in most proteases, and the Asp in tryptase. Rat elastase I has Val in this position. Asp_{244}, at first approximation, might be expected to reduce the size of the substrate binding pocket and to restrict the substrate specificity toward substrates rich in basic amino acids near the cleavage site. A preliminary analysis of Asp in this position was performed using a molecular modeling program (FRODO version 6.5) by substituting Asp for Gly into the known three-dimensional structure of the bovine trypsin-pancreatic trypsin inhibitor complex. The carboxyl side chain of the substituted Asp (Asp_{244} in tryptase) as well as the Asp carboxyl group known to determine trypsin-like substrate specificity (Asp_{217} in tryptase) could be positioned to form near optimal ion pair bonds with the gamma-amino group of Lys or the guanidinium group of Arg, perhaps favoring the latter with its two available nitrogens. In addition, Glu_{245} in human tryptase, like the analogous residue in kallikrein, appears to interact with the side chain of the P3 residue. Thus, basic residues in positions P1 and P3 would appear to be favored. A previous

study with a series of tripeptide-*p*-nitroanilide synthetic substrates having Arg at P1, in fact showed that a substrate with basic residues in P3 and P1 positions, separated by Gly, was favored (18). This restricted substrate specificity also may explain why the classic inhibitors of serine proteases present in serum, urine, lung, and skin fail to inhibit human tryptase. Further studies will be needed to better understand these structure-function relationships.

Acknowledgments

We thank Dr. Tony Wright (Medical College of Virginia, Richmond, VA) for performing the molecular modeling of the substrate binding pocket in tryptase and Dr. Anne-Marie Irani (Medical College of Virginia) for performing the MC_T/MC_{TC} determination.

This work was supported in part by National Institutes of Health grants AI-20487 and AI-27517, Virginia Center for Innovative Technology Award BIO-87013, and Pharmacia Award 87-0697.

References

1. Schwartz, L. B., R. A. Lewis, D. Seldin, and K. F. Austen. 1981. Acid hydrolases and tryptase from secretory granules of dispersed human lung mast cells. *J. Immunol.* 126:1290–1294.

2. Schwartz, L. B. 1985. Monoclonal antibodies against human mast cell tryptase demonstrate shared antigenic sites on subunits of tryptase and selective localization of the enzyme to mast cells. *J. Immunol.* 134:526–531.

3. Schwartz, L. B., A. A. Irani, K. Roller, M. C. Castells, and N. M. Schechter. 1987. Quantitation of histamine, tryptase and chymase in dispersed human T and TC mast cells. *J. Immunol.* 138:2611–2615.

4. Schwartz, L. B., D. D. Metcalfe, J. Miller, H. Earl, and T. Sullivan. 1987. Tryptase levels as an indicator of mast cell activation in systemic anaphylaxis and mastocytosis. *N. Engl. J. Med.* 316:1622–1626.

5. Schwartz, L. B., P. C. Atkins, T. R. Bradford, P. Fleekop, M. Shalit, and B. Zweiman. 1987. Release of tryptase together with histamine during the immediate cutaneous response to allergen. *J. Allergy Clin. Immunol.* 80:850–855.

6. Wenzel, S. E., A. A. Fowler, and L. B. Schwartz. 1988. Activation of pulmonary mast cells by bronchoalveolar allergen challenge: in vivo release of histamine and tryptase in atopic subjects with and without asthma. *Am. Rev. Respir. Dis.* 137:1002–1008.

7. Castells, M., and L. B. Schwartz. 1988. Tryptase levels in nasal lavage fluid as an indicator of the immediate allergic response. *J. Allergy Clin. Immunol.* 82:348–355.

8. Craig, S. S., G. DeBlois, and L. B. Schwartz. 1986. Mast cells in human keloid, small intestine and lung by an immunoperoxidase technique using a murine monoclonal antibody against tryptase. *Am. J. Pathol.* 124:427–435.

9. Irani, A. A., N. M. Schechter, S. S. Craig, G. DeBlois, and L. B. Schwartz. 1986. Two human mast cell subsets with distinct neutral protease compositions. *Proc. Natl. Acad. Sci. USA.* 83:4464–4468.

10. Schwartz, L. B., R. A. Lewis, and K. F. Austen. 1981. Tryptase from human pulmonary mast cells: purification and characterization. *J. Biol. Chem.* 256:11939–11943.

11. Smith, T. J., M. W. Hougland, and D. A. Johnson. 1984. Human lung tryptase, purification and characterization. *J. Biol. Chem.* 259:11046–11049.

12. Cromlish, J. A., N. G. Seidah, M. Marcinkiewcz, J. Hamelin, D. A. Johnson and M. Chretein. 1987. Human pituitary tryptase: molecular forms, NH_2-terminal sequence, immunocytochemical localization and specificity with prohormone and fluorogenic substrates. *J. Biol. Chem.* 262:1363–1373.

13. Schwartz, L. B., and T. M. Bradford. 1986. Regulation of tryptase from human lung mast cells by heparin: stabilization of the active tetramer. *J. Biol. Chem.* 261:7372–7379.

14. Alter, S. C., D. D. Metcalfe, T. R. Bradford, and L. B. Schwartz. 1987. Stabilization of human mast cell tryptase: effects of enzyme concentration, ionic strength and the structure and negative charge density of polysaccharides. *Biochem. J.* 248:821–827.

15. Alter, S. C., P. Yates, H. S. Margolius, and L. B. Schwartz. 1987. Tryptase and kinin generation: tryptase from human mast cells does not activate human urinary prokallikrein. *Int. Arch. Allergy Applied Immunol.* 83:321–324.

16. Schwartz, L. B., M. S. Kawahara, T. E. Hugli, D. Vik, D. T. Fearon, and K. F. Austen. 1983. Generation of C3a anaphylatoxin from human C3a by human mast cell tryptase. *J. Immunol.* 130:1891–1895.

16a. Gruber, B. L., M. J. Marchese, K. Suziki, L. B. Schwartz, Y. Okada, H. Nagase, and N. S. Ramamurthy. 1989. Synovial procollagenase activation by human mast cell tryptase: dependence upon metalloproteinase III activation. *J. Clin. Invest.* In press.

17. Schwartz, L. B., T. R. Bradford, B. L. Littman, and B. U. Wintroub. 1985. The fibrinogenolytic activity of purified tryptase from human lung mast cells. *J. Immunol.* 135:2762–2767.

18. Tanaka, T., B. J. McRae, K. Cho, R. Cook, J. E. Fraki, D. A. Johnson and J. C. Powers. 1983. Mammalian tissue trypsin-like enzymes. Comparative reactivities of human skin tryptase, human lung tryptase and bovine trypsin with peptide 4-nitroanilides and thioester substrates. *J. Biol. Chem.* 258:13552–13557.

19. Seizawa, K., G. H. Caughey, S. C. Lazarus, W. M. Gold, and J. A. Nadel. 1989. Mast cell tryptase causes airway smooth muscle hyperresponsiveness in dogs. *J. Clin. Invest.* 83:175–179.

20. Vanderslice, P., C. S. Craik, J. A. Nadel, and G. H. Caughey. 1989. Molecular cloning of dog mast cell tryptase and a related protease: structural evidence of a unique mode of serine protease activation. *Biochemistry.* 28:4148–4155.

21. Chirgwin, J. M., A. E. Przybyla, R. J. MacDonald and W. J. Rutter. 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry.* 18:5294–5299.

22. Schweizer, J., and K. Goerttler. 1980. Synthesis in vitro of keratin polypeptides directed by mRNA isolated from newborn and adult mouse epidermis. *Eur. J. Biochem.* 112:243–249.

23. Aviv, H., and P. Leder. 1972. Purification of biologically active globin messenger RNA by chromatography on oligo thymidylic acid-cellulose. *Proc. Natl. Acad. Sci. USA.* 69:1408–1412.

24. Mierendorf, R. C., C. Percy, and R. A. Young. 1987. Gene isolation by screening lambda gt11 libraries with antibodies. *Methods Enzymol.* 152:458–469.

25. Irani, A. A., Bradford, T. R., Kepley, C. L., Schechter, N. M. and Schwartz, L. B. 1989. Detection of MC_T and MC_{TC} types of human mast cells by immunohistochemistry using new monoclonal anti-tryptase and anti-chymase antibodies. *J. Histochem. Cytochem.* In press.

26. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular Cloning: A Laboratory Handbook. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. pp. 88–96.

27. Mirendorf, R. C., and D. Pfeffer. 1987. Direct sequencing of denatured plasmid DNA. *Methods Enzymol.* 152:556–563.

28. Toneguzzo, F., S. Glyn, E. Levi, S. Mjolsness, and A. Hayday. 1988. Use of chemically modified T7 DNA polymerase for manual and automated sequencing of supercoiled DNA. *Biotechniques.* 6:460–469.

29. Devereuz, J., P. Haeverli, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* 12:387–395.

30. Kozak, M. 1986. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell.* 44:283–292.

31. Avraham, S., R. L. Stevens, C. F. Nicodemus, M. C. Gartner, K. F. Austen, and J. H. Weiss. 1989. Molecular cloning of a cDNA that encodes the peptide core of a mouse mast cell secretory granule pro-

teoglycan and comparison with the analogous rat and human cDNA. *Proc. Natl. Acad. Sci. USA.* 86:3763–3767.

32. Steitz, T. A., R. Henderson, and D. M. Blow. 1969. Structure of crystalline alpha-chymotrypsin. *J. Mol. Biol.* 46:337–348.

33. Chung, D. W., K. Fujikawa, B. A. McMullen, and E. W. Davie. 1986. Human plasma prekallikrein, a zymogen to a serine protease that contains four tandem repeats. *Biochemistry.* 25:2410–2417.

34. MacDonald, R. J., G. H. Swift, C. Quinto, W. Swain, R. L. Picket, W. Nikovits and W. J. Rutter. 1982. Primary structure of two distinct rat pancreatic preproelastases determined by sequence analysis of the complete cloned messenger ribonucleic acid sequences. *Biochemistry.* 21:1453–1463.

35. Malinkowski, D. P., J. E. Sadler, and E. W. Davie. 1984. Characterization of a complementary deoxyribonucleic acid coding for human and bovine plasminogen. *Biochemistry.* 23:4243–4250.

36. Fujikawa, K., D. W. Chung, L. E. Hendrikson, and E. W. Davie. 1986. Amino acid sequence of human Factor XI, a blood coagulation factor with four tandem repeats that are highly homologous with plasma kallikrein. *Biochemistry.* 25:2417–2424.

37. Hartley, B. S. 1970. Homologies in serine proteases. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 257:77–87.

38. Leytus, S. P., K. R. Loeb, F. S. Hagen, K. Kurachi, and E. W. Davie. 1988. A novel trypsin-like serine protease (hepsin) with a putative transmembrane domain expressed by human liver and hepatoma cells. *Biochemistry.* 27:1067–1074.

39. Foster, D. C., S. Yashitake, and E. W. Davie. 1985. The nucleotide sequence for the gene for human protein C. *Proc. Natl. Acad. Sci. USA.* 82:4673–4677.

40. Gershenfeld, H. K., R. J. Hershenberger, T. B. Shows, and I. L. Weissman. 1988. Cloning and chromosomal assignment of a human cDNA encoding a T cell- and natural killer cell-specific tryptase-like serine protease. *Proc. Natl. Acad. Sci. USA.* 85:1184–1188.

41. Ny, T., F. Elgh, and B. Lund. 1984. The structure of the human tissue-type plasminogen activator gene. Correlation of intron and exon structure to functional and structural domains. *Proc. Natl. Acad. Sci. USA.* 81:5355–5359.

42. Blow, D. M., J. J. Bicktoft, and B. S. Hartley. 1969. Role of a buried acid group in the mechanism of action of chymotrypsin. *Nature (Lond.).* 221:337–340.

43. Woodbury, R. G., N. Katunuma, K. Kobayashi, K. Titani, and H. Neurath. 1978. Covalent structure of a group-specific protease from rat small intestine. *Biochemistry.* 17:811–819.

44. Furie, B., D. H. Bing, R. J. Feldman, D. J. Robinson, J. P. Burnier, and B. C. Furie. 1982. *J. Biol. Chem.* 257:3875–3882.

45. Cohen, G. H., E. W. Silverton, and D. R. Davis. 1981. Refined crystal structure of r-chymotrypsin at 1.9 Angstrom resolution: comparison with other pancreatic serine proteases. *J. Mol. Biol.* 148:449–479.

46. Polgar, L. 1987. Structure and function of serine proteases. *In* Hydrolytic Enzymes. A. Neuberger and K. Broddehurst, editors. Elsevier Science Publishers, Amsterdam. 159–200.

47. Sawyer, L., D. M. Shotton, J. W. Campbell, P. L. Wendell, H. Muirhead, H. C. Watson, R. Diamond, and R. C. Ladner. 1978. The atomic structure of crystalline porcine pancreatic elastase at 2.5 Angstrom resolution: comparisons with the structure of α-chymotrypsin. *J. Mol. Biol.* 118:137–208.

48. Remington, S. J., R. G. Woodbury, R. A. Reynolds, B. W. Matthews, and H. Neurath. 1988. The structure of rat mast cell protease II at 1.9-A resolution. *Biochemistry.* 27:8097–8105.

49. Birktoft, J. J., D. M. Blow, R. Henderson, and J. A. Steitz. 1970. I. Serine proteinases. The structure of α-chymotrypsin. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 257:67–76.

50. Lobe, C. G., B. B. Finlay, W. Paranchych, V. H. Paetleav, and R. C. Bleackley. 1986. Novel serine proteases encoded by two cytoxic T lymphocyte-specific genes. *Science (Wash. DC).* 232:858–861.

51. Krieger, M., L. M. Kay, and R. M. Stroud. 1974. Structure and specific binding of trypsin: comparison of inhibited derivatives and a model for substrate binding. *J. Mol. Biol.* 83:209–230.

52. Fehlhammer, H., W. Bode, and R. Huber. 1977. Crystal structure of bovine trypsinogen at 1.8 Angstrom resolution. *J. Mol. Biol.* 111:415–438.

53. Needleman, S. B., and C. D. Wunsch. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* 48:443–453.