Amino Acid Sequence of Rat Mast Cell Protease I (Chymase)†

Hai Le Trong, David C. Parmelee,‡ Kenneth A. Walsh, Hans Neurath, and Richard G. Woodbury*
Department of Biochemistry, University of Washington, Seattle, Washington 98195
Received May 15, 1987; Revised Manuscript Received June 23, 1987

ABSTRACT: The amino acid sequence has been determined for rat mast cell protease I (RMCP I), a product of peritoneal mast cells. The active enzyme contains 227 residues, including three corresponding to the catalytic triad characteristic of serine protease (His-57, Asp-102, and Ser-195 in chymotrypsin). A computer search for homology indicates 73% and 33% sequence identity of RMCP I with rat mast cell protease II from mucosal mast cells and bovine chymotrypsin A, respectively. When the structure of RMCP I is compared to those of cathepsin G from human neutrophils and two proteases expressed in activated lymphocytes, 48–49% of the sequences are identical in each case. RMCP I has six half-cystine residues at the same positions as in RMCP II, cathepsin G, and the two lymphocyte proteases, suggesting disulfide pairs identical with those reported for RMCP II. A disulfide bond near the active site seryl residue and substrate binding site, present in pancreatic and plasma serine proteases, is not found in RMCP I or in the other cellular proteases. These results indicate that RMCP I and other chymotrypsin-like proteases of granulocyte and lymphocyte origin are more closely related to each other than to the pancreatic or plasma serine proteases.

Mast cells isolated from the peritoneal cavity of rats contain a chymotrypsin-like serine protease first detected by Benditt and Arase (1959). Originally named chymase, this protease has more recently been referred to as rat mast cell protease I (RMCP I)[1] to distinguish it from a similar protease (RMCP II) produced by a distinct type of mucosal mast cell (Woodbury & Neurath, 1978). RMCP I is stored in secretory granules in an active form in association with the proteoglycan heparin. Although it is the most abundant protein in the mast cell granules (Lagunoff & Pritzl, 1976; Woodbury et al., 1981) and details of its substrate specificity have been reported (Yoshida et al., 1980; Powers et al., 1985; Le Trong et al., 1987), its physiological function(s) is (are) largely unknown. We have recently shown that the activity of RMCP I within the insoluble granule matrix is not released into solution and its action is therefore restricted to relatively small proteins that can penetrate the protease–heparin network (Le Trong et al., 1987).

RMCP II is better characterized at the molecular level than RMCP I. The covalent and three-dimensional structures of RMCP II have been determined (Woodbury et al., 1978; Reynolds et al., 1985). On the basis of its amino acid sequence and other information, it was proposed that RMCP II represents a distinct subclass of serine proteases produced by granulocytes and stored in an active form in secretory granules (Woodbury & Neurath, 1980). Several cDNAs recently isolated from cytotoxic T-cell libraries (Lobe et al., 1986; Schmid & Weissmann, 1987) encode serine proteases with structures more similar to RMCP II than to other serine proteases. The cDNA structure of human neutrophil cathepsin G indicates that this protease is another close relative of RMCP II (Salvesen et al., 1987). Thus, it appears that several proteases from granulocytes and lymphocytes have common structural features and form a closely related family.

A partial amino acid sequence analysis of RMCP I indicated that approximately 75% of the amino-terminal 35 residues were identical with those of RMCP II (Woodbury et al., 1978a).

The present study reports the determination of the complete amino acid sequence of RMCP I. The structure is compared to that of RMCP II and to those of other lymphocyte/granulocyte serine proteases in an attempt to document and clarify the structure/function relationships of these enzymes.

Materials and Methods

Rat mast cell protease I (RMCP I) was prepared as described by Everett et al. (1979). *N*-(*p*-Tosyl)-L-phenylalanine chloromethyl ketone–trypsin (TPCK-trypsin) was purchased from Millipore. Citraconic anhydride and BNPS-skatole were purchased from Pierce Chemical Co. Lysyl endopeptidase from *Achromobacter lyticus* was a gift from Dr. T. Masaki, Ibaraki University, Japan. All other reagents and solvents were commercially obtained as analytical or HPLC grades.

Reduction and Carboxymethylation. RMCP I (120 nmol) was reduced by incubation under nitrogen for 2 h at room temperature in 0.4 mL of 6 M guanidine hydrochloride 0.5 M Tris-HCl, 10 mM EDTA, and 45 mM dithiothreitol, pH 8.6. A freshly prepared solution of 100 mM iodoacetic acid in 100 μL of 1 N NaOH was then added, and the mixture was incubated in the dark at room temperature for 30 min. CM-RMCP I was purified by reversed-phase HPLC on an Ultrapore RPSC-C3 column (Altex).

Specific Cleavage of RMCP I. Cleavage at arginyl residues was achieved by digesting N-citraconylated protein with TPCK-trypsin. CM-RMCP I (20 nmol) was N-citraconylated (Atassi & Habeeb, 1972) in 0.5 mL of 6 M guanidine hydrochloride by addition of 16 μL of citraconic anhydride at room temperature over a period of 1 h while maintaining the pH at 8.8 with a pH-salt. The citraconylated protein was then extensively dialyzed against 0.05 M NH_4HCO_3 at pH 8.7. TPCK-trypsin was added in two aliquots over a 4-h period at

† This work was supported in part by grants from the National Institutes of Health (GM-15731 and HL-36114).
* Address correspondence to this author.
‡ Present address: Laboratory of Experimental Carcinogenesis, National Cancer Institute, NIH, Bethesda, MD 20892.

[1] Abbreviations: RMCP I, rat mast cell protease I; RMCP II, rat mast cell protease II; CM, *S*-(carboxymethyl); HPLC, high-performance liquid chromatography; Tris-HCl, tris(hydroxymethyl)aminomethane hydrochloride; EDTA, ethylenediaminetetraacetic acid.

0006-2960/87/0426-6988$01.50/0

37 °C [final protease:substrate ratio 2.2% (w/w)]. Citraconyl groups were removed by incubating the digest in 8.8% (v/v) formic acid for 5 h at 37 °C. The digestion mixture was applied to a tandem series of TSK 2000 SW columns equilibrated in 6 M guanidine hydrochloride and 10 mM phosphate, pH 6.0, and purified further on a C-18 column by reversed-phase HPLC.

Cleavage at lysyl residues was performed by digesting CM-RMCP I (18 nmol) with *Achromobacter* protease I (Masaki et al., 1981) at a protease:substrate ratio of 1:100 (w/w).

Cleavage at tryptophanyl residues was achieved by incubating 25 nmol of CM-RMCP I in 0.5 mL of 75% acetic acid containing 0.9 mg of phenol and 0.5 mg of BNPS-skatole at room temperature, under nitrogen, and in the dark for 46 h as described by Fontana et al. (1973). After treatment of the reaction mixture with 2-mercaptoethanol (4% v/v) at 37 °C for 5 h, it was diluted with 0.5 mL of water, and excess BNPS-skatole was removed by extraction with 1-chlorobutane.

Cleavage at methionine was carried out with cyanogen bromide in 70% formic acid (Gross, 1967). The reaction mixture was lyophilized and separated by size into three pooled fractions on Sephacryl S-200 in 6 M guanidine hydrochloride. Each pooled fraction was then resolved by HPLC on a Waters μBondapak C18 column.

Cleavage at aspartyl residues was achieved by incubating peptides in 2% formic acid for 3 h at 110 °C under vacuum (Inglis, 1983).

Peptides were initially separated by size-exclusion on a tandem series of either TSK-2000 SW (LKB) or TSK-3000 SW (Bio-Rad) columns equilibrated in 6 M guanidine hydrochloride and 10 mM sodium phosphate buffer, pH 6.0. Pooled fractions were then purified by reversed-phase HPLC with an Ultrapore RPSC-C3 column (Altex) or SynChropak RP-P (Synchrom) and eluted with an acetonitrile gradient in 0.1% trifluoroacetic acid (Mahoney & Hermodson, 1980).

Amino Acid Analysis and Automatic Sequence Analysis. Amino acid compositions were determined with a Waters Picotag system (Bidlingmeyer et al., 1984). Most of the sequence analyses were performed on a Beckman 890C spinning-cup sequencer according to Edman and Begg (1967), with polybrene (Tarr et al., 1978) and a program adapted from Brauer et al. (1975). Phenylthiohydantoins were identified by complementary reversed-phase HPLC systems described by Glajch et al. (1985) and Ericsson et al. (1977). Two peptides recovered in low amounts were analyzed in an Applied Biosystems sequencer as described by Hunkapiller et al. (1983).

The sequence of RMCP I was compared with those of other proteins with a VAX/VMS computer using the ALIGN program of Dayhoff et al. (1983).

Results

General Strategy. Approximately 90% of the amino acid sequence of RMCP I was determined from the analysis of intact S-carboxymethylated protein and six nonoverlapping fragments recovered following cleavage of the protein at arginyl residues. The remaining sequence and overlaps were obtained by cleavage of CM-RMCP I at either lysyl, tryptophanyl, or methionyl residues. Subdigestion of a fragment obtained from cleavage at tryptophanyl residues generated further cleavage at aspartyl residues. A summary of this strategy is outlined in Figure 1.

Cleavage at Arginyl Bonds. N-Citraconylated RMCP I was cleaved specifically at arginyl residues with TPCK-trypsin. The mixture of peptides was first separated into pooled fractions (A–D) according to size by gel filtration on two TSK-2000 SW columns in tandem (Figure 2). Each pool was further fractionated on a C18 column by reversed-phase HPLC. The peptides were completely resolved and recovered in yields ranging from 15 to 60% (Table I). The amino acid composition and amino acid sequence of each recovered peptide were determined as shown in Table I and Figure 1.

The largest peptides were recovered from pooled fractions A and B. Pooled fraction A contained a single peptide R6 (M_r 8534) in 40% yield, comprising residues 50–124. Pooled fraction B also contained a single peptide Rγ (M_r 3579) which contained two Arg residues and overlapped R11 and R12, recovered from other fractions (Table I).

Pooled fraction C contained a set of peptides ranging in molecular weight from 2500 to 3500, which were separated completely by reversed-phase HPLC on an RP-P column. A small amount of peptide Rγ was recovered from this fraction in addition to four new peptides (Figure 2). Peptide Rα contained two internal Arg–Pro bonds that were resistant to tryptic cleavage. Peptide Rβ, the product of inefficient cleavage at an internal Arg–Glu bond at 142–143, was obtained in 16% yield.

Pool D contained five peptides, one of which, R7, was found in each of two peaks (Figure 2). Partial oxidation of tryptophan may account for this anomalous chromatographic heterogeneity. A similar anomaly is seen for peptide R9 in pool C (Figure 2), probably due to partial oxidation of a methionine residue.

Cleavage at Lysyl Bonds. Fragments of CM-RMCP I, digested with the lysine-specific protease from *Achromobacter*, were completely separated in good yield by a combination of size-exclusion and reversed-phase HPLC (Figure 3A). The largest peptide, K20 (residues 180–221), was recovered in a yield of 40%. Peptide K14, containing both a methionyl residue and a tryptophanyl residue, was isolated in two forms by reversed-phase HPLC (data not shown). Although several other peptides were recovered, only the five peptides illustrated were used for sequence analysis.

Cleavage at Tryptophanyl Bonds. Peptides resulting from cleavage of CM-RMCP I at tryptophan were separated on TSK-3000 SW columns as shown in Figure 2B. Two large peptides, W1 (residues 1–126) and W2 (residues 127–218), were then purified by reversed-phase HPLC on an RSPC-C3 column.

Peptide W1 was cleaved specifically at aspartyl residues with dilute formic acid (Inglis, 1983). The reaction mixture was separated on an RP-P column to obtain the pure peptide W1-D5.

Sequence Analysis. Edman degradation of intact CM-RMCP I (5 nmol) was conducted and interpreted through 64 cycles. Four residues were tentatively assigned, and five others were not identified. These residues were unambiguously identified by sequence analysis of peptides obtained from other digests. For example, sequence analysis of peptides R4 and K3 identified residues 27 and 48 as Lys and Gly, respectively. Peptide K2 confirmed the sequence of residues 28–46. Analysis of peptide R6 provided evidence for the other unidentified residues and overlapped the amino-terminal region by 14 residues, thus extending the sequence to Val-105 and overlapping K11.

The sequence of a subpeptide, W1-D5, overlapped those of K11 and K14. Analysis of the large peptide W2 yielded 47 residues that overlapped K14 and R9, providing a continuous sequence to Arg-172. Cleavage of CM-RMCP I at methionyl residues with cyanogen bromide generated five principal fragments, only two of which are illustrated in Figure 1. The

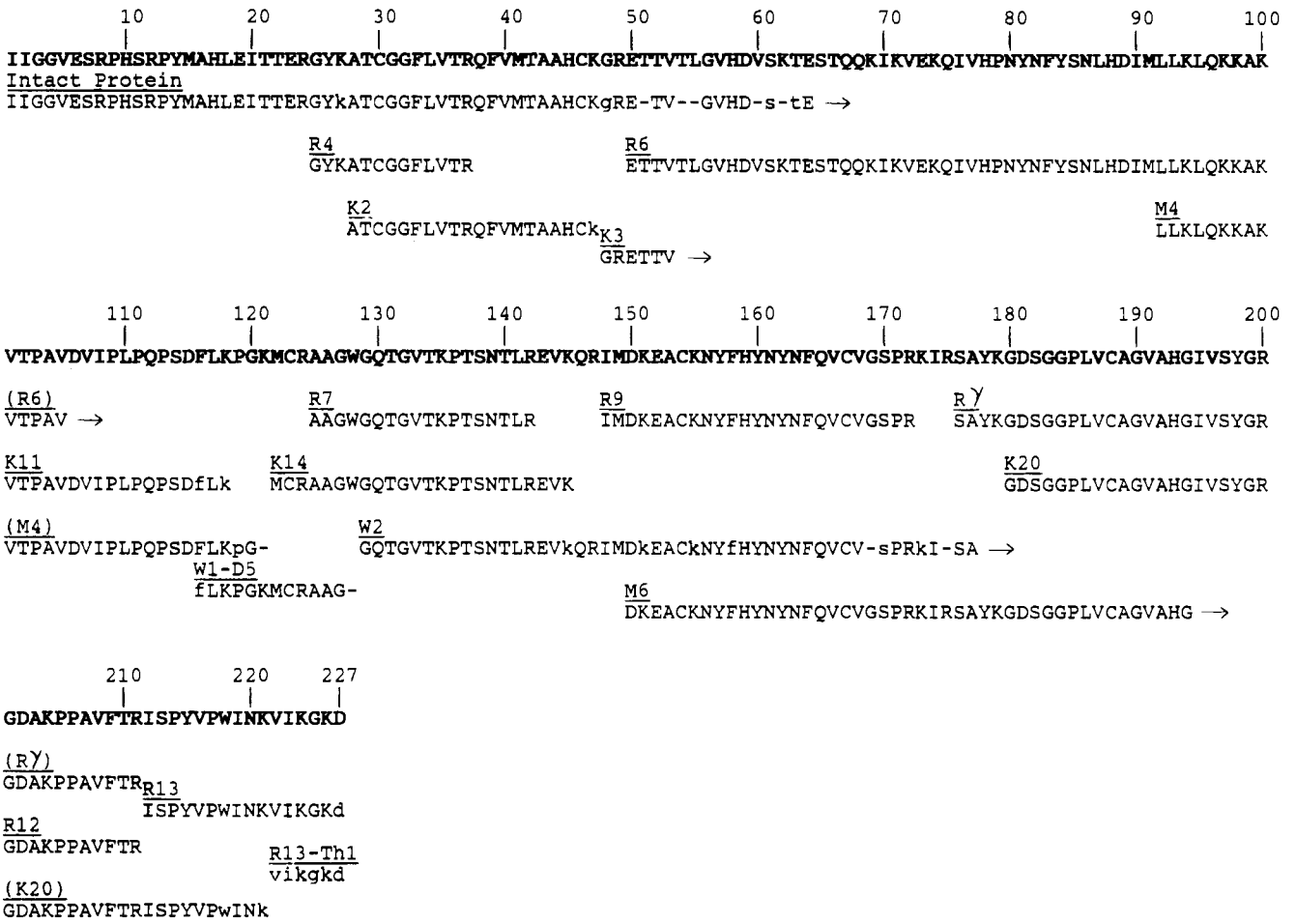

FIGURE 1: Summary of the proof of sequence of RMCP I. The proven sequences of specific peptides (names are underlined) are given in single-letter code below the summary sequence (bold type). The prefixes R, K, W, and M denote peptides generated by cleavage at arginyl, lysyl, tryptophanyl, and methionyl residues, respectively. Peptide W1-D5 was obtained following the subdigestion in dilute formic acid of peptide W1. Peptide Rγ was obtained due to the resistance of the Arg-200–Gly-201 bond to cleavage by trypsin. A lower-case letter indicates that the identification of the residue in that particular sequence analysis was tentative. A dash designates a residue within a peptide that could not be identified. An arrow indicates that the C-terminal residue of the peptide was not identified. Peptide M6 was obtained after cleavage of CM-RMCP I with cyanogen bromide (see Materials and Methods). Peptide R13-Th1 was analyzed by composition only (see text).

sequence of M6 extended the sequence from Arg-172 to Gly-194. Another fragment (M4) confirmed Phe-116.

The remaining sequence was determined as follows. Incomplete cleavage at Arg-200 yielded peptide Rγ (residues 176–211). Together with K20 and R13 this set of peptides yielded a continuous sequence from Ser-176 to Lys-226. Peptide R13 must correspond to the C-terminus of the protein because the peptide was generated by arginyl cleavage and it was the only peptide lacking Arg.

The carboxyl-terminal residue was established to be Asp-227 on the basis of the following data. Sequence analysis of R13 gave a relatively poor yield of the phenylthiohydantoin of aspartic acid in the 16th cycle. Although this observation fit the estimate of two Asx in R13 (Table I), additional evidence was sought. Peptide R13 was subdigested with thermolysin [1:50 (w/w) enzyme:substrate] for 6 h at room temperature. After separation by HPLC, a six-residue peptide (R13-Th1) was found with the following amino acid composition: Val, 0.9; Ile, 1.0; Lys, 2.0; Gly, 1.2; Asp, 1.1. This provided confirmation of the assignment of the C-terminal Asp.

Discussion

The amino acid sequence of RMCP I was deduced largely from a set of arginyl-cleavage fragments that were overlapped by sequence analyses of the intact protein and of peptides generated by cleavage of the protein at lysyl, tryptophanyl, and methionyl residues. Two products of subdigestion, W1-D5 (residues 116–127) and R13-Th1 (residues 222–227), completed the proof (Figure 1). Sixty-five percent of the residues were identified at least twice on different peptides. The amino acid composition derived from the sequence of RMCP I is in reasonable agreement with that determined (Everitt & Neurath, 1979) in acid hydrolysates of the intact protein.

The amino acid sequence clearly establishes the homologous relationship of RMCP I to other mammalian serine proteases. A comparison of the sequences of RMCP I and bovine chymotrypsin A (Figure 4) indicates 33% identity. The residues comprising the characteristic "catalytic triad" in serine proteases (His-57, Asp-102, and Ser-195 in bovine chymotrypsin) are found in corresponding positions in RMCP I. The sequence of RMCP I is most similar to that of RMCP II (73% identity), whereas cathepsin G from human neutrophils (Salvesen et al., 1987), a mouse cytotoxic T-cell protease (CCP I; Lobe et al., 1986), and a protease in activated human lymphocytes (Schmid & Weissmann, 1987) each indicate 48–49% identity with RMCP I. The homology among these proteases of cellular origin clearly suggests that all but neutrophil elastase are more closely related to each other than to the pancreatic or plasma serine proteases (Table II).

Analysis of the alignment scores (Table II) indicates that the granulocyte/lymphocyte proteases possessing chymotrypsin-like specificity have diverged from each other less

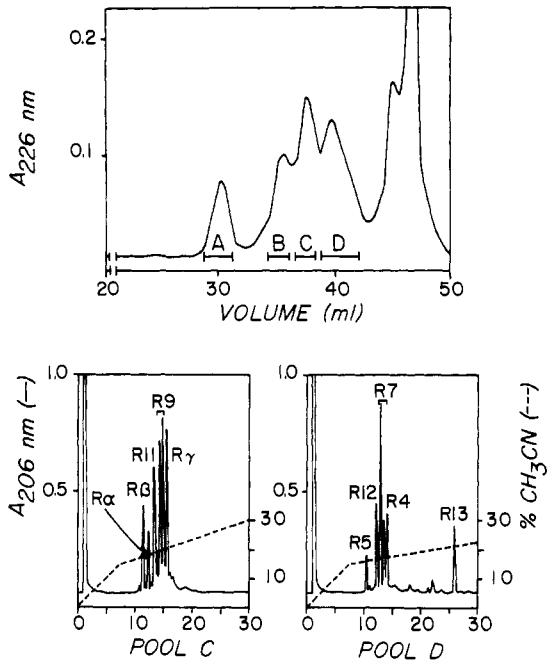

FIGURE 2: Separation of peptides after cleavage at arginyl residues. (Upper) The peptide mixture obtained by digestion of citraconylated CM-RMCP I (20 nmol) with TPCK-trypsin was fractionated on two TSK-2000 SW columns (7.5 × 600 nm) connected in tandem. The peptides were eluted with 6 M guanidine hydrochloride containing 10 mM sodium phosphate buffer (pH 6.0) at a flow rate of 0.5 mL/min. (Lower) Further fractionation of pools C and D performed on a SynChropak RP-P column (4.1 × 250 mm) equilibrated in 0.1% TFA and eluted at room temperature with a linear gradient of acetonitrile at a flow rate of 1.5 mL/min. Pools A and B each contained a single peptide, R6 and Rβ, respectively (not shown).

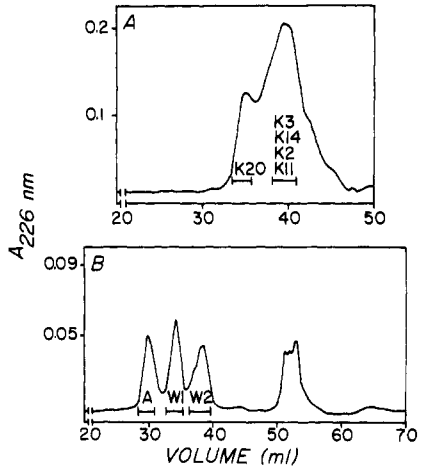

FIGURE 3: Primary separation of mixtures of peptides generated from the cleavage of CM-RMCP I: (A) at Lys (18 nmol of protein) with *Achromobacter* protease I; (B) at Trp (25 nmol of protein) with BNPS-skatole. The separations were performed in 6 M guanidine hydrochloride and 10 mM sodium phosphate buffer (pH 6.0) at a flow rate of 0.5 mL/min on two TSK-2000 SW columns connected in tandem (A) or two TSK-3000 SW columns (B). Each pool of fractions was desalted and further fractionated by reversed-phase HPLC on a C18 column equilibrated in 0.1% trifluoroacetic acid with a linear gradient of acetonitrile. Peptides from the digest at Lys residues were then purified on a SynChropak RP-P column. When more than one peptide was obtained from a single pool from the TSK columns, they are listed from top downward in order of increasing hydrophobicity (their elution order on reversed-phase HPLC). The peptides obtained from cleavage at tryptophanyl residues were purified on an Altex Ultrapore RPSC-C3 column (4.6 × 75 mm); Pooled fraction A contained uncleaved protein.

during evolution than have, for example, the pancreatic serine proteases. Cellular or functional constraints may have slowed

Table I: Amino Acid Compositions of Tryptic Peptides from S-Carboxymethylated, N-Citraconylated RMCP I[a]

	peptide													whole protein	
	Rα (1–24)	R4 (25–37)	R5 (38–49)	R6 (50–124)	R7 (125–142)	R8[b] (143–147)	Bβ (125–147)	R9 (148–172)	R10[b] (173–175)	R11 (176–200)	R12 (201–211)	Rγ (176–211)	R13 (212–227)	from sequence	from hydrolysis[c]
Asx (D/N)	3.1 (3)		0.9 (1)	6.9 (7)	1.0 (1)	(2)	0.8 (1)	3.9 (4)		1.0 (1)	0.7 (1)	2.1 (2)	2.1 (2)	16	16.1
Glx (E/Q)				8.5 (8)	1.3 (1)		3.7 (3)	2.0 (2)						17	17.9
CMC (C)		0.4 (1)	0.6 (1)	1.0 (1)				1.1 (2)		0.8 (1)		0.8 (1)		6	7.6
Ser (S)	2.1 (2)	0.4		4.0 (4)	1.1 (1)		1.6 (1)	1.1 (1)		2.3 (3)		2.6 (3)	1.2 (1)	12	12.1
Gly (G)	2.5 (2)	2.8 (3)	1.3 (1)	2.8 (2)	3.5 (3)		3.3 (3)	1.3 (1)		5.6 (6)	1.4 (1)	6.5 (7)	1.5 (1)	20	23.7
His (H)	1.7 (2)		1.0 (1)	2.3 (3)				1.0 (1)		1.1 (1)		1.2 (1)		8	8.2
Arg (R)	3.1 (3)	1.3 (1)	1.2 (1)	1.3 (1)	1.3 (1)	(1)	2.7 (2)	1.3 (1)	(1)	1.1 (1)	1.2 (1)	2.0 (2)		12	11.2
Thr (T)	2.5 (2)	2.0 (2)	1.1 (1)	5.7 (6)	3.1 (4)		3.3 (4)				1.2 (1)	1.3 (1)		16	16.1
Ala (A)	1.5 (1)	1.2 (1)	2.3 (2)	2.6 (2)	2.3 (2)		1.9 (2)	1.6 (1)		2.8 (3)	2.0 (2)	4.3 (5)		14	15.1
Pro (P)	0.8 (2)	0.5		5.9 (6)	1.3 (1)		1.5 (1)	1.3 (1)		1.2 (1)	1.4 (2)	3.0 (3)	2.2 (2)	15	14.5
Tyr (Y)	1.0 (1)	0.5 (1)		1.9 (2)				2.4 (3)		1.7 (2)		1.5 (2)	0.9 (1)	10	6.5
Val (V)	1.5 (1)	1.1 (1)	1.2 (1)	8.0 (8)	1.1 (1)	(1)	1.9 (2)	2.2 (2)		2.7 (3)	1.2 (1)	3.6 (4)	2.2 (2)	21	21.8
Met (M)	1.0 (1)		0.6 (1)	0.2 (2)				0.2 (1)						5	4.8
Ile (I)	2.4 (3)			3.9 (4)				1.1 (1)	(1)	1.2 (1)		1.2 (1)	2.8 (3)	13	12.2
Leu (L)	1.2 (1)	1.1 (1)		7.3 (7)	1.2 (1)		1.2 (1)			1.2 (1)		1.1 (1)		11	11.3
Phe (F)		0.7 (1)	1.1 (1)	2.1 (2)				1.7 (2)			1.0 (1)	1.2 (1)		7	6.5
Lys (K)		1.2 (1)	1.0 (1)	9.5 (10)	1.3 (1)	(1)	2.0 (2)	2.6 (2)	(1)	1.3 (1)	1.0 (1)	2.5 (2)	3.1 (3)	22	21.9
Trp (W)					ND (1)		ND (1)						ND (1)	2	1.6
total residues	24	13	12	75	18	5	23	25	3	25	11	36	16		
yield (%)	23	30	15	40	60		16	43		25	35	40	20		

[a] Recovered from pooled fractions in Figure 2. Residues per molecule; values of 0.3 or less are not reported; numbers in parentheses are derived from the sequence in Figure 1. Some Arg residues were resistant to digestion, and the resulting peptides containing more than one Arg are identified with Greek letter suffixes. [b] Small hydrophilic peptides were not recovered. [c] Recalculated from the analysis of Everitt and Neurath (1979) with a M_r of 25 076 determined from the amino acid sequence.

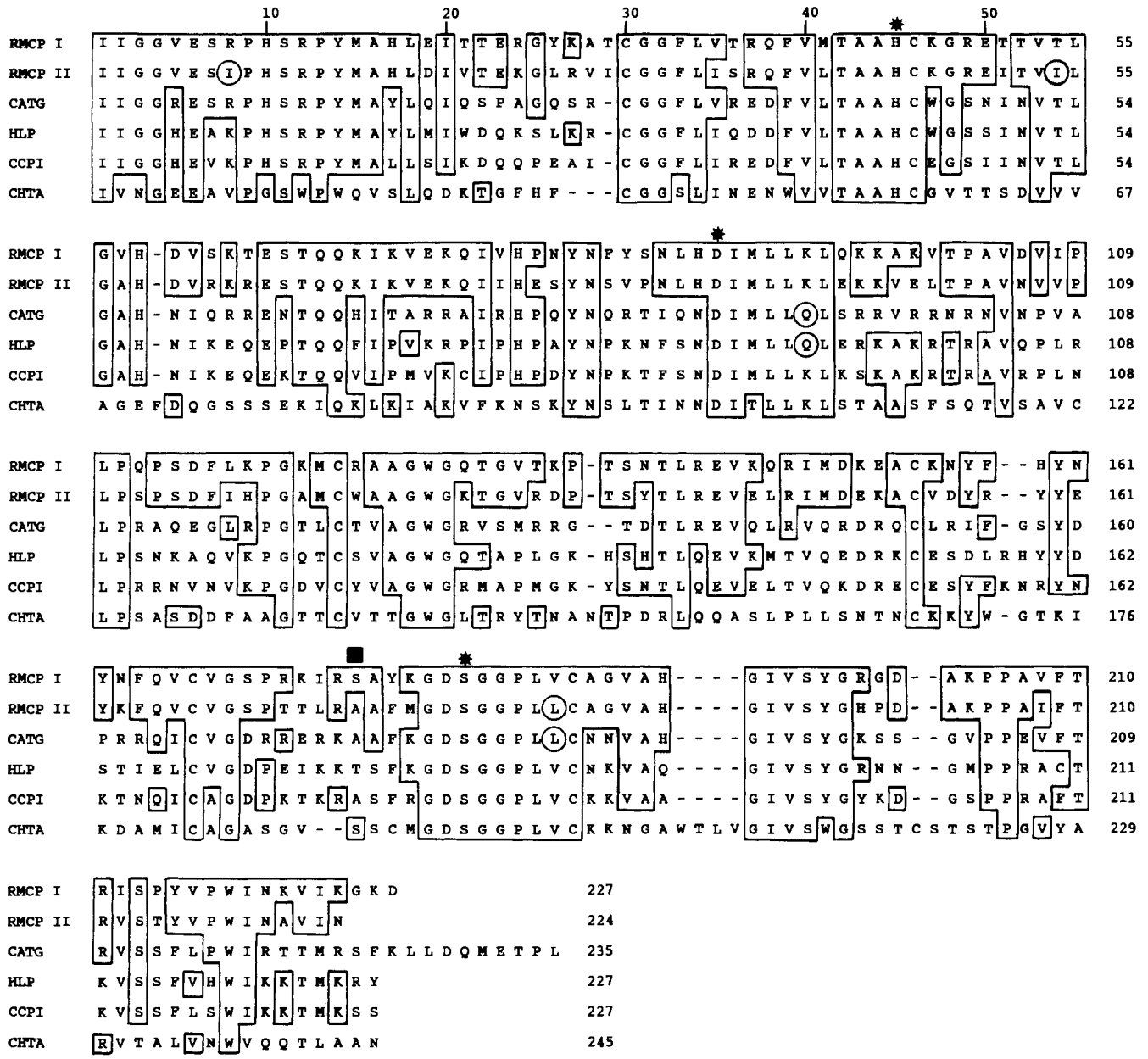

FIGURE 4: Homologous relationships among RMCP I (this study), RMCP II (Woodbury et al., 1978), CATG and human cathepsin G (Salvesen et al., 1987), HLP and human lymphocyte protease (Schmid & Weissman, 1987), CCPI and mouse cytotoxic T-cell specific protease I (Lobe et al., 1986), and CHTA and bovine chymotrypsin A (Hartley, 1970). Optimal alignments were determined with the ALIGN program of Dayhoff et al. (1983). Alignment scores are listed in Table II. Asterisks identify the components of the catalytic triad; a filled square identifies residue 176 in the putative binding pocket.

the rate of divergence of these cellular proteases; alternatively, the similarity of these proteases may indicate their relatively recent divergence from a common ancestral prototype.

Although the primary structures of RMCP I and RMCP II are strikingly similar, the proteases are clearly products of different genes. This has been confirmed by Benfey et al. (1987), who have obtained cDNA and genomic clones of RMCP II and a partial genomic clone of RMCP I. The structures obtained for the genomic clones indicate that RMCP I and RMCP II are representatives of a distinct gene family of serine proteases that, on the basis of hybridization experiments, appears to include 10–15 members (Benfey et al., 1987).

In addition to these sequence similarities, the rat mast cell proteases share other structural features that may relate to their functions. For example, the six half-cystinyl residues in RMCP I align precisely with three disulfide pairs in RMCP II, and both lack the invariant disulfide bond that characterizes the pancreatic and plasma serine proteases (Woodbury et al., 1978). This disulfide bond is also absent from human neutrophil cathepsin G (Salvesen et al., 1987) and from the newly discovered proteases from cytotoxic T cells and natural killer cells (Lobe et al., 1986; Schmid & Weissmann, 1987). In contrast, human neutrophil elastase (Sinha et al., 1987) and an enzyme with putative trypsin-like specificity in cytotoxic T cells (Gershenfeld & Weissmann, 1986) both contain this disulfide bond. The lack of this disulfide bond may be characteristic of granulocyte and lymphocyte proteases possessing chymotrypsin-like specificity. It should be noted that the human neutrophil elastase that contains the disulfide bond is stored in the same granule as cathepsin G that does not (Dewald et al., 1975).

As in the case of RMCP II, it appears that RMCP I is stored fully active within the secretion granules of mast cells. Analysis of the cDNAs of RMCP II, human cathepsin G, and two cytotoxic T-cell enzymes (Benfey et al., 1987; Salvesen

Table II: Alignment Scores[a] of Serine Proteases[b]

cellular	RMCP I	RMCP II	CATG	HLP	CCPI	human neutrophil elastase	CHTA	rat elastase	thrombin
cellular									
rat RMCP I									
rat RMCP II	51								
human CATG	28	39							
human HLP	32	29	36						
mouse CCPI	32	31	38	48					
human neutrophil elastase	21	21	18	19	18				
pancreatic									
bovine CHTA	20	20	18	19	21	18			
rat elastase	19	20	18	17	21	21	20		
plasma									
bovine thrombin	13	12	12	14	12	11	23	12	
bovine factor X	14	11	14	12	14	10	17	15	20

[a] Alignment scores (Dayhoff et al., 1983) are expressed as a number of standard deviation units from the mean for scrambled sequences. [b] Activation peptides were not included in the comparisons of catalytic domains. Abbreviated names follow the nomenclature in Figure 4. Sequences are as displayed in Figure 4 or derived as follows: neutrophil elastase (Sinha et al., 1987), rat pancreatic elastase (MacDonald et al., 1982), thrombin (Magnusson et al., 1975; Irwin et al., 1985), and factor X (Enfield et al., 1980).

et al., 1987; Lobe et al., 1986; Schmid & Weissmann, 1987) indicates that these proteases are synthesized with signal peptides, as one would expect for secreted proteins. It has been suggested that, after cleavage of the signal peptide, a dipeptide remains attached to the amino terminus of both RMCP II and cathepsin G (Benfey et al., 1987; Salvesen et al., 1987). These proteases may therefore exist as proenzymes from which dipeptides are removed during incorporation into secretory granules. Interestingly, the cDNA structure of RMCP II (Benfey et al., 1987) encodes also the tripeptide Thr-Ser-Ser at the C-terminus of the molecule that is absent from the mature enzyme (Woodbury et al., 1978). Similarly, mRNA for human cathepsin G encodes 12 more amino acids than are found at the C-terminus of the active enzyme (Salvesen et al., 1987). Completion of the genomic sequence of RMCP I will indicate if this enzyme is synthesized in a similar precursor form.

RMCP I has nine more lysyl residues than RMCP II, seven of which are located in the C-terminal half of the molecule. This difference may account for a striking difference in their ease of extraction from mast cells. RMCP II, which may be associated with chondroitin sulfate in the granules of mucosal mast cells (Stevens et al., 1986), is freely extracted with physiological saline solutions. In contrast, RMCP I, which interacts tightly with the more extensively sulfated heparin in peritoneal mast cell granules, requires 1 M salt for extraction (Lagunoff & Pritzl, 1976; Woodbury et al., 1981). The three-dimensional structure of RMCP II (Reynolds et al., 1985) displays a relatively uniform distribution of arginyl residues on its surface (S. J. Remington and R. G. Woodbury, unpublished data), but 10 of its 13 Lys residues are clustered on one side of the folded molecule. If it is assumed that the homologous RMCP I molecule is similarly folded, the arginine residues also appear uniformly distributed whereas there are two clusters of lysines, each of approximately 10 residues, on opposite sides of the molecule. It is reasonable to propose that the additional cluster may be related to the extra avidity for heparin and its physiological consequences.

Another structural difference between the two mast cell enzymes is the substitution at position 176 (of RMCP I) of a seryl residue for an alanyl residue in the putative substrate pocket of RMCP II. At the corresponding site, an Asp/Ser substitution accounts for the difference in specificity of trypsin and chymotrypsin (Hartley, 1970). Yoshida et al. (1980) and Powers et al. (1985) have demonstrated that RMCP I and RMCP II have similar substrate specificities toward *p*-nitroanilide peptide substrates. Both enzymes prefer substrates with extended hydrophobic structures; however, the specificity constants, k_{cat}/K_m, for RMCP I were generally several orders of magnitude larger than those for RMCP II. The values for RMCP I were similar to those of chymotrypsin, whereas for RMCP II they were generally closer to those of cathepsin G (Yoshida et al., 1980). These observations also correlate with the finding of Ser-176 in both RMCP I and chymotrypsin in contrast to Ala-176 in RMCP II and cathepsin G.

The amino acid sequence of RMCP I provides further evidence that several proteases of granulocyte and lymphocyte origin comprise a family of serine proteases that are distinct from those secreted by the pancreas and liver (Woodbury et al., 1978; Woodbury & Neurath, 1980). Additionally, it is clear from the comparison of the structure/function relationships of RMCP I and RMCP II that seemingly minor differences in structure in otherwise very similar proteases may have profound influences on the activity of the enzymes and their ultimate physiological role.

ACKNOWLEDGMENTS

We are grateful for the advice and suggestions of Drs. Thomas Marti and Harry Charbonneau during this work and for the assistance of Roger D. Wade, Maria Harrylock, and Lowell H. Ericsson with the amino acid, gas-phase sequencer, and homology analyses, respectively.

REFERENCES

Atassi, M. Z., & Habeeb, A. F. S. A. (1972) *Methods Enzymol. 25*, 546.

Benditt, E. P., & Arase, M. (1959) *J. Exp. Med. 110*, 451.

Benfey, P. N., Yin, F. H., & Leder, P. (1987) *J. Biol. Chem. 262*, 5377.

Bidlingmeyer, B. A., Cohen, S. A., & Tarvin, T. L. (1984) *J. Chromatogr. 336*, 93.

Brauer, A. W., Margolies, M. N., & Haber, E. (1975) *Biochemistry, 14*, 3029.

Dayhoff, M. O., Barker, W. C., & Hunt, L. T. (1983) *Methods Enzymol. 91*, 524.

Dewald, B., Rindler-Ludwig, R., Bretz, U., & Baggiolini, M. (1975) *J. Exp. Med. 141*, 709.

Edman, P., & Begg, G. (1967) *Eur. J. Biochem. 1*, 80.

Enfield, D. L., Ericsson, L. H., Fujikawa, K., Walsh, K. A., Neurath, H., & Titani, K. (1980) *Biochemistry 19*, 659.

Ericsson, L. H., Wade, R. D., Gagnon, J., MacDonald, R. M., Granberg, R. R., & Walsh, K. A. (1977) in *Solid Phase*

Methods in Protein Sequence Analysis (Previero, A., & Coletti-Previero, M. A., Eds.) p 137, Elsevier/North-Holland, New York.

Everitt, M. T., & Neurath, H. (1979) *Biochimie 61*, 653.

Fontana, A., Vito, C., & Toniolo, C. (1973) *FEBS Lett. 32*, 139.

Gershenfeld, H. K., & Weissmann, I. L. (1986) *Science (Washington, D.C.) 232*, 854.

Glajch, J. L., Gluckman, J. C., Charikofsky, J. G., Minor, J. M., & Kirkland, J. J. (1985) *J. Chromatogr. 318*, 23.

Gross, E. (1967) *Methods Enzymol. 11*, 238.

Hartley, B. (1970) *Philos. Trans. R. Soc. London, B 257*, 77.

Hunkapiller, M. W., Hewick, R. M., Dreyer, W. J., & Hood, L. E. (1983) *Methods Enzymol. 91*, 399.

Inglis, A. S. (1983) *Methods Enzymol. 91*, 324.

Irwin, D. M., Ahern, K. G., Pearson, G. D., & MacGillivray, R. T. A. (1985) *Biochemistry 24*, 6854.

Lagunoff, D., & Pritzl, P. (1976) *Arch. Biochem. Biophys. 173*, 554.

Le Trong, H., Neurath, H., & Woodbury, R. G. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 364.

Lobe, C. G., Finlay, B. B., Paranchych, W., Paetkau, V. H., & Bleackley, R. C. (1986) *Science (Washington, D.C.) 232*, 858.

MacDonald, R. J., Swift, G. H., Quinto, C., Swain, W., Pictet, R. L., Nikovits, W., & Rutter, W. J. (1982) *Biochemistry 21*, 1453.

Magnusson, S., Sottrup-Jensen, L., Peterson, T. E., & Claeys, H. (1975) in *Boerhaave Symposium on Prothrombin and Related Coagulation Factors* (Hemker, H. C., & Veltkamp, J. J., Eds.) p 25, Leiden University Press, Leiden, The Netherlands.

Mahoney, W. C., & Hermodson, M. A. (1980) *J. Biol. Chem. 255*, 11199.

Masaki, T., Tanabe, M., Nakamura, K., & Soejima, M. (1981) *Biochim. Biophys. Acta 660*, 44.

Powers, J. C., Tanaka, T., Harper, J. W., Minematus, Y., Barker, L., Lincoln, D., Crumley, K. V., Fraki, J. E., Schechter, N. M., Lazarus, G. G., Nakajima, K., Nakashino, K., Neurath, H., & Woodbury, R. G. (1985) *Biochemistry 24*, 2048.

Reynolds, R. A., Remington, S. J., Weaver, L. H., Fisher, R. G., Anderson, W. F., Ammon, H. L., & Matthews, B. W. (1985) *Acta Crystallogr., Sect. B: Struct. Sci. B41*, 139.

Salvesen, G., Farley, D., Shuman, J., Przybyla, A., Reilly, C., & Travis, J. (1987) *Biochemistry 26*, 2289.

Schmid, J., & Weissmann, C. (1987) *J. Immunol.* (in press).

Sinha, S., Watorek, W., Karr, S., Giles, J., Bode, W., & Travis, J. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 2228.

Stevens, R. L., Lee, T. D. G., Seldin, D. C., Austen, K. F., Befus, A. D., & Bienenstock, J. (1986) *J. Immunol. 137*, 291.

Tarr, G. E., Beecher, J. F., Bell, M., & McKean, D. J. (1978) *Anal. Biochem. 84*, 622.

Woodbury, R. G., & Neurath, H. (1978) *Biochemistry 17*, 4298.

Woodbury, R. G., & Neurath, H. (1980) in *Metabolic Interconversion of Enzymes* (Holzer, H., Ed.) p 145, Springer, Berlin.

Woodbury, R. G., Everitt, M., Sanada, Y., Katunuma, N., Lagunoff, D., & Neurath, H. (1978a) *Proc Natl. Acad. Sci. U.S.A. 75*, 5311.

Woodbury, R. G., Katunuma, N., Kobayashi, K., Titani, K., & Neurath, H. (1978b) *Biochemistry 17*, 811.

Woodbury, R. G., Everitt, M. T., & Neurath, H. (1981) *Methods Enzymol. 80*, 588.

Yoshida, N., Everitt, M. T., Neurath, H., Woodbury, R. G., & Powers, J. C. (1980) *Biochemistry, 19*, 799.