Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 3811–3815, May 1990
Biochemistry

Human mast cell tryptase: Multiple cDNAs and genes reveal a multigene serine protease family

PETER VANDERSLICE*, SUSAN M. BALLINGER*, ELIZABETH K. TAM*†, SANFORD M. GOLDSTEIN‡, CHARLES S. CRAIK§, AND GEORGE H. CAUGHEY*¶‖

*Cardiovascular Research Institute and Departments of ‡Dermatology, §Pharmaceutical Chemistry, and ¶Medicine, University of California, San Francisco, CA 94143

Communicated by John A. Clements, February 13, 1990

ABSTRACT **Three different cDNAs and a gene encoding human skin mast cell tryptase have been cloned and sequenced in their entirety. The deduced amino acid sequences reveal a 30-amino acid prepropeptide followed by a 245-amino acid catalytic domain. The C-terminal undecapeptide of the human preprosequence is identical in dog tryptase and appears to be part of a prosequence unique among serine proteases. The differences among the three human tryptase catalytic domains include the loss of a consensus N-glycosylation site in one cDNA, which may explain some of the heterogeneity in size and susceptibility to deglycosylation seen in tryptase preparations. All three tryptase cDNAs are distinct from a recently reported cDNA obtained from a human lung mast cell library. A skin tryptase cDNA was used to isolate a human tryptase gene, the exons of which match one of the skin-derived cDNAs. The organization of the ≈1.8-kilobase-pair tryptase gene is unique and is not closely related to that of any other mast cell or leukocyte serine protease. The 5′ regulatory regions of the gene share features with those of other serine proteases, including mast cell chymase, but are unusual in being separated from the protein-coding sequence by an intron. High-stringency hybridization of a human genomic DNA blot with a fragment of the tryptase gene confirms the presence of multiple tryptase genes. These findings provide genetic evidence that human mast cell tryptases are the products of a multigene family.**

Tryptase, a major constituent of mast cell secretory granules, has emerged recently as the most specific marker *in vivo* of mast cell activation in humans (1, 2). Tryptase is a trypsin-like serine protease that hydrolyzes peptide bonds on the C-terminal side of basic amino acids. Unlike trypsin, however, tryptase is highly selective in hydrolyzing its peptide and protein targets and resists inactivation by circulating inhibitors (3–6). *In vitro* studies using purified dog and human tryptase preparations suggest potential roles in the lung, skin, and other tissues as a local anticoagulant (7, 8), activator of collagenase (9), promoter of smooth muscle contraction (10), and modulator of neuropeptide activity (11, 12).

The distinctive properties of tryptase have yet to be adequately understood in molecular terms. Although tryptase is stored and released from the mast cell granule in an active form (13, 14), the tryptase preprosequence deduced from dog and human cDNAs suggests that tryptase may be synthesized as a proenzyme (15, 16). Preparations of human and dog tryptase with a single N-terminal sequence are inhomogeneous, yielding two or more closely spaced bands on sodium dodecyl sulfate/polyacrylamide gel electrophoresis (SDS/PAGE) (3, 5, 6, 17). Some, but not all, of this heterogeneity is due to variable N-glycosylation (15, 17). The active form of dog and human tryptase is a tetramer (6, 18, 19), each subunit of which shares antigenic determinants and contains an active site (3). Tryptases from different tissue sources are antigenically cross-reactive (17, 20, 21). However, differences in k_{cat} and K_m toward synthetic peptide substrates and in susceptibility towards protease inhibitors have been observed among preparations of tryptase from lung, skin, and pituitary gland, raising the possibility of functional differences between mast cell tryptases in different locations (4, 17).

To correlate the structure of tryptases with their unique properties and to explore the potential genetic basis of heterogeneity and tissue differences in human tryptases, we cloned and sequenced three different human tryptase cDNAs and a gene corresponding to one of the cDNAs.** A genomic DNA blot probed with a fragment of the cloned gene revealed the presence of multiple tryptase genes.

MATERIALS AND METHODS

Construction and Screening of a Skin cDNA Library. A sample of 4×10^6 mast cells (1.1% of total cells) from scalp skin of a single donor was obtained (22), and poly(A)$^+$ RNA was isolated by LiCl precipitation (23) and oligo(dT)-cellulose chromatography (24). A cDNA library containing 8×10^5 unique recombinants was constructed in phage λ ZAP II vector (Stratagene) (25) and was amplified once in *Escherichia coli* XL1-Blue cells (26) prior to screening. Approximately 10^6 plaque-forming units from the library were screened at 42°C with a dog tryptase cDNA (15) labeled to 2×10^8 cpm/μg by nick-translation (27) as described, except for use of 30% formamide in the hybridization solution (27). Three positive recombinants were identified by autoradiography, plaque-purified, and reprobed. Phagemids containing inserts that hybridized to the cDNA probe were excised from the ZAP II λ phage vector by using R408 helper phage (28), transformed into *E. coli* XL1-Blue cells, and purified by alkaline lysis (29). The sequence of the cDNA inserts was determined by dideoxy chain termination (30) modified for double-stranded DNA (31) by using Sequenase (United States Biochemical) (32). The M13 forward, reverse, and KS primers (Stratagene) were used initially. Subsequent sequencing reactions used oligonucleotide primers designed from previously determined sequence.

Genomic Library Screening. Human tryptase clone II cDNA (see *Results*) was used to screen ≈6×10^5 plaque-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: RMCP II, rat mast cell proteinase II.

†Present address: Division of Pulmonary and Critical Care Medicine, 111J, La Jolla Veterans Administration Medical Center, La Jolla, CA 92161.

‖To whom reprint requests should be addressed at: Cardiovascular Research Institute, University of California, San Francisco, CA 94143-0130.

**The sequences reported in this paper have been deposited in the GenBank data base [accession nos. M33491 (for tryptase I cDNA), M33492 (for tryptase II cDNA), M33493 (for tryptase III cDNA), and M33494 (for tryptase I gene)].

forming units from a human placental genomic library in EMBL-3 (Clontech). The cDNA was labeled with biotin-7-dATP by nick-translation and was hybridized at 50°C to the immobilized phage DNA (33). Positive clones were visualized by using the BluGENE nonradioactive nucleic acid detection system (Bethesda Research Laboratories) (33). Nine of the most strongly hybridizing clones were plaque-purified and rescreened. Phage DNA was purified by the plate lysate method (27) and digested with *Bam*HI to yield genomic fragments, which were separated by agarose gel electrophoresis, transferred to nitrocellulose (34), and hybridized to tryptase cDNA. Hybridizing fragments from two of the clones were ligated into the *Bam*HI site of pBluescript KS+ phagemid (Stratagene), and the nucleotide sequence was determined as described for the cDNAs. The nucleotide sequence of both strands of the gene shown in Fig. 2 was determined with the exception of the sequence of the 3′ 179 nucleotides, which was determined in one direction. The sequence of 521 bases of the second clone was determined in one direction.

Genomic DNA Blotting. Human lymphocytes from a single individual were purified from whole blood by centrifugation on Ficoll–Paque (Pharmacia) (35), and genomic DNA was isolated by proteinase K treatment and phenol extraction (36). Samples of 10 μg of DNA were digested to completion with *Ava* II, *Bal* I, and *Bam*HI (27). The fragments were electrophoresed through a 0.8% agarose gel and transferred to a Hybond-N membrane (Amersham) (34). A 345-base-pair (bp) *Ava* II/*Bam*HI DNA fragment (bases 1160–1504 in Fig. 2) isolated from the genomic clone was labeled to 10^9 cpm/μg by nick-translation (27) and hybridized to the genomic blot at 42°C in 10 ml of hybridization solution containing 50% formamide as described above for the cDNA library (27). The filter was washed three times for 30 min at 68°C in 15 mM NaCl/1.5 mM sodium citrate/0.5% SDS, dried, and exposed to film for 24 hr (27).

RESULTS

Tryptase cDNAs. Three unique clones (I, II, and III) from the skin library were identified, and the sequence of both strands of each insert was determined (Fig. 1). Clone I extends from base 5 shown in Fig. 1 to the poly(A) tail. Amino acids 1–24 of the sequence deduced from clone I are identical to those of the N-terminal 24 amino acids of purified human lung tryptase that could be identified by Edman sequencing (unpublished data). His-44, Asp-91, and Ser-194 correspond to the conserved His-57, Asp-102, and Ser-195 (chymotrypsinogen numbering) residues essential for serine protease catalytic activity. Asp-188 is analogous to Asp-189 of trypsin, which is a primary determinant of the arginine/lysine substrate specificity of trypsin. Clone I tryptase contains eight cysteines and two consensus N-glycosylation sites. A 30-amino acid preprosequence precedes a catalytic domain of 245 amino acids with a calculated M_r of 27,449. Clone II comprises bases 3–1130 shown in Fig. 1. Differences in the deduced amino acid sequence of clone II include conversion of Asn-102 to lysine, eliminating an N-glycosylation site. The nucleotide sequence of clone III extends from base 24 to base 1104 in Fig. 1. The deduced amino acid sequence of clone III includes Arg-Asp-Arg at positions 21–23, a sequence also deduced from a lung tryptase cDNA (16). This differs from the sequence at the same positions deduced for clones I and II (His-Gly-Pro) and the sequence at the same positions determined by N-terminal sequencing of human lung tryptase (His-Xaa-Pro, where Xaa is an unidentified residue) (Fig. 1).

Tryptase Gene. A blot of the *Bam*HI digests of the nine genomic clones hybridized with the tryptase clone II cDNA yielded four different banding patterns (unpublished data), suggesting the existence of multiple tryptase genes. *Bam*HI-digested DNA fragments from two of the clones were subcloned into pBluescript KS+ for sequence analysis. One of the clones encoded exons matching the sequence of tryptase

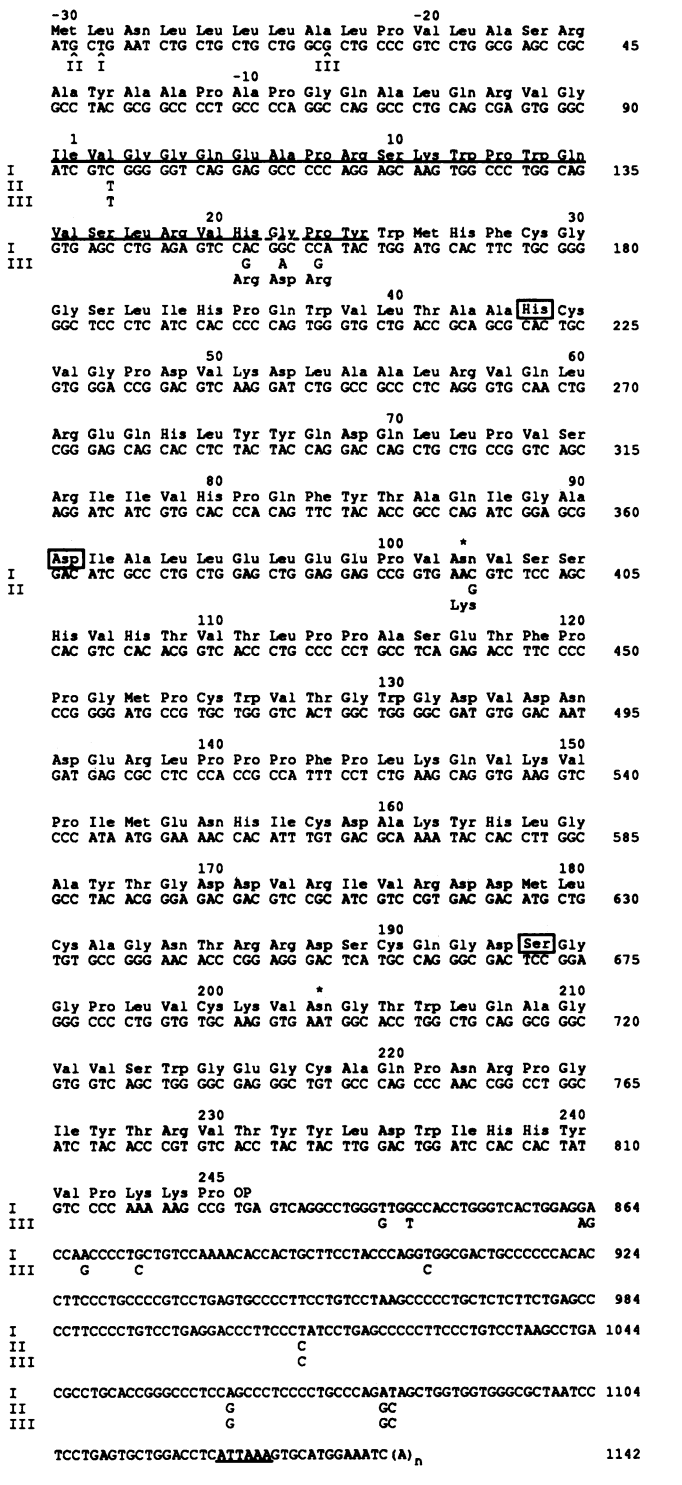

FIG. 1. Sequence of human tryptase cDNAs I, II, and III. Nucleotides are numbered to the right of the figure, and deduced amino acids are numbered above the amino acid sequence. The 5′ end of each cDNA is indicated by a caret with additional 5′ nucleotides inferred from the gene sequence shown in Fig. 2. Differences in sequence between tryptase clone I and clones II or III are indicated below the sequence of tryptase I. The N-terminal amino acid sequence determined for purified human lung tryptase is underlined. The amino acid at the position of Gly-22 (dashed underline) could not be identified by Edman degradation. The histidine, aspartic acid, and serine residues of the catalytic triad are boxed. Consensus N-glycosylation sites are labeled with an asterisk, and the putative polyadenylylation signal sequence is underlined.

clone I cDNA and was sequenced in its entirety (Fig. 2). Partial sequence determined for the second clone revealed 98% identity to the tryptase I gene over the regions sequenced (corresponding to bases −104 to 240 and 388–564 in Fig. 2), but it did not contain exons exactly matching any of the three skin tryptase cDNAs or the lung tryptase cDNA (16).

The tryptase I gene contains five introns separating six exons. All of the *Bam*HI sites of the genomic fragments overlap the tryptase cDNA sequence except the site beginning at base 712 in intron 3. The size of this intron was confirmed by amplifying the region of the intact genomic clone containing the intron by the polymerase chain reaction (39). The locations of introns 2–5 were deduced by comparison of the gene sequence with that of the cDNA. The placement of intron 1 was determined by comparison of the gene sequence with the 5′ noncoding region of a lung tryptase cDNA (16). The 9 nucleotides 5′ to the initiation codon of lung tryptase cDNA are identical to those found 5′ to intron 1. In addition, the intron–exon junction of intron 1 contains 7 of 8 nucleotides of the splice site consensus sequence (40) at the 5′ end and 13 of 15 nucleotides at the 3′ end. The splice sites flanking all of the introns conform to the GT/AG rule (41).

Genomic DNA Blotting. A 345-bp *Ava* II/*Bam*HI DNA fragment from the tryptase gene was used to probe a human genomic Southern blot (Fig. 3 *Upper*). The probe contained no internal restriction sites for *Ava* II, *Bal* I, or *Bam*HI, which were used to digest the genomic DNA. The blot was hybridized to the gene fragment and washed under highly stringent conditions. The 0.4-kb *Ava* II fragment, the 1.5-kb *Bal* I fragment, and the 0.8-kb *Bam*HI fragment predicted from the sequence of the cloned gene appear to be present in the digests as expected (Fig. 3 *Lower*). At least two additional

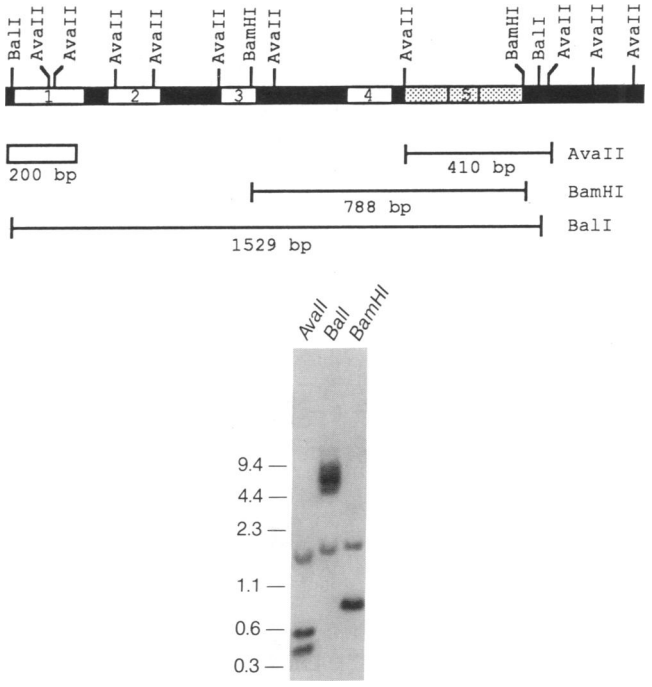

FIG. 3. (*Upper*) Restriction map of tryptase I gene. The box represents the tryptase I gene with restriction sites indicated above the gene. The black regions represent exons, and white regions represent introns (numbered 1–5). The shaded region indicates the portion of the gene used to hybridize to the Southern blot shown in *Lower*. The size of the *Ava* II, *Bam*HI, and *Bal* I restriction fragments predicted to hybridize to the gene fragment is indicated below the gene. (*Lower*) Southern blot of human genomic DNA. An autoradiogram is shown of human genomic DNA digested with the restriction endonucleases indicated, electrophoresed in agarose, transferred to a nylon membrane, and hybridized to a radiolabeled fragment of the tryptase gene. The size (kb) and location of molecular weight markers are shown to the left.

FIG. 2. Sequence of human tryptase I gene. Nucleotides are numbered to the right of the figure; the nucleotide located 25 bases 3′ to the putative TATA box sequence is designated +1 [(marked with a caret) based on the consensus transcription start-site sequence and position relative to the TATA box (37, 38)]. The introns, labeled 1–5, are in lower case letters. The deduced amino acids are above the exon nucleotide sequences and are numbered as in Fig. 1. Putative TATA and CAAT box sequences and the region homologous to a mast cell chymase enhancer element are underlined. The poly(A) attachment site is indicated with a slash (/).

bands appear in the *Ava* II and *Bal* I digests. Even accounting for allelic variation, these results, in conjunction with the presence of multiple cDNAs and genomic clones, indicate the presence of at least two but probably three or more tryptase genes in the human genome.

DISCUSSION

We have identified three distinct tryptase cDNAs from a human skin library, each different from a lung tryptase cDNA recently described (16). We also isolated and sequenced a tryptase gene whose organization is distinct from previously described serine proteases, and we show evidence of additional closely related genes. These findings provide the first evidence of a genetic basis for heterogeneity among mast cell tryptases.

Translation of tryptase mRNA probably begins at the first 5′ ATG, which is preceded by the sequence 5′-TGGCCAGG-3′. This sequence conforms to the consensus start site sequence of eukaryotic mRNAs (42) and to the consensus sequence preceding the initiation codons of mRNAs encoding proteoglycan core protein and other mast cell proteins of the secretory granule (43). Tryptase mRNA encodes a 30-amino acid preprosequence preceding Ile-1 identified to be the N terminus of the mature protein. The hydrophobic N-terminal portion of the prepropeptide is characteristic of a signal sequence. Empiric rules (44) do not clearly identify the C terminus of the signal peptide. Because propeptides tend to be more highly conserved than signal peptides among groups of serine proteases, the finding that the last 11 amino acids of the preprosequences encoded by our human and dog tryptase cDNAs (15) are identical suggests that the C-terminal portion of the preprosequence functions as a propeptide. An intriguing feature of the proposed propeptide is its termination in glycine, the amino acid also found in the sequence deduced from a tryptase-related dog mast cell cDNA (15) and from a lung tryptase cDNA (16), but which is otherwise unique. In serine proteases such as the pancreatic zymogens, the prosequence serves as an activation peptide, which is removed by activating proteases that recognize specific propeptide sequences. The propeptide hydrolysis site is a basic amino acid in the pancreatic serine proteases and apparently is an acidic residue in most leukocyte and mast cell granule-associated serine proteases (15), including mast cell chymase, which is packaged in mast cell granules with tryptase (45, 46). Thus the C-terminal glycine of the proposed tryptase propeptide suggests that tryptase may be activated by proteases that differ from those that activate chymase and other known types of serine proteases (15).

The skin tryptase cDNAs, like dog tryptase cDNA (15), encode 245-amino acid catalytic domains in contrast to the 244-amino acid domain encoded by a human lung tryptase cDNA (16). The catalytic domains of human skin tryptases I, II, and III are each 78% identical to dog tryptase and are 91%, 91%, and 92% identical, respectively, to the human lung tryptase. Unlike human skin tryptase clones I and III and lung tryptase cDNA, skin clone II and dog tryptase contain one rather than two consensus N-glycosylation sites. Thus, variable N-glycosylation of different tryptases or tryptase subunits probably contributes to the size heterogeneity seen on SDS/PAGE (3, 5, 6, 17).

The ≈1.8-kb tryptase I gene contains six exons separated by five small introns. The most unique feature of the tryptase gene organization is the placement of intron 1 (Fig. 4). An intron has been observed at this site in the dog pancreatic lipase gene (61) but not in other serine protease genes. Like most other serine protease genes, the catalytic triad residues are encoded by separate exons. The placement and phases (47) of the four introns within the coding region of the gene are analogous to those in the genes of trypsin (48) and glandular kallikrein (49) but are more dissimilar to those of other genes encoding serine proteases of hemostasis and of leukocyte and mast cell granules, including mast cell chymase (RMCP II) (50). Assuming descendence from a common ancestral gene, the tryptase gene may have diverged from that of trypsin more recently than from the genes of other granule-associated leukocyte serine proteases.

The 5′ flanking region of the tryptase gene (Fig. 2) is likely to contain regions that regulate the exceptionally high levels of mast cell-specific expression (as much as 23% of mast cell protein) (3). In addition to the TATA (37) and CAAT (38) box sequences, the 5′ flanking region contains a sequence (nucleotides −200 to −181) that shares 14 of 20 nucleotides in common with the consensus enhancer element of protease genes expressed in the pancreas (62, 63). An analogous sequence is contained in a region 5′ to the rat mast cell chymase gene that can function as an enhancer element and bind specifically to mast cell trans-acting factors (64). The availability of the 5′ flanking region of the tryptase gene will

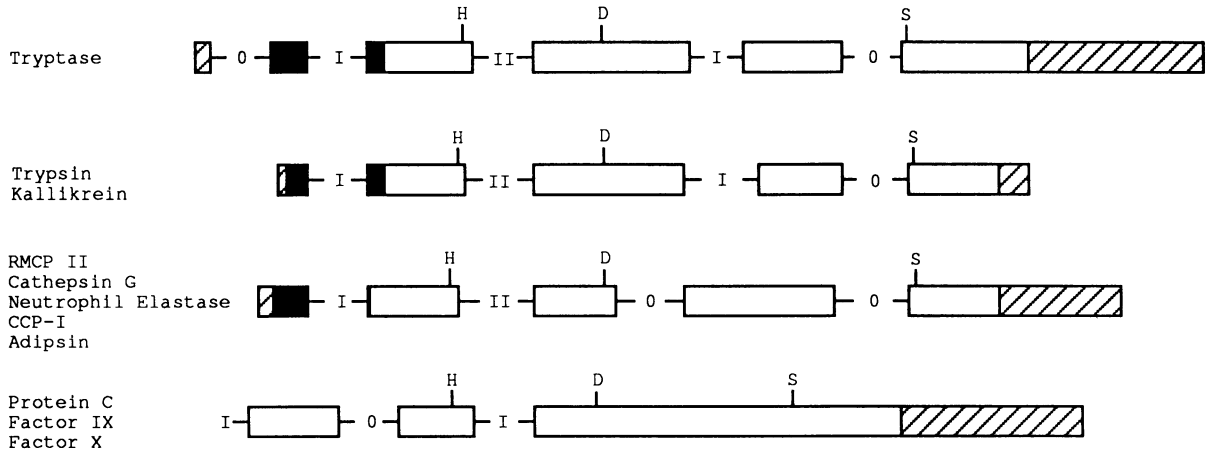

FIG. 4. Comparison of the gene structure of tryptase with that of other serine proteases. Exons are represented by open boxes, and the introns (not to scale), by lines. Putative prepropeptides are indicated by black boxes, and 5′ and 3′ noncoding regions, by hatched boxes. The location of the codons for the catalytic triad residues, histidine (H), aspartic acid (D), and serine (S), are indicated. The exons are drawn to the scale of the first member of each group [ie., tryptase, trypsin, RMCP II (rat mast cell proteinase II), and protein C]. Protease genes with homologous intron/exon organizations are listed below their respective homologue. Only the gene structure of the catalytic domain of protein C, factor IX, and factor X is shown. The phases of the introns (0, I, or II) (47) are indicated. Gene structures were derived from the following: trypsin (48), kallikrein (49), RMCP II (50), cathepsin G (51), neutrophil elastase (52), CCP-I (cytotoxic T cell-specific protein I) (53), adipsin (54, 55), protein C (56, 57), factor IX (58, 59), and factor X (60).

facilitate the identification of elements involved in mast cell-specific expression.

The multiple bands hybridizing to the tryptase gene fragment on the genomic blot support the evidence provided by multiple tryptase cDNAs and genomic clones that different tryptase genes exist in the human genome. Different tryptases may be expressed by different populations of mast cells, just as mast cell chymases I and II in the rat are made exclusively by mast cells of connective tissue and mucosa, respectively (65). Although the finding of different tryptase cDNAs in skin and lung suggests the possibility of tissue-specific differences in the type of tryptase expressed, this remains to be proven. Some, or even all, of the four human tryptase cDNAs described to date may encode separate components of the tryptase tetramer. It is unlikely that the differences among the three skin cDNAs obtained from a single donor and the substantially different lung cDNA reported previously (16) are due solely to allelic variation. Therefore, mast cell tryptase joins other serine proteases such as glandular kallikrein (49), trypsin (48), and RMCP II (50), which are members of multigene families.

This work was supported by Grants HL-01736, HL-24136, and AM31901 from the National Institutes of Health and GM8608086 from the National Science Foundation. G.H.C. is an R.J.R. Nabisco Research Scholar and P.V. is a Parker B. Francis Fellow in Pulmonary Research.

1. Schwartz, L. B., Metcalfe, D. D., Miller, J. S., Earl, H. & Sullivan, T. (1987) *N. Engl. J. Med.* **316,** 1622–1626.
2. Wenzell, S. E., Fowler, A. A., III, & Schwartz, L. B. (1988) *Am. Rev. Respir. Dis.* **137,** 1002–1008.
3. Schwartz, L. B., Lewis, R. A. & Austen, K. F. (1981) *J. Biol. Chem.* **256,** 11939–11943.
4. Tanaka, T., McRae, B. J., Cho, K., Cook, R., Fraki, J. E., Johnson, D. A. & Powers, J. C. (1983) *J. Biol. Chem.* **258,** 13552–13557.
5. Smith, T. J., Hougland, M. W. & Johnson, D. A. (1984) *J. Biol. Chem.* **259,** 11046–11051.
6. Caughey, G. H., Viro, N. F., Ramachandran, J., Lazarus, S. C., Borson, D. B. & Nadel, J. A. (1987) *Arch. Biochem. Biophys.* **258,** 555–563.
7. Maier, M., Spragg, J. & Schwartz, L. B. (1983) *J. Immunol.* **130,** 2352–2356.
8. Schwartz, L. B., Bradford, T. R., Littman, B. H. & Wintroub, B. U. (1985) *J. Immunol.* **135,** 2762–2767.
9. Gruber, B. L., Schwartz, L. B., Ramamurthy, N. S., Irani, A. M. & Marchese, M. J. (1988) *J. Immunol.* **140,** 3936–3942.
10. Sekizawa, K., Caughey, G. H., Lazarus, S. C., Gold, W. M. & Nadel, J. A. (1989) *J. Clin. Invest.* **83,** 175–179.
11. Caughey, G. H., Leidig, F., Viro, N. F. & Nadel, J. A. (1988) *J. Pharmacol. Exp. Ther.* **244,** 133–137.
12. Franconi, G., Graf, P. D., Lazarus, S. C., Nadel, J. A. & Caughey, G. H. (1989) *J. Pharmacol. Exp. Ther.* **248,** 947–951.
13. Schwartz, L. B., Lewis, R. A., Seldin, D. & Austen, K. F. (1981) *J. Immunol.* **126,** 1290–1294.
14. Caughey, G. H., Lazarus, S. C., Viro, N. F., Gold, W. M. & Nadel, J. A. (1988) *Immunology* **63,** 339–344.
15. Vanderslice, P., Craik, C. S., Nadel, J. A. & Caughey, G. H. (1989) *Biochemistry* **28,** 4148–4155.
16. Miller, J. S., Westin, E. H. & Schwartz, L. B. (1989) *J. Clin. Invest.* **84,** 1188–1195.
17. Cromlish, J. A., Seidah, N. G., Marcinkiewicz, M., Hamelin, J., Johnson, D. A. & Chretien, M. (1987) *J. Biol. Chem.* **262,** 1363–1373.
18. Schwartz, L. B. & Bradford, T. R. (1986) *J. Biol. Chem.* **261,** 7372–7379.
19. Alter, S. C., Metcalfe, D. D., Bradford, T. R. & Schwartz, L. B. (1987) *Biochem. J.* **248,** 821–827.
20. Irani, A. A., Schechter, N. M., Craig, G., DeBlois, G. & Schwartz, L. B. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 4464–4468.
21. Harvima, I. T., Schechter, N. M., Harvima, R. J. & Fraki, J. E. (1988) *Biochem. Biophys. Acta.* **957,** 71–80.
22. Goldstein, S. M., Kaempfer, C. K., Proud, D., Schwartz, L. B., Irani, A. M. & Wintroub, B. U. (1987) *J. Immunol.* **139,** 2724–2729.
23. Auffray, C. & Rougeon, F. (1980) *Eur. J. Biochem.* **107,** 303–314.
24. Aviv, H. & Leder, P. (1972) *Proc. Natl. Acad. Sci. USA* **69,** 1408–1412.
25. Hyunh, T. V., Young, R. A. & Davis, R. W. (1985) in *DNA Cloning Techniques: A Practical Approach*, ed. Glover, D. (IRL, Oxford), pp. 49–78.
26. Bullock, W. O., Fernandez, J. M. & Short, J. M. (1987) *BioTechniques* **5,** 376–379.
27. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) in *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY), 1st Ed.
28. Short, J. M., Fernandez, J. M., Huse, W. D. & Sorge, J. A. (1988) *Nucleic Acids Res.* **16,** 7583–7600.
29. Birnboim, H. C. & Doly, J. (1979) *Nucleic Acids Res.* **7,** 1513–1523.
30. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
31. Chen, E. Y. & Seeburg, P. H. (1985) *DNA* **4,** 165–170.
32. Tabor, S. & Richardson, C. C. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 4767–4771.
33. Gebeyehu, G., Rao, P. Y., Soochan, P., Simms, D. A. & Klevan, L. (1987) *Nucleic Acids Res.* **15,** 4513–4534.
34. Southern, E. M. (1975) *J. Mol. Biol.* **98,** 503–517.
35. Bøyun, A. (1968) *Scand. J. Clin. Lab. Invest.* **21** Suppl. 97, 77–89.
36. Yen, P. H., Marsh, B., Mohandas, T. K. & Shapiro, L. J. (1984) *Somatic Cell Mol. Genet.* **10,** 561–571.
37. Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C. & Chambon, P. (1980) *Science* **209,** 1406–1414.
38. Benoist, C., O'Hare, K., Breathnach, R. & Chambon, P. (1980) *Nucleic Acids Res.* **8,** 127–142.
39. Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. & Erlich, H. A. (1988) *Science* **239,** 487–491.
40. Padgett, R. A., Grabowski, P. J., Konarska, M. M., Seiler, S. & Sharp, P. A. (1986) *Annu. Rev. Biochem.* **55,** 1119–1150.
41. Breathnach, R. & Chambon, P. (1981) *Annu. Rev. Biochem.* **50,** 349–383.
42. Kozak, M. (1984) *Nucleic Acids Res.* **12,** 857–872.
43. Avraham, S., Stevens, R. L., Nicodemus, C. F., Gartner, M. C., Austen, K. F. & Weis, J. H. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 3763–3767.
44. Von Heijne, G. (1986) *Nucleic Acids Res.* **14,** 4683–4690.
45. Craig, S. S., Schechter, N. M. & Schwartz, L. B. (1988) *Lab. Invest.* **58,** 682–691.
46. Caughey, G. H., Lazarus, S. C., Viro, N. F., Gold, W. M. & Nadel, J. A. (1988) *Immunology* **63,** 339–344.
47. Rogers, J. (1985) *Nature (London)* **315,** 458–459.
48. Craik, C. S., Choo, Q.-L., Swift, G. H., Quinto, C., MacDonald, R. J. & Rutter, W. J. (1984) *J. Biol. Chem.* **259,** 14255–14264.
49. Mason, A. J., Evans, B. A., Cox, D. R., Shine, J. & Richards, R. I. (1983) *Nature (London)* **303,** 300–307.
50. Benfey, P. N., Yin, F. H. & Leder, P. (1987) *J. Biol. Chem.* **262,** 5377–5384.
51. Hohn, P. A., Popescu, N. C., Hanson, R. D., Salvesen, G. & Ley, T. J. (1989) *J. Biol. Chem.* **264,** 13412–13419.
52. Takahashi, H., Nukiwa, T., Yoshimura, K., Quick, C. D., States, D. J., Holmes, M. D., Whang-Peng, J., Knutsen, T. & Crystal, R. G. (1988) *J. Biol. Chem.* **263,** 14739–14747.
53. Lobe, C. G., Upton, C., Duggan, B., Ehrman, N., Letellier, M., Bell, J., McFadden, G. & Bleackley, R. C. (1988) *Biochemistry* **27,** 6941–6946.
54. Min, H. Y. & Spiegelman, B. M. (1986) *Nucleic Acids Res.* **14,** 8879–8892.
55. Phillips, M., Djian, P. & Green, H. (1986) *J. Biol. Chem.* **261,** 10821–10827.
56. Foster, D. C., Yoshitake, S. & Davie, E. W. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 4673–4677.
57. Plutzky, J., Hoskins, J. A., Long, G. L. & Crabtree, G. R. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 546–550.
58. Yoshitake, S., Schach, B. G., Foster, D. C., Davie, E. W. & Kurachi, K. (1985) *Biochemistry* **24,** 3736–3750.
59. Anson, D. S., Choo, K. H., Rees, D. J. S., Giannelli, F., Gould, K., Huddleston, J. A. & Brownlee, G. G. (1984) *EMBO J.* **3,** 1053–1060.
60. Leytus, S. P., Foster, D. C., Kurachi, K. & Davie, E. W. (1986) *Biochemistry* **25,** 5098–5102.
61. Mickel, F. S., Weidenbach, F., Swarovsky, B., LaForge, K. S. & Scheele, G. A. (1989) *J. Biol. Chem.* **264,** 12895–12901.
62. Ornitz, D. M., Palmiter, R. D., Hammer, R. E., Brinster, R. L., Swift, G. H. & MacDonald, R. J. (1985) *Nature (London)* **313,** 600–602.
63. Boulet, A. M., Erwin, C. R. & Rutter, W. J. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 3599–3603.
64. Sarid, J., Benfey, P. N. & Leder, P. (1989) *J. Biol. Chem.* **264,** 1022–1026.
65. Gibson, S. & Miller, H. R. P. (1986) *Immunology* **58,** 101–104.