Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 9480–9484, December 1989
Immunology

Cloning of cDNAs that encode human mast cell carboxypeptidase A, and comparison of the protein with mouse mast cell carboxypeptidase A and rat pancreatic carboxypeptidases

(secretory granule/protease)

DALE S. REYNOLDS*†‡, DANIEL S. GURLEY*†, RICHARD L. STEVENS*†, DAVID J. SUGARBAKER§¶, K. FRANK AUSTEN*†, AND WILLIAM E. SERAFIN*†

Departments of *Medicine and §Surgery, Harvard Medical School, and the †Department of Rheumatology and Immunology and the ¶Division of Thoracic Surgery, Brigham and Women's Hospital, Boston, MA 02115

Contributed by K. Frank Austen, August 24, 1989

ABSTRACT **Human skin and lung mast cells and rodent peritoneal mast cells contain a carboxypeptidase in their secretory granules. We have screened human lung cDNA libraries with a mouse mast cell carboxypeptidase A (MC-CPA) cDNA probe to isolate a near-full-length cDNA that encodes human MC-CPA. The 5′ end of the human MC-CPA transcript was defined by direct mRNA sequencing and by isolation and partial sequencing of the human MC-CPA gene. Human MC-CPA is predicted to be translated as a 417 amino acid preproenzyme which includes a 15 amino acid signal peptide and a 94 amino acid activation peptide. The mature human MC-CPA enzyme has a predicted size of 36.1 kDa, a net positive charge of 16 at neutral pH, and 86% amino acid sequence identity with mouse MC-CPA. DNA blot analyses showed that human MC-CPA mRNA is transcribed from a single locus in the human genome. Comparison of the human MC-CPA with mouse MC-CPA and with three rat pancreatic carboxypeptidases shows that these enzymes are encoded by distinct but homologous genes.**

Carboxypeptidase activity has been detected in the secretory granules of rat (1, 2) and mouse (3) peritoneal cavity mast cells and in human skin and lung mast cells (4, 5). The pH optima of the rodent and human mast cell carboxypeptidases are neutral to basic, which distinguishes these secretory granule exopeptidases from lysosomal carboxypeptidases that are present in various cells, including fibroblasts (6). The rodent mast cell-derived enzyme has been termed mast cell carboxypeptidase A (MC-CPA) because it preferentially cleaves substrates with carboxyl-terminal aliphatic or aromatic amino acids (1). Recently, we determined the amino-terminal amino acid sequence of mouse MC-CPA. An oligonucleotide probe was prepared and used to isolate cDNAs that encode mouse MC-CPA from a Kirsten sarcoma virus-immortalized mast cell cDNA library (7, 8). The deduced amino acid sequence from these cDNAs revealed that the mouse exopeptidase is translated as a 417 amino acid preproenzyme that possesses a 15 amino acid signal peptide and a 94 amino acid activation peptide (7).

Because the amino-terminal amino acid sequences of human (5) and mouse (7) MC-CPA are nearly identical, a mouse MC-CPA cDNA was used as a probe to screen a human lung cDNA library. We have now isolated and sequenced cDNAs,‖ as well as a portion of the human gene, that encode a human enzyme which has 86% amino acid sequence identity with the mature mouse MC-CPA and which has a predicted amino-terminal amino acid sequence identical to that reported for the amino terminus of the human skin MC-CPA (5).

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

MATERIALS AND METHODS

Isolation of cDNAs That Encode Human MC-CPA. A 592-base-pair (bp) cDNA which corresponds to residues 45 to 637 of the 1.5-kilobase (kb) mouse MC-CPA mRNA transcript (7) was radiolabeled (9) and used to screen an amplified human lung cDNA library (library HL1004b; Clontech). Filters (Magnagraph, Micron Separations, Westboro, MA) were probed at 37°C in 50% (vol/vol) formamide/750 mM NaCl/75 mM sodium citrate/2× Denhardt's buffer/0.1% $NaDodSO_4$/1 mM EDTA/100 μg of salmon sperm DNA per ml/10 mM sodium phosphate, pH 7.0 (1× Denhardt's buffer is 0.02% bovine serum albumin/0.02% Ficoll/0.02% polyvinylpyrrolidone). The filters were washed under conditions of low stringency (37°C; 30 mM NaCl/3 mM sodium citrate/0.1% $NaDodSO_4$/1 mM EDTA/10 mM sodium phosphate, pH 7.0) to isolate the human lung cDNA designated cDNA HCPA-I. A 585-bp 5′→*Eco*RV fragment of cDNA HCPA-I was then used to screen ≈2 × 10^6 recombinants from the amplified human lung cDNA library HL1066b, which has a reported complexity of 1.4 × 10^6 recombinants (Clontech). Filters were washed at high stringency (55°C) to isolate 21 additional human lung MC-CPA cDNAs. These cloned phage were then probed with a 131-bp 5′→*Sac* I restriction fragment of the mouse MC-CPA cDNA under conditions of low stringency to identify clones that contained the 5′ region of the human MC-CPA cDNA sequence. Phage λ DNA was obtained by a rapid miniprep procedure (10), and cDNA inserts were ligated into M13mp18 and sequenced (11). The antisense strand of cDNA HCPA-III and both strands of cDNA HCPA-I and cDNA HCPA-II were sequenced in their entirety. The cDNA HCPA-III was found to have attached on its 5′ end an ≈900-bp region encoding an irrelevant cDNA. cDNA HCPA-IV was only partially sequenced to determine that it did not extend further 5′ than cDNA HCPA-II.

Primer Extension Analysis. The 5′ end of the relatively rare MC-CPA transcript in whole human lung tissue RNA was defined by primer extension analysis and RNA sequencing (12–14) as modified below. Grossly normal human lung tissue (30 g) was dissected from the margins of a surgical specimen and was homogenized in a blender in guanidine isothiocyanate buffer (15). Total RNA was extracted and poly(A)$^+$ RNA was

Abbreviations: CPA1, CPA2, and CPB, rat pancreatic carboxypeptidases A1, A2, and B; MC-CPA, mast cell carboxypeptidase A.

‡To whom reprint requests should be addressed at: Harvard Medical School, The Seeley G. Mudd Building, Room 624, 250 Longwood Avenue, Boston, MA 02115.

‖The sequence reported in this paper has been deposited in the GenBank data base (accession no. M27717).

selected (16). One hundred nanograms of a gel-purified antisense 26-mer oligonucleotide (5′-TCATCCTGGGGCT-TCACGCGGAACAC-3′) was labeled with T4 polynucleotide kinase (New England Biolabs) (13), and the enzyme was inactivated at 65°C for 5 min. The radiolabeled oligonucleotide (10 ng) was mixed with ≈30 μg of poly(A)$^+$ human lung RNA and annealed for 1 hr at 71°C (12) in a total volume of 40 μl. Extension/termination mixes (13.2 μl) containing 33 units of avian myeloblastosis virus reverse transcriptase (Stratagene), 0.5 mM deoxynucleotides (Ultrapure Set, Pharmacia), and either no dideoxynucleotides or 0.42 mM adenine, cytosine, guanine, or thymine dideoxynucleotide (Pharmacia) were prepared on ice (12). Aliquots (8 μl) of the annealed RNA/primer were added to the extension/termination mixes, incubated at 46°C for 1 hr, stopped, and analyzed on a 10% acrylamide/7 M urea sequencing gel as described (7). Size markers were from a DNA sequencing ladder.

Partial Characterization of the Human MC-CPA Gene. Approximately 5 × 10^5 recombinants from a human genomic DNA library in λEMBL-3 (HL1067J; Clontech) were screened (17) with a ^{32}P-labeled 26-mer oligonucleotide (5′-CAAAGCG-GACAGGAGCAATTGCAAGA-3′) that was complementary to nucleotide residues 39 to 64 of the sense strand of the consensus human MC-CPA cDNA. From five primary hybridizing plaques, one recombinant was isolated, and the 5′ portion of the human MC-CPA gene was subcloned in Bluescript plasmid (Stratagene) and sequenced on both strands.

DNA Blot Analyses. Samples (20 μg per lane) of human genomic placenta DNA (Clontech) were digested for 18 hr at 37°C with *Bam*HI, *Eco*RI, or *Hin*dIII, resolved by 0.65% agarose gel electrophoresis, and transferred to Zetabind (Cuno) (18). The blot was probed under the conditions described above with a combination of the 928-bp cDNA HCPA-II and the 1026-bp *Bst*XI→3′ fragment of cDNA HCPA-III (which together comprise a nearly full-length MC-CPA cDNA probe), or separately with a 131-bp 5′→*Bst*XI fragment of cDNA HCPA-I. The DNA blot was washed under conditions of high stringency as above, except that final washes were at 65°C.

RESULTS

Isolation and Sequencing of cDNAs That Encode Human MC-CPA. A human lung cDNA library was probed under conditions of low stringency with a mouse MC-CPA cDNA to isolate the 1193-bp human MC-CPA cDNA, which was designated cDNA HCPA-I. This cDNA was found to have high nucleotide sequence similarity to the 3′ end of the mouse MC-CPA cDNA (data not shown). Subsequent screening of a second human lung cDNA library at high stringency with a 585-bp 5′→*Eco*RV restriction fragment of cDNA HCPA-I yielded 21 hybridizing phage plaques, at a frequency of approximately 1 positive clone per 80,000–100,000 recombinant phage screened. Two of these clones (cDNAs HCPA-II and -IV) were subsequently found to hybridize to a 5′ restriction fragment of the mouse MC-CPA cDNA. These two clones, and one additional clone (cDNA HCPA-III) that did not hybridize to the 5′ mouse probe, were sequenced. The consensus nucleotide sequence for the four characterized human MC-CPA cDNA clones is shown in Fig. 1 in uppercase letters, and the sequencing strategy is shown in Fig. 2. Relative to the full-length mouse cDNA, the consensus human MC-CPA cDNA sequence lacked a 5′ untranslated region and ≈4 bp of the coding region. Two overlapping polyadenylylation sites (Fig. 1) occurred 22 and 17 nucleotides, respectively, upstream of the poly(A) tract, which was greater than 60 nucleotides in cDNA HCPA-I. cDNA HCPA-III contained the polyadenylylation sites but had 19 additional nucleotides on the 3′ end (shown in parentheses). The last four of these nucleotides in cDNA HCPA-III may represent the start of a poly(A) tail for this clone. No other differences were noted between the nucleotide sequences of the cDNA clones. By combining the consensus cDNA sequence with the sequence determined directly from the human lung mRNA transcript (see below),

```
   1 atg aGG CTC ATC CTG CCT GTG GGT TTG ATT GCT ACC ACT CTT GCA ATT GCT CCT GTC CGC TTT GAC AGG GAG AAG GTG TTC
      M   R   L   I   L   P   V   G   L   I   A   T   T   L   A ▲ I   A   P   V   R   F   D   R   E   K   V   F   -83
  82 CGC GTG AAG CCC CAG GAT GAA AAA CAA GCA GAC ATC ATA AAG GAC TTG GCC AAA ACC AAT GAG CTT GAC TTC TGG TAT CCA
      R   V   K   P   Q   D   E   K   Q   A   D   I   I   K   D   L   A   K   T   N   E   L   D   F   W   Y   P   -56
 163 GGT GCC ACC CAC CAC GTA GCT GCT AAT ATG ATG GTG GAT TTC CGA GTT AGT GAG AAG GAA TCC CAA GCC ATC CAG TCT GCC
      G   A   T   H   H   V   A   A   N   M   M   V   D   F   R   V   S   E   K   E   S   Q   A   I   Q   S   A   -29
 244 TTG GAT CAA AAT AAA ATG CAC TAT GAA ATC TTG ATT CAT GAT CTA CAA GAA GAG ATT GAG AAA CAG TTT GAT GTT AAA GAA
      L   D   Q   N   K   M   H   Y   E   I   L   I   H   D   L   Q   E   E   I   E   K   Q   F   D   V   K   E    -2
 325 GAT|ATC CCA GGC AGG CAC AGC TAC GCA AAA TAC AAT AAT TGG GAA AAG ATT GTG GCT TGG ACT GAA AAG ATG ATG GAT AAG
      D ▲|I   P   G   R   H   S   Y   A   K   Y   N   N   W   E   K   I   V   A   W   T   E   K   M   M   D   K    26
 406 TAT CCT|GAA ATG GTC TCT CGT ATT AAA ATT GGA TCT ACT GTT GAA GAT AAT CCA CTA TAT GTT CTG AAG ATT GGG GAA AAG
      Y   P |E   M   V   S   R   I   K   I   G   S   T   V   E   D   N   P   L   Y   V   L   K   I   G   E   K    53
 487 AAT GAA AGA AGA AAG GCT ATT TTT ATG GAT TGT GGC ATT(CAC)GCA[CGA](GAA)TGG GTC TCC CCA GCA TTC TGC CAG TGG TTT
      N   E   R   R   K   A   I   F   M   D   C   G   I  (H)  A  [R] (E)  W   V   S   P   A   F   C   Q   W   F    80
 568 GTC TAT CAG GCA ACC AAA ACT TAT GGG AGA AAC AAA ATT ATG ACC AAA CTC TTG GAC CGA ATG AAT TTT TAC ATT CTT CCT
      V   Y   Q   A   T   K   T   Y   G   R   N   K   I   M   T   K   L   L   D   R   M   N   F   Y   I   L   P   107
 649 GTG TTC AAT GTT GAT GGA TAT ATT TGG TCA TGG ACA AAG AAC CGC ATG TGG AGA AAA AAT CGT TCC AAG AAC CAA AAC TCC
      V   F   N   V   D   G   Y   I   W   S   W   T   K   N   R   M   W   R   K   N   R   S   K   N   Q   N   S   134
 730 AAA TGC ATC GGC ACT GAC CTC[AAC][AGG]AAT TTT AAT GCT TCA TGG AAC TCC ATT CCT AAC ACC AAT GAC CCA TGT GCA GAT
      K   C   I   G   T   D   L  [N]  [R]  N   F   N   A   S   W   N   S   I   P   N   T   N   D   P   C   A   D   161
 811 AAC TAT CGG GGC TCT GCA CCA GAG TCC GAG AAA GAG ACG AAA GCT GTC ACT AAT TTC ATT AGA AGC CAC CTG AAT GAA ATC
      N   Y   R   G   S   A   P   E   S   E   K   E   T   K   A   V   T   N   F   I   R   S   H   L   N   E   I   188
 892 AAG GTT TAC ATC ACC TTC(CAT)TCC[TAC]TCC CAG ATG CTA TTG TTT CCC TAT GGA TAT ACA TCA AAA CTG CCA CCT AAC CAT
      K   V   Y   I   T   F  (H)  S  [Y]  S   Q   M   L   L   F   P   Y   G   Y   T   S   K   L   P   P   N   H   215
 973 GAG GAC TTG GCC AAA GTT GCA AAG ATT GGC ACT GAT GTT CTA TCA ACT CGA TAT GAA ACC CGC TAC ATC TAT GGC CCA ATA
      E   D   L   A   K   V   A   K   I   G   T   D   V   L   S   T   R   Y   E   T   R   Y   I   Y   G   P   I   242
1054 GAA TCA ACA ATT TAC CCG ATA TCA GGT TCT TCT TTA[GAC]TGG GCT TAT GAC CTG GGC ATC AAA CAC ACA TTT GCC TTT GAG
      E   S   T   I   Y   P   I   S   G   S   S   L  [D]  W   A   Y   D   L   G   I   K   H   T   F   A   F   E   269
1135 CTC CGA GAT AAA GGC AAA TTT GGT[TTT]CTC CTT CCA GAA TCC CGG ATA AAG CCA ACG TGC AGA GAG ACC ATG CTA GCT GTC
      L   R   D   K   G   K   F   G  [F]  L   L   P   E   S   R   I   K   P   T   C   R   E   T   M   L   A   V   296
1216 AAA TTT ATT GCC AAG TAT ATC CTC AAG CAT ACT TCC TAA    AGAACTGCCCTCTGTTTGGAATAAGCCAATTAATCCTTTTTTGTGCCTTTCA
      K   F   I   A   K   Y   I   L   K   H   T   S  Stop                                                       308

1307 TCAGAAAGTCAATCTTCAGTTATCCCCAAATGCAGCTTCTATTTCACCTGAATCCTTCTCTTGCTCATTTAAGTCCCATGTTACTGCTGTTTGCTTTTACTTACTTT
1414 CAGTAGCACCATAACGAAGTAGCTTTAAGTGAAACCTTTTAACTACCTTTCTTTGCTCCAAGTGAAGTTTGGACCCAGCAGAAAGCATTATTTTGAAAGGTGATATA
1521 CAGTGGGGCACAGAAAACAAATGAAAACCCTCAGTTTCTCACAGATTTTCACCATGTGGCTTCATCAATTTATGTGCTAATACAATAAAATAAAATGCACTTAAA(C
1627 GAAATTTTAAGTAGAAAA)AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
```

FIG. 1. Consensus nucleotide and amino acid sequences of human MC-CPA. The arrows indicate the sites where the hydrophobic signal peptide and the activation peptide are predicted to be cleaved from the preproenzyme. The amino-terminal portion of the mature protein that has been determined by amino acid sequencing (5) is boxed. Two overlapping polyadenylylation sites are underlined. Potential zinc-binding (circled) and substrate-binding (small boxes) amino acids are also indicated. The numbers to the left are the number of bp from the start of the consensus sequence; to the right are the number of amino acids, with no. 1 assigned to the first amino acid of the mature enzyme. The first four nucleotides (lowercase letters) were determined from sequencing the MC-CPA mRNA from human lung poly(A)$^+$ RNA. The rest of the sequence (uppercase letters) represents the consensus nucleotide sequence of the MC-CPA cDNAs.

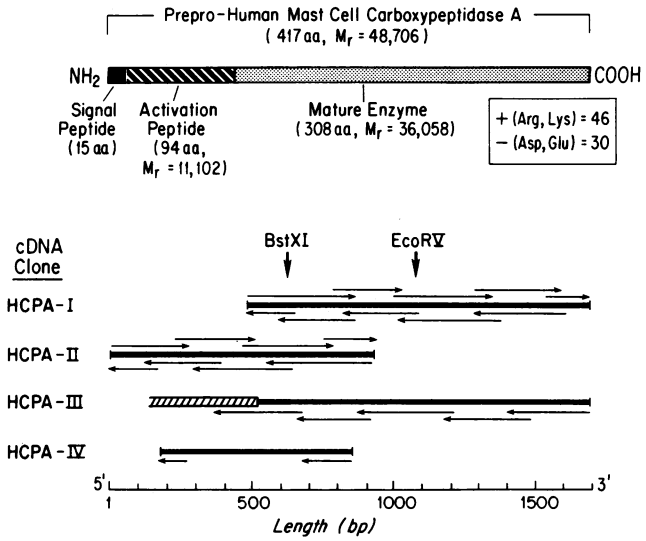

FIG. 2. Schematic of cDNA sequencing strategy and functional domains of human MC-CPA. aa, Amino acids.

the sequence for the entire coding region of the MC-CPA transcript could be determined (Fig. 1). MC-CPA is predicted to be initially translated as a 417 amino acid protein with a 48.7-kDa protein core.

Primer Extension Analysis and Sequencing of the Human MC-CPA mRNA. To characterize the 5′ end of the MC-CPA transcript, primer extension analysis and direct sequencing of the MC-CPA mRNA were performed. An antisense 26-mer oligonucleotide complementary to the human MC-CPA cDNA sequence was used with poly(A)$^+$ RNA from human lung tissue as the template. The oligonucleotide was complementary to the sequence at 72–97 nucleotides from the 5′ end of cDNA HCPA-II. As shown in Fig. 3 *Left*, 86 nucleotides were extended onto the primer, resulting in an extension product of 112 nucleotides. By this analysis, cDNA HCPA-II lacked 4 translated and 11 untranslated nucleotides at its 5′ terminus. Primer extension RNA sequencing (data not shown) confirmed the first 60 nucleotides of the consensus MC-CPA cDNA sequence (Fig. 1) and provided the 4 additional nucleotides needed to complete the translated portion of the MC-CPA sequence (Fig. 1, lowercase letters). The 5′ untranslated portion was sequenced to within 1 nucleotide of the transcription-initiation site. A minor primer extension product of 131 nucleotides was noted in two independent experiments, indicating that transcription may also initiate from a site located 19 nucleotides farther 5′ in the MC-CPA gene for a small fraction of the MC-CPA mRNA transcripts present in human lung (Fig. 3 *Left*).

Partial Characterization of the Human MC-CPA Gene. The 5′ sequence of the human MC-CPA mRNA was confirmed by sequencing the 5′ portion of the human MC-CPA gene (Fig. 3 *Right*). The region characterized included the first two exons and a portion of the 5′ flanking DNA. The nucleotide sequences of the first and second exons were identical to the consensus sequence for MC-CPA which had been obtained from both cDNA and RNA sequencing. The TATAAA-box-like sequence CATAAA was found to be located 31 bp upstream of the major transcription-initiation site (Fig. 3 *Right*). A CATCAA sequence was found to be located 32 bp upstream of the putative minor transcription-initiation site.

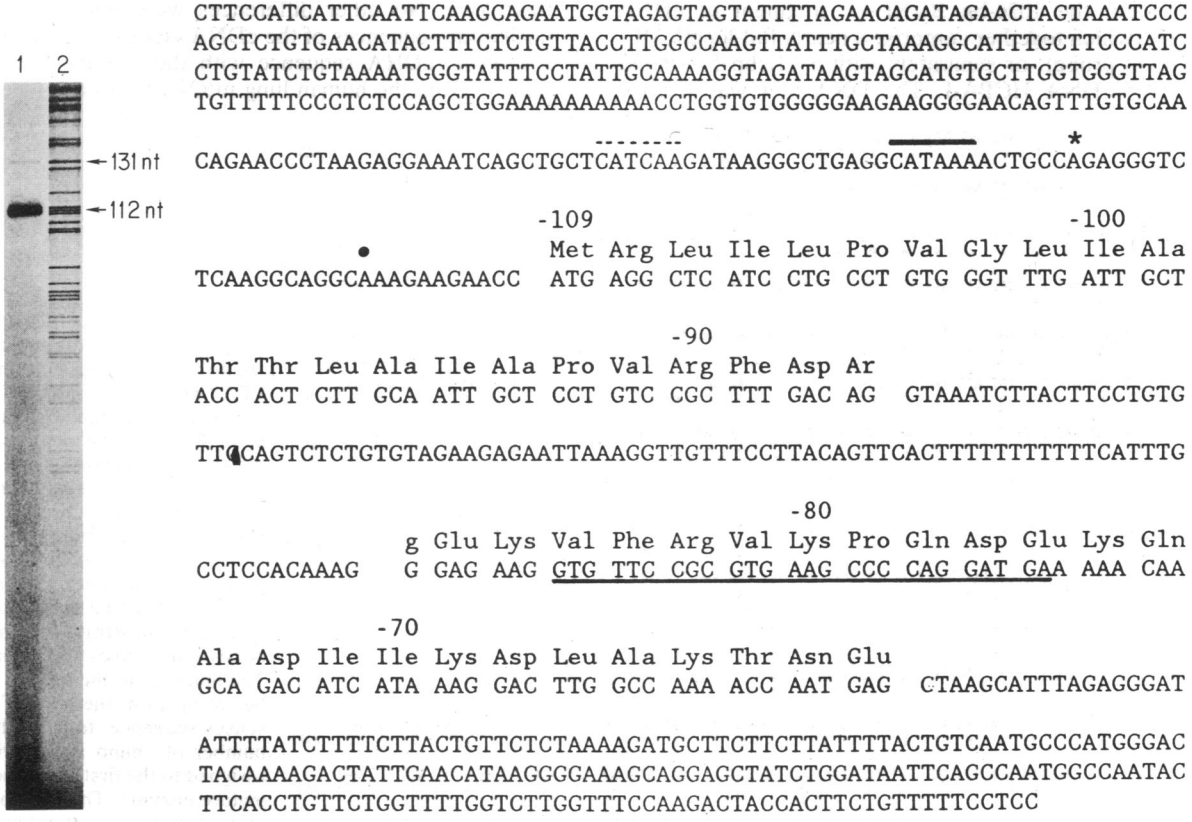

FIG. 3. (*Left*) Primer extension analysis of the 5′ end of the MC-CPA mRNA from human lung. A primer extension reaction was performed with human lung poly(A)$^+$ RNA (lane 1). Size markers (lane 2) denote the number of nucleotides in the extended products. nt, Nucleotides. (*Right*) Partial nucleotide and deduced amino acid sequences of a human MC-CPA genomic clone. The primary transcription-initiation site (●), based on the major primer extension product, is preceded by a CATAAA sequence 31 bp upstream in the flanking DNA (overlined). A putative minor transcription-initiation site (*) is also indicated and is preceded by a CATCAA sequence 32 bp upstream (broken overline). The oligonucleotide used for primer extension was complementary to the underlined sequence.

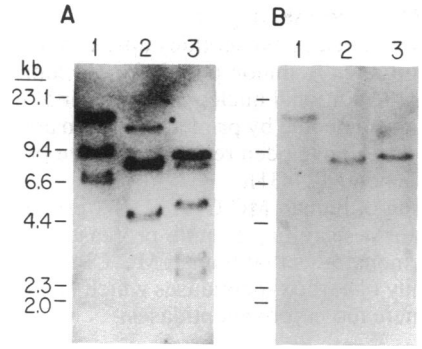

FIG. 4. Human genomic DNA blot analysis. Human placenta DNA was digested with *Bam*HI (lanes 1), *Eco*RI (lanes 2), or *Hin*dIII (lanes 3) and probed with two MC-CPA partial-length cDNA fragments which together span the MC-CPA cDNA sequence (*A*) or with a 131-bp 5′→*Bst*XI fragment of cDNA HCPA-I (*B*).

DNA Blot Analyses. Human genomic placenta DNA was probed with two MC-CPA cDNA fragments (the 928-bp cDNA HCPA-II and the 1026 bp *Bst*XI→3′ fragment of cDNA HCPA-III) which together span nearly the full length of the cDNA sequence. Multiple DNA fragments were detected in each lane, regardless of which restriction enzyme was used (Fig. 4*A*). When the same blot was then probed with the 131-bp 5′→*Bst*XI fragment of cDNA HCPA-I, hybridization to a single DNA fragment was seen in each lane, indicating that this gene is encoded at a single locus in the human genome.

DISCUSSION

The carboxypeptidases that have been isolated and characterized from mouse (3), rat (1), and human (4, 5) mast cells are 33- to 36-kDa metalloexopeptidases. The cationic MC-CPA is stored along with cationic serine proteases in an ionic complex with anionic proteoglycans in the mast cell's secretory granules (2, 3, 19, 20). The close approximation of the serine endopeptidase and carboxypeptidase enzymes in this complex is thought to facilitate the sequential cleavage and degradation of common protein substrates (21). In humans, MC-CPA is more prominent in skin mast cells than in lung mast cells (4, 5). Human skin mast cells also contain larger amounts of chymotryptic serine proteases than lung mast cells do (22).

In the present study, a human lung cDNA library was screened with a mouse MC-CPA cDNA probe (7) to obtain

```
HMC-CPA MRLILPVGLIATTLAIAPVRFDREKVFRVKPQDEKQA-DIIK-DLAKTNELDFW -58
MMC-CPA ::FF:LMAV:Y::::::::H::::::::::L:N::H:-SVL:-N:TQSI:::::
CPB     :L:L:ALVSV:LAH:SEE-H::GNR:Y::SVHG:DHV-NL:Q-E::N:K:I:::
CPA1    :KRL:ILS:LLEAVCGNE-N:VGHQ:L:ISAA::A:VQKVKELEDLEHLQ::::
CPA2    :::T:LLAALLGYIYCQE-T:VGDQ:LEII:SH:E:IRTLLQLEAEEHL::::

HMC-CPA YPGATHHVAANMMVDFRVSEKESQAIQSALDQNKMHYE------ILIHDLQEEI -10
MMC-CPA ::D:I:DI:V::T:::::::::::T:::T:E:H:I:::------:::::::::
CPB     K:DSATQ:KPLTT:::H:KAEDVADVENF:EE:EV:::------V::SNVRNAL
CPA1    R----DAAR:GIPI:V::PFPSI:SVKAF:EYHGIS::IMIEDVQLLLDEEKQQ
CPA2    K----SPTIPGETVHV::PFASI:AVKVF:ESQGID:SIMIEDVQVLLDQER:E

HMC-CPA EKQFDVKEDIPGRHSYAKYNNWEKIVAWTEKMMDKYPEMVSRIKIGSTVEDNPL  45
MMC-CPA ::::::::::A::::::::D:D:::S:::::LE:H::::::::::::::::::
CPB     :S:::SHTRAS:-:::T:::K::T:E::IQQVATDN:DL:TQSV::T:F:GRNM
CPA1    MSAFQARALSTDSFN::T:HTLDE:YEFMDLLVAEH:QL::K:Q::N:F:GR:I
CPA2    MLFNQQR:R-GGNFNFEA:HTL:E:YQEMDNLVAEN:GL::KVNL::SF:NR:M

HMC-CPA YVLKIGEKNERRKAIFMDCGIHAREWVSPAFCQWFVYQATKTYGRNKIMTKLLD  99
MMC-CPA ::::::K:DGE:::::::::::::::I::::::::::::::S::K:::::::::
CPB     ::::::KTRPNKP:::I:::F:::::I:::::::::RE:VR::NQEIH:KQ:::
CPA1    H:::FSTGGTN:P::WI:T:::S::::TQ:SGV::AKKI::D::QDPTF:AV::
CPA2    N:::FSTGG-DKP::WL:A::::::::TQ:TAL:TANKIASD::TDPAI:S::N

HMC-CPA RMNFYILPVFNVDGYIWSWTKNRMWRKNRSKNQNSKCIGTDLNRNFNASWNSIP 153
MMC-CPA :::::V:::::::::::::QD::::::::R::::T:::::::::DV::D:S:
CPB     ELD::V:::V:I:::VYT:::D:::::T::TMAG:S:L:VRP::::::G:CEVG
CPA1    N:DIFLEI:T:P::FAYTHKT::::::T::HT:G:L:V:V:P:::WD:GFGMAG
CPA2    TLDIFL:::T:P:::VF:Q:T::::::T:::RSG:G:V:V:P:::WD:NFGGPG

HMC-CPA NTNDPCADNYRGSAPESEKETKAVTNFIRSHLNEIKVYITFHSYSQMLLFPYGY 207
MMC-CPA :::K::LNV:::P:::::::::::::::::::S::A:::::::::::I::::
CPB     ASRS::SET:C:P:::::::::LAD:::NN:ST::A:L:I::::::M:Y::S:
CPA1    ASSN::SET:::KF:N::V:V:SIVD:VT::-GN::AF:SI:::::L::Y::::
CPA2    ASSS::S:S:H:PK:N::V:V:SIVD::K::-GKV:AF::L:::::L:M:::::

HMC-CPA TSKLPPNHEDLAKVAKIGTDVLSTRYETRYIYGPIESTIYPISGSSLDWAYDLG 261
MMC-CPA ;F::::::Q::L:::R:A::A:::::::::::::A::::KT:::::::V::::
CPB     DY:::E:Y:E:NALV:GAAKE:A:LHG:K:T:::GAT::::AA:G:D::S::Q:
CPA1    ::EPA:DQAE:DQL::SAVTA:TSLHG:KFK::S:ID:::QA:::TI::T:SQ:
CPA2    KCTK:DDFNE:DE::QKAAQA:KRLHG:S:KV:::C:V::QA::G:I:::::::

HMC-CPA IKHTFAFELRDKGKFGFLLPESRIKPTCRETMLAVKFIAKYILKHTS         308
MMC-CPA ::::::::::::::S::::::::::::K::::S::::::::::N::
CPB     ::YS:T:::::T:F::::::::Q:RQ::E:::::::Y::N:VRE:LY
CPA1    ::YS:T:::::T:LR:::::A:Q:I::AE::W::LLT:MDHTV::PY
CPA2    ::YS:::::::TAFY:::::AKQ:L::AE::W:GL:T:MEHVRD:PY
```

FIG. 5. Comparison of the deduced amino acid sequences of human MC-CPA (HMC-CPA) with mouse MC-CPA (MMC-CPA) and with the rat pancreatic carboxypeptidases B, A1, and A2 (CPB, CPA1, CPA2). Colons indicate identity with HMC-CPA. The alignment of CPB, CPA1, and CPA2 is from Clauser *et al.* (24). The numbering of the amino acids is based on the human mast cell enzyme, with position no. 1 (arrow) set to the first amino acid of the mature enzyme.

cDNAs that encode human MC-CPA (Fig. 1). The 5′ end of the MC-CPA coding sequence was then defined by primer extension analysis and RNA sequencing (Fig. 3 *Left*). The sequence of the lung MC-CPA mRNA transcript showed that the cDNA consensus sequence lacked 11 untranslated and 4 translated nucleotides at its 5′ terminus. The 5′ end of the MC-CPA sequence was then confirmed by cloning and partially sequencing the MC-CPA gene (Fig. 3 *Right*).

The predicted structure of human MC-CPA is highly similar to that of mouse MC-CPA (7). The "−3, −1 rule" of von Heijne (23) for the cleavage of signal peptides indicates that a 15 amino acid hydrophobic leader sequence precedes a 94 amino acid activation peptide in both the human and mouse enzymes. The cleavage site between the activation peptide and the amino terminus of the mature human MC-CPA was predicted by alignment with mouse MC-CPA (Fig. 5). The amino-terminal sequence of the mature human MC-CPA predicted from this alignment was identical to the amino-terminal 28 amino acid sequence reported for the human skin MC-CPA (5) (see boxed area in Fig. 1). Thus, the protein core of the mature human MC-CPA enzyme consists of 308 amino acids, has a size of 36.1 kDa, and is strongly positively charged at neutral pH (Arg + Lys = 46; Asp + Glu = 30). These features of the predicted structure are summarized in Fig. 2.

For the mature forms of the enzymes, the amino acid sequence identities of human MC-CPA compared to mouse MC-CPA (7) and to rat pancreatic CPB (24), CPA1 (25), and CPA2 (26) were 86%, 55%, 43%, and 43%, respectively. The hydrophobic signal peptide and the activation peptide of human MC-CPA were less conserved with respect to the corresponding regions of each of these enzymes. Two intrachain disulfide bonds (Cys-64/Cys-77 and Cys-136/Cys-159) are predicted by comparison with rat pancreatic CPB (24). Putative zinc-binding amino acids (His-67, Glu-70, His-195; circled in Fig. 1) and substrate-binding amino acids (Arg-69, Asn-142, Arg-143, Tyr-197, Asp-255, Phe-278; boxed in Fig. 1) in human MC-CPA correspond to identical residues in mouse MC-CPA and in rat pancreatic CPA1, CPA2, and CPB (24).

It is known that rat pancreatic CPA1 and CPA2 differ from CPB in their substrate binding sites. Both CPA1 and CPA2 have an Ile residue at position 243, whereas CPB has a Gly residue. Even though the mast cell carboxypeptidases both have higher overall structural similarity to pancreatic CPB, the corresponding residue in human MC-CPA (and in mouse MC-CPA) is an Ile at position 242, and thus is more similar to CPA1 and CPA2. The bovine pancreatic carboxypeptidases A and B (27, 28) differ at amino acid position 255. This residue has been shown to be crucial for positioning aromatic or aliphatic substrate amino acids for cleavage by bovine pancreatic carboxypeptidase A (which has an Ile) or for positioning basic amino acids for cleavage by carboxypeptidase B (which has an Asp). Because both human and mouse mast cell carboxypeptidases have the hydrophobic Leu at the corresponding position (residue 254; see Fig. 5), they would be expected to more closely resemble bovine pancreatic carboxypeptidase A than carboxypeptidase B in their substrate specificity. This agrees with the substrate preferences previously described for the mast cell carboxypeptidases (3, 5) and suggests that this human enzyme should be designated as mast cell carboxypeptidase A.

DNA blot analysis showed that human MC-CPA is encoded by a single gene (Fig. 4). This gene is likely to be relatively large, because when restriction-digested genomic DNA was probed with the near-full-length MC-CPA cDNA, hybridization to multiple large restriction fragments was detected. Primer extension analysis (Fig. 3 *Left*) and direct sequencing of the MC-CPA transcript from human lung poly(A)$^+$ RNA (not shown) showed that transcription of MC-CPA is initiated 11 nucleotides upstream of the translation-initiation site. A minor population of lung transcripts which have 19 additional nucleotides in their 5′ untranslated region was also detected by primer extension analysis. Other eukaryotic genes have been reported to have two functional promoter elements (29–31).

The cloning of human MC-CPA cDNA provides a molecular probe for a secretory granule protease of human mast cells and demonstrates that human MC-CPA is a member of a larger family of carboxypeptidases which includes the mast cell and pancreatic carboxypeptidases.

We thank Dr. C. Nicodemus (Harvard Medical School) for review of this manuscript. This work was supported by grants AI-22531, AI-23401, AI-23483, AR-35907, CA-09141, and HL-36110 from the National Institutes of Health and by the Irvington Foundation for Medical Research. R.L.S. is an Established Investigator of the American Heart Association. W.E.S. is the Burroughs Wellcome Investigator in the Immunopharmacology of Allergic Diseases.

1. Everitt, M. T. & Neurath, H. (1980) *FEBS Lett.* **110,** 292–296.
2. Schwartz, L. B., Riedel, C., Schratz, J. J. & Austen, K. F. (1982) *J. Immunol.* **128,** 1128–1133.
3. Serafin, W. E., Dayton, E. T., Gravallese, P. M., Austen, K. F. & Stevens, R. L. (1987) *J. Immunol.* **139,** 3771–3776.
4. Goldstein, S. M., Kaempfer, C. E., Proud, D., Schwartz, L. B., Irani, A.-M. & Wintroub, B. U. (1987) *J. Immunol.* **139,** 2724–2729.
5. Goldstein, S. M., Kaempfer, C. E., Kealey, J. T. & Wintroub, B. U. (1989) *J. Clin. Invest.* 83, 1630–1636.
6. Guy, C. J. & Butterworth, J. (1978) *Clin. Chim. Acta* **87,** 63–69.
7. Reynolds, D. S., Stevens, R. L., Gurley, D. S., Lane, W. S., Austen, K. F. & Serafin, W. E. (1989) *J. Biol. Chem.* **264,** in press.
8. Reynolds, D. S., Serafin, W. E., Faller, D. V., Wall, D. A., Abbas, A. K., Dvorak, A. M., Austen, K. F. & Stevens, R. L. (1988) *J. Biol. Chem.* **263,** 12783–12791.
9. Feinberg, A. P. & Vogelstein, B. (1983) *Anal. Biochem.* **132,** 6–13.
10. Tsonis, P. A. & Manes, T. (1988) *Biotechniques* **6,** 950–951.
11. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
12. Geliebter, J., Zeff, R. A., Melvold, R. W. & Nathenson, S. G. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 3371–3375.
13. Stoflet, E. S., Koeberl, D. D., Sarkar, G. & Sommer, S. S. (1988) *Science* **239,** 491–494.
14. Wise, R. J., Karn, R. C., Larsen, S. H., Hodes, M. E., Gardell, S. J. & Rutter, W. J. (1984) *Mol. Biol. Med.* **2,** 307–322.
15. Chomczynski, P. & Sacchi, N. (1987) *Anal. Biochem.* **162,** 156–159.
16. Aviv, H. & Leder, P. (1972) *Proc. Natl. Acad. Sci. USA* **69,** 1408–1412.
17. Wong, W. W., Klickstein, L. B., Smith, J. A., Weis, J. H. & Fearon, D. T. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 7711–7715.
18. Southern, E. M. (1975) *J. Mol. Biol.* **98,** 503–517.
19. Yurt, R. & Austen, K. F. (1977) *J. Exp. Med.* **146,** 1405–1419.
20. Serafin, W. E., Katz, H. R., Austen, K. F. & Stevens, R. L. (1986) *J. Biol. Chem.* **261,** 15017–15021.
21. Kokkonen, J. O., Vartiainen, M. & Kovanen, P. T. (1986) *J. Biol. Chem.* **261,** 16067–16072.
22. Schwartz, L. B., Irani, A.-M. A., Roller, K., Castells, M. C. & Schechter, N. M. (1987) *J. Immunol.* **138,** 2611–2615.
23. von Heijne, G. (1984) *J. Mol. Biol.* **173,** 243–251.
24. Clauser, E., Gardell, S. J., Craik, C. S., MacDonald, R. J. & Rutter, W. J. (1988) *J. Biol. Chem.* **263,** 17837–17845.
25. Quinto, C., Quiroga, M., Swain, W. F., Nikovits, W. C., Jr., Standring, D. N., Pictet, R. L., Valenzuela, P. & Rutter, W. J. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 31–35.
26. Gardell, S. J., Craik, C. S., Clauser, E., Goldsmith, E. J., Stewart, C.-B., Graf, M. & Rutter, W. J. (1988) *J. Biol. Chem.* **263,** 17828–17836.
27. Lipscomb, W. N., Hartsuck, J. A., Reeke, G. N., Quiocho, F. A., Bethge, P. H., Ludwig, M. L., Steitz, T. A., Muirhead, H. & Coppola, J. C. (1968) *Brookhaven Symp. Biol.* **21,** 24–90.
28. Schmid, M. F. & Herriott, J. R. (1976) *J. Mol. Biol.* **103,** 175–190.
29. Selby, M. J., Barta, A., Baxter, J. D., Bell, G. I. & Eberhardt, N. L. (1984) *J. Biol. Chem.* **259,** 13131–13138.
30. Faye, G., Leung, D. W., Tatchell, K., Hall, B. D. & Smith, M. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 2258–2262.
31. Grez, M., Land, H., Giesecke, K., Schutz, G., Jung, A. & Sippel, A. E. (1981) *Cell* **25,** 743–752.