

Cloning and characterization of the novel gene for mast cell carboxypeptidase A.

D S Reynolds, ... , D S Gurley, K F Austen

J Clin Invest. 1992;89(1):273-282. <https://doi.org/10.1172/JCI115571>.

Research Article

No gene for a hematopoietic cell carboxypeptidase has previously been characterized. Mast cell carboxypeptidase A (MC-CPA) is a prominent secretory granule marker of mast cell differentiation and phenotype. The 32-kb human MC-CPA gene was isolated, localized to chromosome 3, and found to contain 11 exons. No significant homology was found between the 5' flanking region of the MC-CPA gene and those of three rat pancreatic carboxypeptidase genes (carboxypeptidase A1 and A2, and carboxypeptidase B [CPB]). In contrast, the intron/exon organization of the MC-CPA gene was conserved, most closely resembling the CPB gene. MC-CPA is unique among carboxypeptidases in having a CPA-like substrate-binding pocket and enzymatic activity despite overall protein and gene structures more similar to CPB. Evolutionary tree analysis of the carboxypeptidase gene family showed that, before the mammalian species radiation, a common MC-CPA/CPB ancestor diverged by gene duplication from the lineage leading to CPA, and then underwent another gene duplication to form separate but similar gene structures for MC-CPA and CPB. MC-CPA mRNA was prominent in dispersed lung cells enriched for mast cells but was undetectable in other nontransformed populations of several lineages, demonstrating that transcription of MC-CPA, a novel carboxypeptidase gene, provides a specific molecular marker for mast cells among normal hematopoietic cell populations.

Find the latest version:

<https://jci.me/115571/pdf>



Cloning and Characterization of the Novel Gene for Mast Cell Carboxypeptidase A

Dale S. Reynolds, Daniel S. Gurley, and K. Frank Austen

Department of Medicine, Harvard Medical School, and the Department of Rheumatology and Immunology, Brigham and Women's Hospital, Boston, Massachusetts 02115

Abstract

No gene for a hematopoietic cell carboxypeptidase has previously been characterized. Mast cell carboxypeptidase A (MC-CPA) is a prominent secretory granule marker of mast cell differentiation and phenotype. The 32-kb human MC-CPA gene was isolated, localized to chromosome 3, and found to contain 11 exons. No significant homology was found between the 5' flanking region of the MC-CPA gene and those of three rat pancreatic carboxypeptidase genes (carboxypeptidase A1 and A2, and carboxypeptidase B [CPB]). In contrast, the intron/exon organization of the MC-CPA gene was conserved, most closely resembling the CPB gene. MC-CPA is unique among carboxypeptidases in having a CPA-like substrate-binding pocket and enzymatic activity despite overall protein and gene structures more similar to CPB. Evolutionary tree analysis of the carboxypeptidase gene family showed that, before the mammalian species radiation, a common MC-CPA/CPB ancestor diverged by gene duplication from the lineage leading to CPA, and then underwent another gene duplication to form separate but similar gene structures for MC-CPA and CPB. MC-CPA mRNA was prominent in dispersed lung cells enriched for mast cells but was undetectable in other nontransformed populations of several lineages, demonstrating that transcription of MC-CPA, a novel carboxypeptidase gene, provides a specific molecular marker for mast cells among normal hematopoietic cell populations. (*J. Clin. Invest.* 1992. 89:273–282.) Key words: secretory granule • neutral protease • gene family • evolutionary tree • intron loss

Introduction

The mast cell carboxypeptidase A (MC-CPA)¹ enzymes of the rat, mouse, and human are homologous secretory granule metalloexopeptidases that have pH optima in the neutral to basic

Address correspondence to Dr. Dale S. Reynolds, Department of Rheumatology and Immunology, Brigham and Women's Hospital, The Seeley G. Mudd Building, Room 624, 250 Longwood Ave., Boston, MA 02115.

Received for publication 24 June 1991 and in revised form 28 August 1991.

1. Abbreviations used in this paper: CPA, carboxypeptidase A; CPA1 and CPA2, rat pancreatic carboxypeptidases A1 and A2; CPB, carboxypeptidase B; CTL, cytotoxic T lymphocyte; EBV-LCL, Epstein-Barr virus-transformed B lymphoblastoid cell line; MC-CPA, mast cell carboxypeptidase A; NK, natural killer; PCR, polymerase chain reaction; TcR, T cell receptor.

J. Clin. Invest.

© The American Society for Clinical Investigation, Inc.

0021-9738/92/01/0273/10 \$2.00

Volume 89, January 1992, 273–282

range (1–3). The MC-CPAs resemble bovine pancreatic carboxypeptidase A (CPA) in cleaving COOH-terminal aromatic and aliphatic amino acid residues, although human MC-CPA differs from bovine CPA by failing to cleave des-Arg bradykinin and substance P (4). MC-CPAs can cleave newly exposed COOH-terminal residues after endopeptidase cleavage by chymotryptic mast cell serine proteases, thereby sequentially degrading common protein substrates (5, 6). Immunohistochemical analysis of human mast cells from various tissue sites has shown that MC-CPA epitopes are segregated to connective tissue phenotype mast cells prominent in gut submucosa and skin (7), which also contain larger quantities of chymotryptic and tryptic serine proteases (8, 9). In transformed mouse mast cell lines, the content of MC-CPA enzymatic activity increases with differentiation toward the connective tissue mast cell phenotype (10). Immature bone marrow-derived mouse mast cells obtained *in vitro* contain little MC-CPA in their granules, but they exhibit marked transcription of the MC-CPA gene (11), and greatly increase MC-CPA translation in response to coculture with mouse 3T3 fibroblasts (2). MC-CPA mRNA expression in mouse mast cells obtained *in vivo* distinguishes among phenotypic subclasses in being high in connective tissue mast cells from the serosal cavity but absent to barely detectable in mucosal mast cell-rich tissue from the intestines of helminth-infected mice (12).

The complete sequence of the human MC-CPA mRNA (13) was determined by cloning the cDNA from a human lung library using a mouse MC-CPA cDNA probe (11), coupled with direct sequencing of the 5' end of the MC-CPA transcript in human lung poly(A)⁺ RNA. The complete deduced amino acid sequence of human MC-CPA revealed a 317 amino acid preproenzyme, which includes a 15 amino acid hydrophobic leader peptide, a 94 amino acid activation peptide, and a mature enzyme of 308 amino acids. Alignment of the deduced MC-CPA amino acid sequence with rat pancreatic carboxypeptidases A1 (CPA1; 14), A2 (CPA2; 15), and B (CPB; 14) showed overall amino acid sequence homologies of 43, 43, and 55%, respectively. The substrate-binding pocket residues of MC-CPA most closely resembled those of the pancreatic CPAs (13).

Here we report the cloning and restriction endonuclease mapping of the 32-kb human MC-CPA gene, its complete DNA sequence except for gaps in the four largest introns, and its localization to human chromosome 3. Because the MC-CPA gene is the first carboxypeptidase gene to be characterized for any hematopoietic cell in general, and for the mast cell in particular, its organization was compared with the pancreatic carboxypeptidase genes, and its molecular ancestry was examined by an evolutionary analysis of the carboxypeptidase gene family. We also show by transcription analysis that this novel mammalian carboxypeptidase gene is normally substantially expressed only in mast cells as contrasted with other hematopoietic cell populations.

Methods

Cloning and structural characterization of the human MC-CPA gene. A human genomic DNA clone containing the 5' end of the human MC-CPA gene was previously isolated from a human genomic DNA library in the modified λ EMBL-3 (SP6/T7) vector (HL1067J; Clontech Laboratories, Inc., Palo Alto, CA) (13). To obtain clones containing the complete human MC-CPA gene, this library was screened with two human MC-CPA cDNA probes corresponding to the middle and 3' thirds of the 1,630-bp human MC-CPA cDNA sequence. The middle-portion probe was a 585-bp 5'→EcoRV fragment of a partial-length cDNA, corresponding to residues 489–1,073 of the complete human MC-CPA consensus cDNA sequence. The 3' probe was a 571-bp EcoRV→3' cDNA fragment that lacked a poly(A) tail. These cDNAs were used sequentially to screen filters lifted from plating 1×10^6 recombinant phage. Filters were probed (11) at 37°C and washed under conditions of high stringency (58°C; 30 mM NaCl, 3 mM sodium citrate, 0.1% SDS, 1 mM EDTA, and 10 mM sodium phosphate, pH 7.0) to isolate 11 recombinants using the middle-portion cDNA probe, and 2 additional recombinants using the 3' probe. To obtain a preliminary map of the gene, restriction-digested λ DNA (16) for each clone was resolved by agarose gel electrophoresis and blotted to Zetabind (Cuno Inc., Meriden, CT). Blots were probed (17) sequentially with several 18-mer oligonucleotides radiolabeled with [γ - 32 P]dATP (6,000 Ci/mmol; DuPont-New England Nuclear, Boston, MA) using T4 polynucleotide kinase (New England Biolabs, Beverly, MA). Before reprobing, blots were stripped by washing in 30 mM NaCl and 3 mM sodium citrate at 80°C for 30 min. Based on preliminary mapping, four clones were chosen for detailed structural characterization. Overlapping restriction fragments of the inserts were subcloned into the plasmid Bluescript-SK-II⁺ (Stratagene Inc., La Jolla, CA) for DNA sequencing and fine restriction mapping. DNA sequencing was performed with synthesized oligonucleotide primers and a standard kit (Sequenase; United States Biochemical Corp., Cleveland, OH). All exons, 5'-flanking DNA, intron/exon junctions, and the majority of the intron sequence data presented were confirmed by sequencing both strands. Compressions were resolved by use of the Sequenase kit deoxyinosine triphosphate mixes and by sequencing the opposite strand.

Chromosomal localization. DNA isolated from normal human (150 ng/reaction) and hamster (150 ng/reaction) cells, and from a panel of 18 human/hamster somatic cell hybrid cell lines (350 ng/reaction) was purchased (CPD-100; BIOS Corp., New Haven, CT). Karyotype information for the hybrids was supplied by BIOS Corp. Polymerase chain reactions (PCR) without template DNA were performed to demonstrate the absence of overt contamination. Two synthetic 18-mer oligonucleotide primers (5'-GCCAGATACCCAAATTCA-3' and 5'-CACAGAGCTGGGATTTAC-3') were used at 0.2 μ M for amplification of a 129-bp region of 5'-flanking DNA (positions -399 to -271). MgCl₂ was used at 2.5 mM in 0.1 ml reactions under mineral oil. Other conditions were according to Perkin-Elmer Cetus Instruments (Norwalk, CT) instructions using a standard PCR kit. Reactions were denatured at 94°C for 1 min, annealed at 52°C for 1 min, and extended at 72°C for 1 min, for 35 cycles. PCR products were analyzed by electrophoresis of 30 μ l of each reaction in ethidium bromide-stained 3% SeaKem ME agarose (FMC Corp., Rockland, ME) gels, using HaeIII-digested PhiX-174 DNA (New England Biolabs) for size markers.

Amino acid sequence alignment. Carboxypeptidase amino acid sequences were obtained from a search of the biomedical literature and from the National Biomedical Research Foundation/Protein Identification Resource (NBRF/PIR) (release 25) and the Translated GenBank (release 64). The "fasta" search program of Pearson and Lipman (18) was used to search the above data bases for amino acid sequences having homology to human MC-CPA. Carboxypeptidase amino acid sequences were aligned with a multiple alignment algorithm (19), using a Dayhoff cost matrix with penalties of 2.5 for opening a gap and 0.5 for each residue in a gap. Minor adjustments in the generated alignment were performed by inspection to improve the alignment of introduced gaps. DNA sequence data compilation, database searches, and amino

acid sequence alignments were performed with Sequence Assembly Manager (SAM) and EUGENE software from the Molecular Biology Information Resource, Department of Cell Biology, Baylor College of Medicine (Houston, TX), at the Molecular Biology Computer Research Resource (MBCRR), Dana-Farber Cancer Institute (Boston, MA).

Cell lines, tissues, and purified cell populations. Cell lines were obtained from the American Type Culture Collection (Rockville, MD) unless otherwise noted, and were grown in enriched RPMI-1640 culture medium (10) with 10% fetal bovine serum. A primary human dermal fibroblast culture HF-15 (established by Dr. J. Moss, National Institutes of Health, Bethesda, MD) was provided by Dr. C. Ling, Harvard Medical School (Boston, MA). Whole lung tissue was obtained from the normal margins of fresh surgical pathology specimens, provided by Dr. D. Sugarbaker, Brigham and Women's Hospital (Boston, MA). Dispersed human lung cells were obtained from a 50-g human lung specimen by twice digesting minced tissue with pronase plus chymopapain, followed by two digestions with collagenase plus elastase (20). After each digestion for 30 min at room temperature the minced lung tissue was passed over a Nytex filter; the free cells in the filtrate were pelleted and washed in Tyrode's buffer. Total RNA was isolated from 1.3×10^7 cells from the third filtrate, which were ~10% mast cells containing granules that stained metachromatic with Toluidine blue. Human skin from a fresh surgical pathology specimen was minced with scissors on ice and homogenized in guanidine solution (21) in a blender for RNA isolation. PBMC (22), plastic-adherent monocytes (23), and neutrophils of >95% purity (22) were obtained from healthy donors, as described. Hypodense eosinophils from a patient with hypereosinophilia secondary to experimental IL-2 therapy for a malignancy were obtained by isolation of PBMC followed by culture in enriched medium plus 10 pM recombinant human GM-CSF to further select for eosinophils (24). On day 4 of culture 5×10^7 eosinophils of >98% purity were obtained as determined by Wright-Giemsa staining. Enriched IL-2-expanded natural killer (NK) cells and T cells were obtained by first partially depleting normal human PBMC of B cells and T cells using CD20- and CD5-specific monoclonal antibodies and immunomagnetic beads. The polyclonal NK cells were then activated and expanded by culture in medium containing 1 nM IL-2 for 6 d, and >95% pure CD56⁺CD6⁻ NK cells (4×10^6 cells) were isolated by flow cytometry and cell sorting. Polyclonal CD56⁻CD6⁺ T lymphocytes (>95% purity, 4×10^6 cells) from the same culture were also isolated by flow cytometry and cell sorting (Dr. M. Robertson, Dana-Farber Cancer Institute). Lectin-activated human peripheral blood T cells were obtained from CD6⁺ T cells (25) cultured with Con A in enriched medium for 7 d (Dr. P. Anderson, Brigham and Women's Hospital). Three cloned human CD8⁺CD4⁻ cytolytic T lymphocyte (CTL) lines were also examined. Two were mycoplasma extract specific, γ/δ T cell receptor (TcR)-expressing clones designated DG.SF37 and DG.SF68, harvested 10 d after stimulation with PHA (Dr. C. Morita, Dana-Farber Cancer Institute). One clone was an α/β TcR-expressing CTL line specific for HLA-B35, designated 8S1, harvested 10 d after stimulation with irradiated allogeneic EBV-transformed B lymphoblastoid cell line (EBV-LCL) cells (Dr. S. Porcelli, Dana-Farber Cancer Institute).

RNA blot analysis. Total RNA was extracted from cells and tissues (21), and poly(A)⁺ RNA was selected (26) where indicated. Poly(A)⁺ RNAs from human liver and brain were purchased (Clontech). Unless otherwise stated, each lane contained either 15 μ g of total RNA or 5 μ g of poly(A)⁺ RNA from the specified cell or tissue source. RNA samples were denatured in formaldehyde/formamide, electrophoresed in 1% formaldehyde-agarose gels, and blotted to Zetabind (27). An RNA blot containing 15 μ g/lane of total RNA from the human megakaryocytic cell line CMK (28) was provided by Dr. H. Avraham, New England Deaconess Hospital (Boston, MA). A 131-bp gene-specific MC-CPA cDNA probe (13) was labeled by random priming to a specific activity of $5\text{--}8 \times 10^8$ cpm/ μ g with [α - 32 P]dCTP (3,000 Ci/mmol; New England Nuclear). After prehybridization for 4 h, RNA blots were hybridized overnight at 42°C for MC-CPA and 37°C for actin. RNA blots were

washed under conditions of high stringency (65°C for MC-CPA) and stripped at 80°C for 30 min in fresh hybridization solution (11) before being reprobed to demonstrate actin mRNA. Autoradiographic films were exposed with two enhancing screens at -70°C for 4 d for MC-CPA and 8 h for actin.

Dot matrix analysis. Caltech software employing a standard algorithm (29) was used with ratios of matches/window size of 4/5 and 8/10 for each pair of sequences compared.

Results

Isolation and structural characterization of the human MC-CPA gene. A clone containing the 5' end of the MC-CPA gene, including the first two exons and 350 bp of 5'-flanking DNA, was previously isolated from a human genomic DNA library (13). This library was rescreened with two partial-length MC-CPA cDNA probes that corresponded to approximately the middle and 3' thirds of the MC-CPA transcript to obtain 11 and 2 additional recombinant phage clones, respectively. To obtain a preliminary map and determine which of the 14 clones would be sufficient to span the entire MC-CPA gene, blots containing restriction-digested λ phage DNA from each of the clones were probed with radiolabeled oligonucleotides which were predicted to correspond to putative exons 1, 2, 3, 4, 6, 9, 10, or 11 of the MC-CPA gene, based on alignment of the MC-CPA cDNA sequence with the rat CPB gene. Four clones, designated λ HMCCPA-1 to -4, were predicted to contain inserts that spanned putative exons 1-3, 3-9, 4-10, and 10-11, respectively, of the MC-CPA gene. From detailed restriction mapping of these four clones a map of the complete human MC-CPA gene was constructed (Fig. 1). The restriction sites present in the overlapping recombinant phage inserts were consistent with a previous genomic DNA blot analysis using a full-length MC-CPA cDNA probe (13).

With the exception of gaps in the four largest introns, the nucleotide sequence of the complete human MC-CPA gene (Fig. 2 A) was determined including 1,800 bp on the 5' flank (Fig. 2 B). By comparison with the consensus nucleotide sequence for the complete MC-CPA mRNA obtained from cDNA and direct mRNA sequencing (13), the human MC-CPA gene was determined to have 11 exons spanning 32 kb. No differences were noted between the consensus MC-CPA cDNA sequence and the combined sequences of the exons. The TATAAA-like sequence CATAAA is present 31 bp 5' to the major transcription initiation site (13). Exons 1-11 are 79, 68, 76, 125, 103, 102, 111, 91, 203, 85, and 582 bp in length, respectively, and the 5' and 3' untranslated regions of exons 1 and 11 have 11 and 397 bp, respectively. The 10 introns of the MC-CPA gene vary in length from 97 bp (intron 1) to 9.9 kb (intron 10). The splice donor and acceptor sites conform to the

GT-AG rule (30) and are homologous to described splice donor and acceptor consensus sequences (31, 32).

Localization of the MC-CPA gene to human chromosome 3. PCR amplification of a 129-bp region (Fig. 2 B, *underlined*) of the MC-CPA gene 5' flanking DNA was performed using as template DNA isolated from normal human and hamster cells, or from each of a panel of 18 human/hamster somatic cell hybrid cell lines. Only the normal human DNA, and DNA from 3 of the 18 somatic cell hybrids yielded the expected ~ 129-bp fragment. The three positive hybrids contained the following sets of human chromosomes: 3 only; 3, 5, 6, 10, 19, and 21; and 3, 5, 12, 14, 20, 22, and Y. Only human chromosome 3 was present in all three of the positive hybrid cell lines, and chromosome 3 was absent in all 15 negative hybrid cell lines. Only the Y chromosome was also absent in all 15 negative hybrids, but it was present in only one of the 3 positive hybrids. The ~ 129-bp PCR products derived using as template the normal human DNA, the cloned human MC-CPA gene insert in phage clone λ HMCCPA-1 (0.1 ng phage DNA), and the positive human/hamster hybrid cell lines had identical sizes as determined by agarose gel electrophoresis (not shown). These data indicated that the MC-CPA gene is located on human chromosome 3.

Comparison of MC-CPA to other members of the carboxypeptidase gene family. The data for the intron/exon organization of the human MC-CPA gene are summarized in Fig. 3 along with schematic structures for the previously characterized rat pancreatic carboxypeptidase genes CPA1 (14), CPA2 (15), and CPB (14). With the exception of intron 3 (see below) the MC-CPA and pancreatic carboxypeptidase genes have similar overall organizations, although the lengths of corresponding introns and total lengths of the genes vary considerably. To obtain a detailed comparison of the intron/exon organization of the MC-CPA gene with the pancreatic carboxypeptidase genes, amino acid sequences for the five known prepropeptides (Fig. 4 A) and for the mature enzymes of 10 members of the carboxypeptidase gene family (Fig. 4 B) were aligned. This alignment was used to compare the relative positions of the introns for the MC-CPA and pancreatic carboxypeptidase genes. Intron numbers are designated above the corresponding amino acid residues in Fig. 4, A and B. For the MC-CPA and rat CPA1, CPA2, and CPB genes, all intron/exon junctions occur at homologous amino acid codons, except for intron 3. Furthermore, in almost every case, corresponding introns for the four genes fall after the same nucleotides of the codons. Intron 3 of the human MC-CPA gene occurs at a site homologous to that of intron 3 of the rat CPB gene, after the second nucleotide of the Glu (E) codon at amino acid position -20

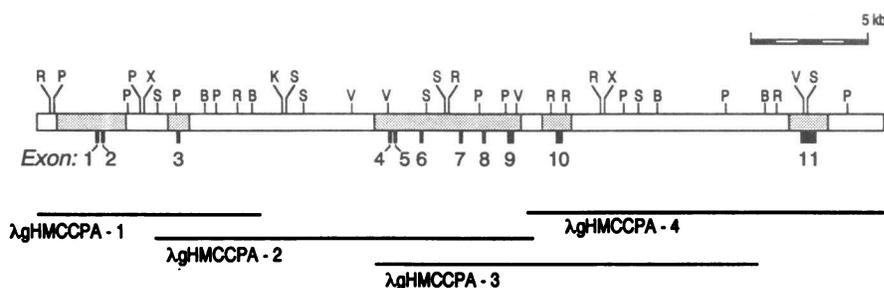


Figure 1. Restriction endonuclease map and schematic structure of the human MC-CPA gene. Exons 1-11 are depicted as numbered black boxes below the line. The five stippled areas were completely sequenced (see Fig. 2, A and B). The relative positions of the four overlapping genomic DNA clones, which together spanned the gene, are indicated (*bottom*). The positions of restriction sites are shown above the line (B, BamHI; R, EcoRI; V, EcoRV; S, SacI; P, PstI; K, KpnI). There was no Sall restriction site.

B

-1800 gaagtcctctgttgacagcacatgaacatcttcaactgtctagagagcatggagctgttttttagcacctaccacgaagtcctcaacatctctgtctactaagcctcacagccactt
-1680 gtttagagtcacattagcattccattttcaaaatgaggaaacaaagccagcaataaactaggcagctgctcaaggggatgtagctagcgtgtgggagccacattgaagtgtatctgtc
-1560 gaaccctaaagtcaaggtctcttcttctcactgttccctggtaaaagtattcttttagtcaaaagcaatggagcaagcattgtctttgggagttctcagtgctgggtgagcttggccaggtt
-1440 caagaacatgaaggcatttgaaaacctctctcttcttaccctgaagctggagccgaaatccatttggggaaataaaaaaagtcctcagcttaccacacacctacagatagag
-1320 aagaatggaattactgctgctgtatataaacaggggcgtgtctccaacacactattttgatagatcaaaagcgtgactctgagaagttttatctacagtcagctatcacgtttgtcaat
-1200 gtgaagaggtgttttcaaaaaggaaattcctaataagtttaactgattagataaaatcactactcttttaggttaccatgatcttaggtaaaggtgtattctccataatcaaccag
-1080 ttgtgtcaagactaaacatccctccagagatggttgtgaatggctctatagggtgctttcattaattcaactctcaaaaggcattatttggcaattcacagcaccaggtctgaagacag
-960 tgtccaatataaacattacagcattgatattttctgtcgtcaggtagttctcaagttgcttagaaatagggtgacacctgaggcagagagagaagggcaagagtttagggc
-840 agtgcttaggtcttttagtgttatttttaccagagaagtagcaggaataatacaaaaatgttatcttgaataatcaaatgggtaacattcaatgttcgcatagacttac
-720 agcatttttagtacatttatattttgtaatttaaaatttttgggttttttttgcctccaagtagctgggattaaggaccctgcccccagccggcgaatttttttatttttagtaga
-600 gacaggtttcaccacgtggccaggtggtcttgaactcctgacctcagatgatccaccgcgcggcctcccaactgctgggattacagcgtgagccaccgtgccagctcaattta
-480 aatgttttaattcacaacttaccattctagcttcaaaaatataattttacattatgacaccactccaagattttatggccagatacccaaatcctcagctccaagtgaat
-360 tttaaaactaattttctccatcattcaattcaagcagaatggtagagtagtattttagaacagatagaactagtaaatcccagctctgtgaacatactttctctgttacctggccaagt
-240 tatttgcataaggcatttctccatcctgtatctgtaaaatgggtatttctctattgcaaaaggtagataagtagcatgtcttgggtgggttagtgttttccctctccagctggaaaa
-120 aaaaacctgggtgtgggggaagaaggggaacagtttgtgcaacagaacctaaagaggaat(cagctgct)atcagataagggtgagccataaaactgccagaggggtctcaagcaggc

Figure 2. (A) Nucleotide sequence of the human MC-CPA gene. Nucleotide sequences of the introns and 3'-flanking DNA are shown in lowercase letters. The sequences of the 11 exons are shown in capital letters. The first nucleotide shown is the transcription initiation site (*). The deduced amino acid sequence of human MC-CPA appears above the exon sequences, with position 1 set to the amino terminus of the mature protein (4). Introns 2, 3, 9, and 10 were not completely sequenced and the approximate lengths in basepairs of the gaps are indicated. The stop codon in exon 11 is indicated (END). Two overlapping polyadenylation sites which begin 13 nucleotides from the 3' end of exon 11 are overlined. The gene-specific 131-bp cDNA probe used in RNA blot analyses corresponded to regions of exons 6 and 7 (underlined). (B) Nucleotide sequence of the MC-CPA gene 5'-flanking DNA. Numbers on the left indicate positions relative to the transcription initiation site. Sequence motifs with homology to the 5'-flanking DNA of the human cathepsin G gene (circled) and the human skin mast cell tryptase gene (boxed) are indicated. A GATA-binding transcription factor recognition site AGATAA (bracket) is present immediately 5' to the TATAAA-like sequence CATAAA (bold underline), which begins 31 bp 5' to the transcription initiation site. The 129-bp region that was amplified by PCR in chromosomal localization experiments is underlined. The sequence data for the 5'- and 3'-flanking DNA and MC-CPA gene are available under EMBL/GenBank/ DDBJ accession numbers M73716, M73717, M73718, M73719, and M73720.

(Fig. 4 A). In contrast, codons for an additional seven amino acids (IMIEDVQ) precede intron 3 in the CPA2 gene (15). These same seven amino acids are found in the large exon 3 of the CPA1 gene, but no equivalent third intron is present (14).

For the 10 aligned mature carboxypeptidase amino acid sequences (Fig. 4 B), 34 amino acid residues were completely invariant. Based on x-ray crystallographic studies (33, 34) of bovine CPA (35) and CPB (36), several of the invariant residues correspond to predicted zinc-binding, and catalytic and substrate-binding amino acids.

RNA blot analysis using a gene-specific probe for MC-CPA. Similar-sized 1.7-kb MC-CPA transcripts were detected in poly(A)⁺ RNA from lung (Fig. 5, lane 1) and skin (lane 2)

tissue. MC-CPA mRNA expression was more prominent in a total RNA sample from dispersed lung cells enriched to contain 10% mast cells (lane 18), than in poly(A)⁺ RNA from whole lung tissue. Poly(A)⁺ RNA from brain and liver tissue did not contain detectable MC-CPA mRNA (data not shown). Several transformed tumor cell lines were examined. MC-CPA transcripts were present in two myelomonocytic tumor cell lines, KG-1 (lane 5) and HL-60 (lane 15), and at a low level in the T cell lymphoma line Jurkat (lane 8). MC-CPA mRNA was undetectable in two other T lymphocytic tumor cell lines, CEM (lane 9) and H9 (lane 10), the monocytic tumor cell lines THP-1 (lane 3) and U-937 (lane 4), the Burkitt lymphoma line Raji (lane 6), an EBV-LCL (lane 7), and the epithelial tumor

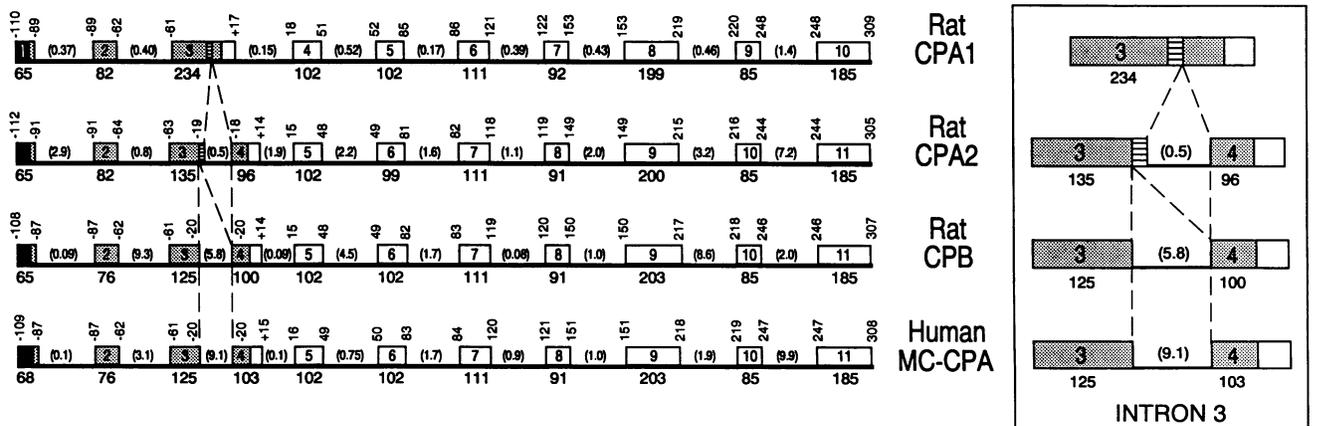


Figure 3. Gene organization of human MC-CPA compared with rat pancreatic carboxypeptidases CPA1, CPA2, and CPB. Exons are indicated as numbered boxes, with the number of protein-coding nucleotides in each exon indicated below. Exons encoding the hydrophobic leader peptides (solid), activation peptides (stippled), and mature enzyme sequences (open boxes) are indicated. Vertically arranged numbers indicate the amino acids encoded at the intron/exon junctions, with position 1 set to the first amino acid of the mature enzymes. Intron lengths (in kilobases) are in parentheses. Dashed lines between the genes show the relative positions of homologous amino acids, based on alignment of the amino acid sequences (see Fig. 4 A). The small, horizontally barred area (inset) represents seven amino acids, IMIEDVQ, present in exon 3 of CPA1 and CPA2 (see Fig. 4 A). The gene structures for CPA1, CPA2, and CPB are taken from Clauser et al. (14) and Gardell et al. (15).

A

```

HUMAN MC-CPA  MRLILPVLGIATTLAIAPVRFDRKVFVRVQPEKQA-DIIK-DLAKTNELDFWYPGATHVHAANMMVDFRVSEKESQAIQSALDQNKMHYE-----ILIHDLQEEIEKQFDVKED  -1
MOUSE MC-CPA  ::FF:LMAV:Y:::HS:::L:N:H:-SVL:-N:TQSI:::D:I:DI:V:T:::T:::T:E:H:I:::-----:DE
RAT CPB       :L:L:ALVSV:LAH:SEE-H::GNR:Y::SVHG:DHV-NL:Q-E::N:K:I:::K:DSATQ:KPLTT:::H:KAEDVADVENF:EE:EV:::-----V::SNVRNAL:S:::SHTR
RAT CPA1     :KRL:ILS:LLEAVCGNE-N:VGHQ:L:ISAA::A:VQVKLELDLEHLQ:::R----DAAR:GIPI:V::PFPSI:SVKAF:EYHGIS::IMIEDVQLLL:EEKQMSAFQAR
RAT CPA2     :::T:LLAALLGYIYQCE-T:VG DQ:LEII:SH:E:IRTLQLLEAEHL:::K----SPTIPGETVHV::PFASI::VKVF:ESQID:SIMIEDVQLL:QER:EMLFNQQR:R
  
```

B

```

HUMAN MC-CPA  IPGRHSYAKYNNWEKIVAWTEKMMKYPPEMVSRIKIGSTVEDNPLYVLKIGEKN  54
MOUSE MC-CPA  :A:::D:D:::S:::LE:H:::K:D
RAT MC-CPA    :A:::D:N:::S:::VE:H:::R:D
COW CPB      -TTG::E::T::E:::QVASEN:DLI:SA::T:FLG:TI:L:V:KPG
RAT CPB      -ASG::T::K::T:E::IQVATDN:DL:TQSV::T:F:GRNM:::KTR
RAT CPA1     ALSTDSFN::T:HTLDE:YEFMDLLVAEH:QL:K:Q::N:F:GR:IH:::FSTGG
RAT CPA2     -G:NFNFEA:HTL:E:YQEMDNLVAEH:GL:KVN:L::SF:NR:MN:::FSTGG
COW CPA      ARSTNTFN::T:HTLDE:YDFMDLLVAEH:QL:KLQ::RSY:GR:I:::FSTGG
CRAYFISH CPB ---MDWTS:HDYDE:N::LDSLATD::LA:VEDV:LSY:GRMKL:L:KGG
BACTERIA CP  -DFPPADSR:H:YAEMN:AIDARTAAN:SIM:KRV::K:YQGRVIAV:VSDNV

HUMAN MC-CPA  ---ERRKAI FMDCGIHAREVSPAFQWFVYQATKTYGRNKIMTKLLDRMNFYILP  107
MOUSE MC-CPA  ---GE:::I:::S:::K:::V::
RAT MC-CPA    ---GE:::A:S:::K:N:::V::
COW CPB      ---SNKP:V:::F:::I:::RE:VR:::EIH:EF:KLD:V::
RAT CPB      ---PNKP:I:::F:::I:::RE:VR:::NOEIH:KQ:ELD:V::
RAT CPA1     ---TN:P:WI:T::S:::TQ:SGV:AKKV:D::QOPTF:AV:N:DIFLEI
RAT CPA2     ---DKP:WL:A:::TQ:TAL:TANKIASD:TOPAI:S::NTLDIFL:
COW CPA      ---SN:P:WI:L::S:::ITQ:TGV:AKKF:EN:Q:PSF:AI::S:DIFLEI
CRAYFISH CPB ---ADKPI:I:G:::IA:STVYI:NEFVSN---SATYDDI:SNV::VM:
BACTERIA CP  AAD:AEPEVLFTAHQ:::HLTVEMALYLLRELQGG::SDSRI:QAVNGRELW:V:

HUMAN MC-CPA  VFNVDG--YIWSWTKNRMWRKNRSKNQNSK-CIGTDLNRRNFNASWNS-IPNTNDPC  159
MOUSE MC-CPA  :D:::QD:::R:::T:::DV:D::S:::K::
RAT MC-CPA    :D:::D:::PS:T:::DV:D::S:::DN::
COW CPB      :V:I:::YT::T:::TRAG:S::T:::D:G:C:::GAS:N::
RAT CPB      :V:I:::VYT::D:::T:TMAG:S::L:VRP:::D:G:C:::GAS:N::
RAT CPA1     :T:P:::FAYTHKT:::T:HT:G:L::V:V:P:::WD:GFGM-AGASSN::
RAT CPA2     :T:P:::VF:Q:T:::T:RSG:G::V:V:P:::WD:NFGM-PGASSS::
COW CPA      :T:PN:::FAFTHSE::L:::VTSS:L::V:V:A:::WD:GFGM-AGASSS::
CRAYFISH CPB TI:P:::AYTF:DD:L:::T:ETG:SVLG:K:A:P:::WSFH:DE-VGASSS::
BACTERIA CP  DM:P:::GE:DIASGSY:S:::QP:AG:S-AV:::WAYK:GCCGSSSS:S:

HUMAN MC-CPA  ADNYRGSAPESKETKAVTNFIRSHLNE----IKVYITFHYSQMLLFPYGYTSKL  211
MOUSE MC-CPA  LNV::P:::S:::A:::I:::F::
RAT MC-CPA    LSV::P:::S:::A:::I:::F::
COW CPB      SET:C:::A:::S:::AD::N::SS---:A:L:I:::M:Y::S:DY::
RAT CPB      SET:C:P:::LAD::NN:ST---:A:L:I:::M:Y::S:DY::
RAT CPA1     SET::KF:N::V:V:SIVD:VT:G:::AF:SI:::L:Y:::EP
RAT CPA2     S:S:H:PK:N::V:V:SIVD::K:GK---:V:AF:L:::L:M:::KCTK
COW CPA      SET:H:KYAN::V:V:SIVD:VKN:G:::F:AFLSI:::L:Y:::TQS
CRAYFISH CPB S:I:A:PE:F::V:MRN:RDQ:LEYAAN---:L:::LWMY:W:F::D:
BACTERIA CP  SET::A:A:AP:::V:AD:V::RVVGGKQQ:TAA:D::T::ELV:W:F::YND

HUMAN MC-CPA  PPNHEDLAKVAKIGTDVLSYTRYETRYIYGPIESTIYPIGSSSLDWAYDLG-IKHTF  266
MOUSE MC-CPA  :Q::L::R:A::A:::A:::KT:::V::
RAT MC-CPA    :Q::L::R:A::S:::A:::KT:::V::
COW CPB      :K:NVE:NTL::GAVKK:ASLHG:T:S::GAT:::A::G:D:::Q:::YS:
RAT CPB      :E:Y:E:NALV:GAAKE:A:LHG:K:T::GAT:::AA:G:D:::S:::YS:
RAT CPA1     A:DQAE:DQL::SAVTA:TSLHG:KFK::S:ID::QA:::T::T:SQ:::YS:
RAT CPA2     :DDFNE:DE::QKAAQ:KRLHG:S:KV:::C:V::QA::G:I:::S:::YS:
COW CPA      I:DKTE:NQ:::SAVAA:KSL:G:S:K::S:IT::QA::G:I::S:NQ:::YS:
CRAYFISH CPB :DDMQ::DTL:TNAV:A:TAVHG::EI:SSTN::AAA:G:D:::KGE:GV:YAY
BACTERIA CP  TAPGMTADDRNFAAVGQKMAASNG:TAEQSSDL::ITD::ID::LWGSQ-KIFGY

HUMAN MC-CPA  AFEL--RDKGKFGFLLPESRIKPTCRETMLAVKFTAKYILKHTS  308
MOUSE MC-CPA  :S:::K:::S:::N::
RAT MC-CPA    :S:::K:::S:::N::
COW CPB      T:::RY::V::Q:Q::E:::I:YVTS:V:E:L-
RAT CPB      T:::T:F:::Q:RQ::E:::Y::N:VRE:LY
RAT CPA1     T:::T:LR:::A:Q:I::AE:W::LLT:MDHTV:PY
RAT CPA2     :TAFY:::AKQ:L::AE:W:GL:T:MEHTVRD:PY
COW CPA      T:::T:RY:::A:Q:I::AQ:W:G:LT:MEHTVNN-
CRAYFISH CPB TI:::T:NY:::NQ:I::GE:::FEG::VV:NFKDITY
BACTERIA CP  T::MYP:SASGG::YP:DEV:ERETSRRND::LQLIENADCMYRSIGKEAQYCS
  
```

Figure 4. (A) Amino acid sequence alignment for five known carboxypeptidase leader and activation peptides, and location of introns 1–3 for the MC-CPA, CPB, CPA1, and CPA2 genes. Dashes represent insertion of a gap and colons represent identity with human MC-CPA. Introns 1–3 for the human MC-CPA and rat CPB, CPA1, and CPA2 genes occur at the amino acid codons indicated (numbers above the amino acid sequences). The rat CPA1 gene has no similarly placed third intron. The complete structure of the mouse MC-CPA gene has not been determined. (B) Mature enzyme amino acid sequence alignment for 10 phylogenetically diverse carboxypeptidases, and location of introns 4–10 for the MC-CPA, CPB, CPA1, and CPA2 genes. Predicted catalytic and substrate-binding (□) and zinc-binding amino acids (●) are strongly conserved. The critical specificity-determining residue corresponding to Ile²⁵⁵ of bovine CPA is indicated (▽). Numbers to the right indicate amino acid position numbers for human MC-CPA and cow CPA, with position 1 set to the start of the mature enzymes. For introns 4–10 of the MC-CPA gene (numbers above the amino acid sequence), the corresponding introns of the rat CPB, CPA1, and CPA2 genes occur at the homologous amino acid codon in every case.

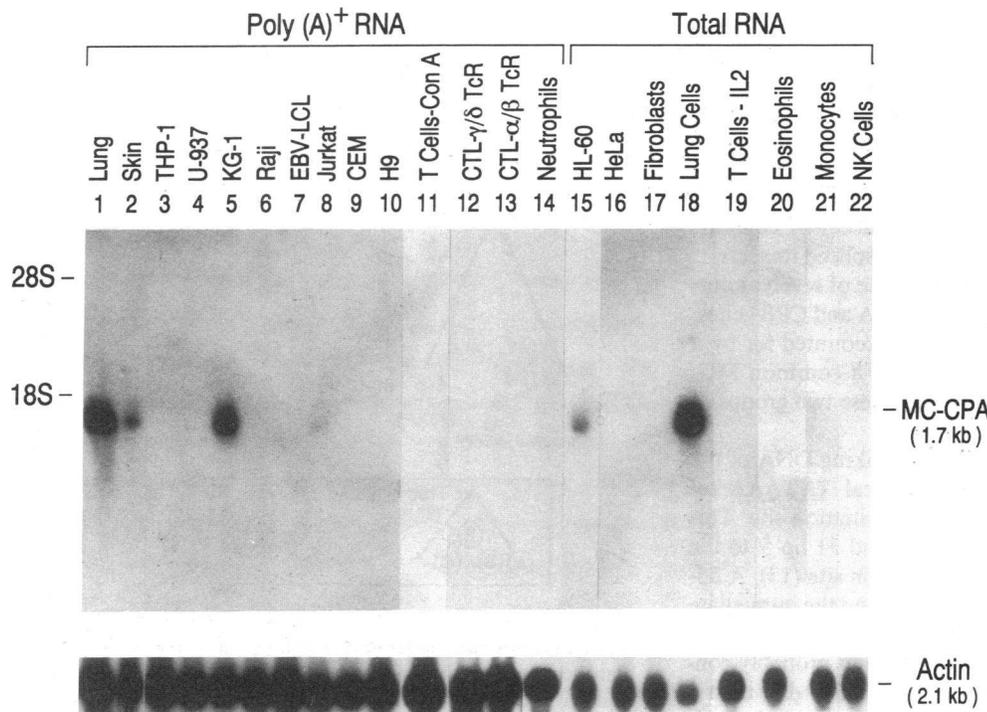


Figure 5. RNA blot analysis with a gene-specific probe for MC-CPA. Poly(A)⁺ RNA or total RNA from human tissues, cell lines, and purified cell populations was probed with a 131-bp, gene-specific cDNA probe for MC-CPA or with a β -actin cDNA probe. Film was exposed for 4 d for MC-CPA and for 8 h for actin. The positions of the 18S and 28S ribosomal RNAs and the sizes of the hybridizing bands are shown.

cell line HeLa (lane 16). An RNA blot of 15 μ g of total RNA/lane from the megakaryocytic cell line CMK, cultured with and without 10 ng/ml phorbol myristate acetate (PMA) to induce greater phenotypic differentiation (28), was also examined. MC-CPA mRNA was detected in both PMA-treated and untreated CMK cells, similar in intensity to that seen in HL-60 cells (data not shown).

MC-CPA mRNA expression was examined in purified, nontransformed cell populations of several lineages, including other mature, bone marrow-derived populations that express secretory granules. None of these populations contained detectable MC-CPA mRNA, including unfractionated PBMC (poly(A)⁺ RNA; data not shown), purified adherent monocytes (lane 21), neutrophils ($\sim 1 \mu$ g poly(A)⁺ RNA; lane 14), hypodense eosinophils (lane 20), CD56⁺CD6⁻ polyclonal NK cells (lane 22), CD56⁻CD6⁺ polyclonal T cells activated for 6 d in IL-2 (from the same culture as the NK cells; lane 19), polyclonal T cells activated for 7 d with Con A (lane 11), and a primary dermal fibroblast culture (lane 17). Three cloned CD8⁺CD4⁻ CTL lines also did not contain detectable MC-CPA mRNA. These included two PHA-activated, mycoplasma extract-specific, γ/δ TcR-expressing CTL lines designated DG.SF68 (lane 12) and DG.SF37 (15 μ g total RNA/lane; data not shown), and one HLA-B35-specific, α/β TcR-expressing CTL line designated 8S1 (lane 13).

Discussion

In this study we describe the cloning and characterization of the human MC-CPA gene and demonstrate its mast cell-specific transcription among several normal bone marrow-derived cell populations. The relationship of MC-CPA to other members of the carboxypeptidase gene family was established by analysis of the MC-CPA gene structure and evolutionary tree analysis of the carboxypeptidase gene family.

Using MC-CPA cDNA probes the complete 32-kb MC-CPA gene was isolated, restriction mapped (Fig. 1), subcloned into plasmids, and partially sequenced (Fig. 2, A and B). The 11 exons of the MC-CPA gene contained a short, 11-nucleotide 5'-untranslated region followed by the coding sequences for the leader peptide (within exon 1), activation peptide (exons 1–4), and mature enzyme (exons 4–11). Based on x-ray crystallographic studies (33, 34) of bovine CPA (35) and CPB (36), predicted zinc- and substrate-binding amino acid residues of MC-CPA are distributed over several exons. The predicted zinc-binding residues corresponding to His⁶⁹, Glu⁷², and His¹⁹⁶ of bovine CPA are in exons 6, 6, and 9, respectively, whereas catalytic and substrate-binding residues corresponding to Arg⁷¹, Arg¹²⁷, Asn¹⁴⁴, Arg¹⁴⁵, Tyr¹⁹⁸, Asp²⁵⁶, Glu²⁷⁰, and Phe²⁷⁹ of bovine CPA are in exons 6, 8, 8, 8, 9, 11, 11, and 11, respectively.

The intron/exon organization of the MC-CPA gene (Fig. 3) was compared with the pancreatic carboxypeptidase genes by determining the relative positions of intron breakpoints in the aligned preproenzyme amino acid sequences (Fig. 4, A and B). With the exception of intron 3, placement of introns was highly conserved, indicating that this gene family derived from a single primordial carboxypeptidase gene. The splice junction for intron 10 of the MC-CPA gene occurs within the codon for the predicted substrate-binding residue Tyr²⁴⁷ (Tyr²⁴⁸ of bovine CPA). Splice sites often are located at codons for surface residues, where a junctional mutation may not significantly alter protein structure (37). Tyr²⁴⁸ of bovine CPA may have arisen by a junctional mutation early in carboxypeptidase evolution and then been conserved, as it is an invariant residue in the carboxypeptidase gene family (Fig. 4 B) and occurs at the splice junction for intron 10 in all four carboxypeptidase genes (14, 15).

A distinguishing structural feature of the carboxypeptidase genes is the presence and location of the third intron. Exon 3 of

the CPA1 gene encodes the last 61 amino acids of the activation peptide and the first 17 amino acids of the mature enzyme (14). The corresponding region in the CPA2 (15), CPB (14), and MC-CPA genes is divided by intron 3 (see Fig. 3, inset). Intron 3 in the CPA1 gene must have been deleted after a gene duplication leading to separate CPA1 and CPA2 genes. Examples have been described of precise intron loss in members of other gene families (38, 39). Intron loss may occur by homologous recombination with a DNA copy of the spliced transcript, as in pseudogene formation (40). The absence of seven amino acid codons in the third exons of the MC-CPA and CPB genes, as opposed to CPA1 and CPA2, could be accounted for by a junctional sliding event having occurred in a common MC-CPA/CPB or CPA1/CPA2 ancestor, after these two groups of genes diverged through gene duplication.

A preliminary 167-bp sequence of 5' flanking DNA of the mouse MC-CPA gene (41) shows a classical TATAAA sequence 32 bp 5' to the gene transcription initiation site. This corresponds to the CATAAA sequence found 31 bp 5' to the human MC-CPA gene transcription initiation site (13). A 55-bp region including the CATAAA sequence and the immediate 5'-flanking DNA is strongly conserved across species from human to mouse (44 of 55 matches) (13, 41) and probably contains MC-CPA-specific regulatory elements. By dot matrix analysis, a 9-bp sequence was identified within this 55-bp conserved region [CAGCTGCTC] at positions -59 to -51 which is also present at positions -207 to -199 in the 5'-flanking DNA of the human cathepsin G gene (321 bp) (42). Cathepsin G is expressed in the same populations of human mast cells that contain MC-CPA (43). Also found within this 55-bp conserved region of the MC-CPA gene 5' flank is a recognition site for GATA-binding transcription factors [AGATAA] (44) at positions -46 to -41. Dot matrix comparison of the 5'-flanking regions of the human MC-CPA gene (1,800 bp) with the 5'-flanking DNA sequences of the genes for rat CPA1 (307 bp), CPA2 (422 bp), and CPB (607 bp), a mouse mast cell chymase termed MMCP-4 (1,853 bp) (12), and a mouse mast cell tryptase termed MMCP-6 (1,583 bp) (45), showed no significant homology. A 14-bp sequence from -378 to -365 of the human MC-CPA gene [CAGTCTCCAAGTGC] was identified which has 13 of 13 matches with positions -202 to -190 [CAGTCTCC-AGTGC] in the 5' flank of a human skin mast cell tryptase gene (224 bp) (46), with insertion of a 1-bp gap in the tryptase sequence. These 13 bp from the human tryptase gene include the first 11 bp of a 20-bp element identified (46) as being homologous to the consensus core enhancer of pancreatic serine protease genes (47). A similar sequence was noted in the rat chymase RMCP-II gene promoter region (48).

X-ray crystallographic studies (33, 34) of bovine pancreatic carboxypeptidases have shown that the amino acid residue at position 255 of CPA and CPB is crucial in determining substrate specificity, being Ile in CPA and Asp in CPB. Rat pancreatic CPA1 and CPA2 each have an Ile, and rat CPB has an Asp at this site. Since the crayfish hepatopancreas contains homologues of trypsin and CPB but lacks both chymotrypsin and CPA, Titani et al. suggested that CPB may actually predate CPA in the evolution of this gene family (49). In 1979 Narahashi and Yoda described a bacterial metalloproteinase from *Streptomyces griseus* that combined the specificities of mammalian CPA and CPB (50) and had a substrate-binding pocket with an Ile adjacent to an Asp residue, corresponding to positions 254 and 255 (51), respectively, of bovine CPA. It was

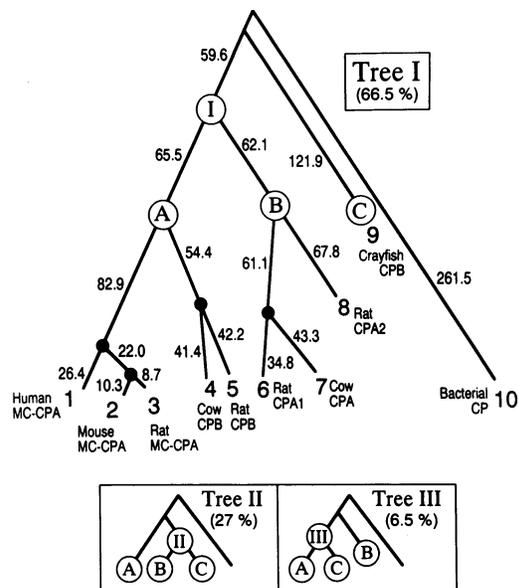


Figure 6. Minimal nucleotide replacement tree for the 10 carboxypeptidase amino acid sequences shown in Fig. 4. Numbers on the branches represent the least number of nucleotide substitutions that could give rise to the amino acid sequences at the tips. There are four groups of orthologues ([1-3], [4, 5], [6, 7], and [8]) that are paralogous to each other, which arose from three gene duplications at nodes A, B, and I. For the different enzymes the critical substrate specificity-determining residues corresponding to Ile²⁵⁵ of bovine CPA are Leu (branches 1-3), Asp (branches 4, 5, and 9), Ile (branches 6-8), and Asp with an adjacent Ile (branch 10). Statistical ambiguity for the placement of the branch leading to crayfish CPB was found, with the most likely arrangement shown in tree I. The value 66.5% represents the statistical support for node I from 100 bootstraps. All other nodes had 100% support from those bootstraps and were unambiguously defined based on the set of 10 amino acid sequences analyzed. The remaining 33.5% of statistical support was distributed to trees II (27%) and III (6.5%), which differ from tree I only in the manner of divergence of the crayfish CPB lineage. The odds are therefore ~ 2.5:1 in favor of the divergence of crayfish CPB taking place before the gene duplication giving rise to group A (tree I vs. tree II). Likewise, the odds are ~ 10:1 in favor of tree I vs. tree III, in which the gene duplication giving rise to group B is the first event. The root of the tree was not defined, but was drawn as the midpoint of the longest lineage. Closed circles mark divergences attributable to mammalian speciation. The bootstrapping statistical analysis yielded noninteger values for nucleotide replacements which were rounded off to one decimal place.

speculated (51) that the primordial carboxypeptidase may have been like the *S. griseus* enzyme, with CPA and CPB of higher organisms having lost one or the other specificity as they evolved to fulfill more specific functions. Each of the MC-CPAs has a Leu at the position corresponding to Ile²⁵⁵ of bovine CPA (11, 13, 52). The unique substrate-binding pockets of the MC-CPAs confer CPA-like enzymatic specificity (1-3) despite overall protein (11, 13, 52) and gene intron/exon structures most similar to pancreatic CPB. Further, the 5'-flanking DNA of the MC-CPA gene is distinct from that of either CPA or CPB, implying that distinct regulatory mechanisms are associated with its mast cell-specific distribution. Taken together, these findings demonstrate that the MC-CPA gene is a novel member of the carboxypeptidase gene family.

To establish the molecular ancestry of MC-CPA within the carboxypeptidase gene family, an evolutionary tree was constructed for the 10 mature carboxypeptidase amino acid sequences, using the Fitch algorithm for generation of minimal nucleotide replacement trees from amino acid sequence data (53). The most parsimonious nucleotide replacement tree that could give rise to the different amino acid sequences is shown in Fig. 6. The number of nucleotide substitutions required to give the amino acid sequences are shown, and are depicted as the lengths of the branches. There are four groups of orthologues ([1-3], [4, 5], [6, 7], and [8]), representing four modern mammalian carboxypeptidase genes and their species homologues. These groups are paralogous to each other, arising from three gene duplications at A, B, and I. In tree I, the crayfish CPB lineage diverges before separate A and B groups are formed by gene duplication. A common MC-CPA/CPB ancestor diverged from the lineage leading to CPA by a gene duplication, and then underwent a second duplication to form separate but similar MC-CPA and CPB gene structures as defined by intron/exon organization. Distinct ancestors of these three modern mammalian carboxypeptidase genes, CPA, CPB, and MC-CPA, were established before the radiation of mammalian species which began ~ 100 million years ago (40). Because the mouse, rat, and human MC-CPAs all have a Leu residue at the critical specificity-determining position corresponding to Ile²⁵⁵ of bovine CPA, conversion of this residue to a Leu in the lineage leading to the MC-CPAs probably occurred before mammalian speciation as well.

Using a gene-specific MC-CPA cDNA probe under conditions of high stringency, similar-sized MC-CPA transcripts are present in poly(A)⁺ RNA from lung and skin tissues, and in total RNA from dispersed lung cells enriched to contain 10% mast cells (Fig. 5). In transformed cell lines MC-CPA mRNA was barely detectable in poly(A)⁺ RNA from one of three T lymphoblastic tumor cell lines, and was also detected in three granulated tumor lines of bone marrow origin: the megakaryocytic cell line CMK and two immature myelomonocytic cell lines, HL-60 and KG-1 (54, 55). None of several purified, non-transformed cell populations contained detectable MC-CPA mRNA, including neutrophils, eosinophils, monocytes, NK cells, unfractionated PBMC, polyclonal T lymphocytes, and cloned CTL lines. Therefore MC-CPA mRNA is a specific molecular marker for mast cells among normal hematopoietic cell populations.

Acknowledgments

We thank Dr. Walter Fitch, University of California at Irvine, for performing the statistical evolutionary tree analysis using his algorithms; Dr. Temple Smith, MBCRR, Dana-Farber Cancer Institute, for helpful discussions; and Dr. William Serafin, Vanderbilt Medical School, for providing the skin RNA.

This work was supported by grants HL-36110, AI-31599, AR-36308, AR-35907, AI-22531, and RR-05950 from the National Institutes of Health, and in part by a grant from the Hyde and Watson Foundation.

References

1. Everitt, M. T., and H. Neurath. 1980. Rat peritoneal mast cell carboxypeptidase: localization, purification, and enzymatic properties. *FEBS (Fed. Eur. Biochem. Soc.) Lett.* 110:292-296.
2. Serafin, W. E., E. T. Dayton, P. M. Gravalles, K. F. Austen, and R. L. Stevens. 1987. Carboxypeptidase A in mouse mast cells: identification, characterization, and use as a differentiation marker. *J. Immunol.* 139:3771-3776.

3. Goldstein, S. M., C. E. Kaempfer, D. Proud, L. B. Schwartz, A.-M. Irani, and B. U. Wintroub. 1987. Detection and partial characterization of a human mast cell carboxypeptidase. *J. Immunol.* 139:2724-2729.
4. Goldstein, S. M., C. E. Kaempfer, J. T. Kealey, and B. U. Wintroub. 1989. Human mast cell carboxypeptidase. Purification and characterization. *J. Clin. Invest.* 83:1630-1636.
5. Kokkonen, J. O., M. Vartiainen, and P. T. Kovanen. 1986. Low density lipoprotein degradation by secretory granules of rat mast cells: sequential degradation of apolipoprotein B by granule chymase and carboxypeptidase A. *J. Biol. Chem.* 261:16067-16072.
6. Kokkonen, J. O., and P. T. Kovanen. 1989. Proteolytic enzymes of mast cell granules degrade low density lipoproteins and promote their granule-mediated uptake by macrophages in vitro. *J. Biol. Chem.* 264:10749-10755.
7. Weidner, N., and K. F. Austen. 1991. Ultrastructural and immunohistochemical characterization of normal mast cells at multiple body sites. *J. Invest. Dermatol.* 96:26S-30S.
8. Irani, A. A., N. M. Schechter, S. S. Craig, G. DeBlois, and L. B. Schwartz. 1986. Two types of human mast cells that have distinct neutral protease compositions. *Proc. Natl. Acad. Sci. USA.* 83:4464-4468.
9. Schwartz, L. B., A.-M. 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.
10. Reynolds, D. S., W. E. Serafin, D. V. Faller, D. A. Wall, A. K. Abbas, A. M. Dvorak, K. F. Austen, and R. L. Stevens. 1988. Immortalization of murine connective tissue-type mast cells at multiple stages of their differentiation by coculture of splenocytes with fibroblasts that produce Kirsten sarcoma virus. *J. Biol. Chem.* 263:12783-12791.
11. Reynolds, D. S., R. L. Stevens, D. S. Gurley, W. S. Lane, K. F. Austen, and W. E. Serafin. 1989. Isolation and molecular cloning of mast cell carboxypeptidase A: a novel member of the carboxypeptidase gene family. *J. Biol. Chem.* 264:20094-20099.
12. Serafin, W. E., T. P. Sullivan, G. A. Conder, A. Ebrahimi, P. Marcham, S. S. Johnson, K. F. Austen, and D. S. Reynolds. 1991. Cloning of the cDNA and gene for mouse mast cell protease-4: demonstration of its late transcription in mast cell subclasses and analysis of homology to subclass-specific neutral proteases of the mouse and rat. *J. Biol. Chem.* 266:1934-1941.
13. Reynolds, D. S., D. S. Gurley, R. L. Stevens, D. J. Sugarbaker, K. F. Austen, and W. E. Serafin. 1989. Cloning of human mast cell carboxypeptidase A, and comparison with mouse mast cell carboxypeptidase A and rat pancreatic carboxypeptidases. *Proc. Natl. Acad. Sci. USA.* 86:9480-9484.
14. Clauser, E., S. J. Gardell, C. S. Craik, R. J. MacDonald, and W. J. Rutter. 1988. Structural characterization of the rat carboxypeptidase A1 and B genes. *J. Biol. Chem.* 263:17837-17845.
15. Gardell, S. J., C. S. Craik, E. Clauser, E. J. Goldsmith, C. B. Stewart, M. Graf, and W. J. Rutter. 1988. A novel rat carboxypeptidase, CPA2: characterization, molecular cloning, and evolutionary implications on substrate specificity in the carboxypeptidase gene family. *J. Biol. Chem.* 263:17828-17836.
16. Tsonis, P. A., and T. Manes. 1988. Rapid phage DNA isolation without the use of enzymes. *Biotechniques.* 6:950-951.
17. Wong, W. W., L. B. Klickstein, J. A. Smith, J. H. Weis, and D. T. Fearon. 1985. Identification of a partial cDNA clone for the human receptor for complement fragments C3b/C4b. *Proc. Natl. Acad. Sci. USA.* 82:7711-7715.
18. Pearson, W. R., and D. J. Lipman. 1986. Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA.* 85:2444-2448.
19. Feng, D. F., and R. F. Doolittle. 1987. Progressive sequence alignment as a prerequisite to correct phylogenetic trees. *J. Mol. Evol.* 25:351-360.
20. Caulfield, J. P., R. A. Lewis, A. Hein, and K. F. Austen. 1980. Secretion in dissociated human pulmonary mast cells. Evidence for solubilization of granule contents before discharge. *J. Cell Biol.* 85:299-311.
21. Chomczynski, P., and N. Sacchi. 1987. Single-step method of RNA extraction by acid-guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* 162:156-159.
22. Boyum, A. 1968. Isolation of mononuclear cells and granulocytes from human blood: isolation of mononuclear cells by one centrifugation, and of granulocytes by combining centrifugation with sedimentation at 1 g. *Scand. J. Clin. Lab. Invest.* 21(Suppl 97):77-89.
23. Czop, J. K., and J. Kay. 1991. Isolation and characterization of β -glucan receptors on human mononuclear phagocytes. *J. Exp. Med.* 173:1511-1520.
24. Lam, B. K., W. F. Owen, Jr., K. F. Austen, and R. J. Soberman. 1989. The identification of a distinct export step following the biosynthesis of leukotriene C₄ by human eosinophils. *J. Biol. Chem.* 264:12885-12889.
25. Anderson, P., M. Caligiuri, J. Ritz, and S. F. Schlossman. 1989. CD3-negative natural killer cells express zeta TCR as part of a novel molecular complex. *Nature (Lond.)* 341:159-162.
26. Aviv, H., and P. Leder. 1972. Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose. *Proc. Natl. Acad. Sci. USA.* 69:1408-1412.
27. Thomas, P. S. 1980. Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. *Proc. Natl. Acad. Sci. USA.* 77:5201-5205.

28. Komatsu, N., T. Suda, M. Moroi, N. Tokuyama, Y. Sakata, M. Okada, T. Nishida, Y. Hirai, T. Sato, A. Fuse, and Y. Miura. 1989. Growth and differentiation of a human megakaryoblastic cell line, CMK. *Blood*. 74:42-48.
29. Maizel, J. V., Jr., and R. P. Lenk. 1981. Enhanced graphic matrix analysis of nucleic acid and protein sequences. *Proc. Natl. Acad. Sci. USA*. 78:7665-7669.
30. Breathnach, R., and P. Chambon. 1981. Organization and expression of eukaryotic split genes coding for proteins. *Annu. Rev. Biochem.* 50:349-383.
31. Mount, S. M. 1982. A catalogue of splice junction sequences. *Nucleic Acids Res.* 10:459-472.
32. Shapiro, M. B., and P. Senapathy. 1987. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res.* 15:7155-7177.
33. Rees, D. C., M. Lewis, and W. N. Lipscomb. 1983. Refined crystal structure of carboxypeptidase A at 1.54-angstrom resolution. *J. Mol. Biol.* 168:367-387.
34. Schmid, M. F., and J. R. Herriott. 1976. Structure of carboxypeptidase B at 2.8-angstrom resolution. *J. Mol. Biol.* 103:175-190.
35. Bradshaw, R. A., K. A. Walsh, and H. Neurath. 1971. Amino acid sequence of bovine carboxypeptidase A. Isolation and characterization of selected peptic and nagarse peptides and the complete sequence of fragment F-I. *Biochemistry*. 10:961-972.
36. Titani, K., L. H. Ericsson, K. A. Walsh, and H. Neurath. 1975. Amino acid sequence of bovine carboxypeptidase B. *Proc. Natl. Acad. Sci. USA*. 72:1666-1670.
37. Craik, C. S., S. Sprang, R. Fletterick, and W. J. Rutter. 1982. Intron-exon splice junctions map at protein surfaces. *Nature (Lond.)*. 299:180-182.
38. Perler, F., A. Efstratiadis, P. Lomedico, W. Gilbert, R. Kolodner, and J. Dodgson. 1980. The evolution of genes: the chicken preproinsulin gene. *Cell*. 20:555-566.
39. Chang, C., and E. M. Meyerowitz. 1986. Molecular cloning and DNA sequence of the *Arabidopsis thaliana* alcohol dehydrogenase gene. *Proc. Natl. Acad. Sci. USA*. 83:1408-1412.
40. Gilbert, W., M. Marchionni, and G. McKnight. 1986. On the antiquity of introns. *Cell*. 46:151-154.
41. Reynolds, D. S. 1989. Development of immortalized mouse mast cell lines and molecular characterization of mast cell secretory granule proteases. Ph.D. thesis. Harvard University, Cambridge, MA. 162.
42. Hohn, P. A., N. C. Popeson, R. D. Hanson, G. Salvesen, and T. J. Ley. 1989. Genomic organization and chromosomal localization of the human cathepsin G gene. *J. Biol. Chem.* 264:13412-13419.
43. Schechter, N. M., A. M. Irani, J. L. Sprows, J. Abernathy, B. Wintroub, and L. B. Schwartz. 1990. Identification of a cathepsin G-like proteinase in the MCTC type of human mast cell. *J. Immunol.* 145:2652-2661.
44. Zon, L. I., M. S. Gurish, R. L. Stevens, C. Mather, D. S. Reynolds, K. F. Austen, and S. H. Orkin. 1991. GATA-binding transcription factors in mast cells regulate the promoter of the mast cell carboxypeptidase gene. *J. Biol. Chem.* In press.
45. Reynolds, D. S., D. S. Gurley, K. F. Austen, and W. E. Serafin. 1991. Cloning of the cDNA and gene of mouse mast cell protease-6: transcription by progenitor mast cells and mast cells of the connective tissue subclass. *J. Biol. Chem.* 266:3847-3853.
46. Vanderslice, P., S. M. Ballinger, E. K. Tam, S. M. Goldstein, C. S. Craik, and G. H. Caughey. 1990. Human mast cell tryptase: multiple cDNAs and genes reveal a multigene serine protease family. *Proc. Natl. Acad. Sci. USA*. 87:3811-3815.
47. Boulet, A. M., C. R. Erwin, and W. J. Rutter. 1986. Cell-specific enhancers in the rat exocrine pancreas. *Proc. Natl. Acad. Sci. USA*. 83:3599-3603.
48. Sarid, J., P. N. Benfey, and P. Leder. 1989. The mast cell-specific expression of a protease gene, RMCP II, is regulated by an enhancer element that binds specifically to mast cell trans-acting factors. *J. Biol. Chem.* 264:1022-1026.
49. Titani, K., L. H. Ericsson, S. Kumar, F. Jakob, H. Neurath, and R. Zwilling. 1984. Amino acid sequence of crayfish (*Astacus fluviatilis*) carboxypeptidase B. *Biochemistry*. 23:1245-1250.
50. Narahashi, Y., and K. Yoda. 1979. Purification and some properties of a new metallo carboxypeptidase of *Streptomyces griseus* K-1. *J. Biochem. (Tokyo)*. 86:683-694.
51. Narahashi, Y. 1990. The amino acid sequence of zinc-carboxypeptidase from *Streptomyces griseus*. *J. Biochem. (Tokyo)*. 107:879-886.
52. Cole, K. R., S. Kumar, H. L. Trong, R. G. Woodbury, K. A. Walsh, and H. Neurath. 1991. Rat mast cell carboxypeptidase: amino acid sequence and evidence of enzyme activity within mast cell granules. *Biochemistry*. 30:648-655.
53. Fitch, W. M., and J. S. Farris. 1974. Evolutionary trees with minimum nucleotide replacements from amino acid sequences. *J. Mol. Evol.* 3:263-278.
54. Leglise, M. C., G. A. Dent, L. H. Ayscue, and D. W. Ross. 1988. Leukemic cell maturation: phenotypic variability and oncogene expression in HL60 cells: a review. *Blood Cells (NY)*. 13:319-337.
55. Furlley, A. J., B. R. Reeves, S. Mizutani, L. J. Altass, S. M. Watt, M. C. Jacob, P. van den Elsen, C. Terhorst, and M. F. Greaves. 1986. Divergent molecular phenotypes of KG1 and KG1a myeloid cell lines. *Blood*. 68:1101-1107.