A Novel Mouse Gene Family Coding for Cationic, Cysteine-rich Peptides

REGULATION IN SMALL INTESTINE AND CELLS OF MYELOID ORIGIN*

(Received for publication, October 23, 1989)

André J. Ouellette and John C. Lualdi

From the Cell Biology Unit, Shriners Burns Institute, Surgical Services, Massachusetts General Hospital, and the Departments of Surgery and Biological Chemistry, Harvard Medical School, Boston, Massachusetts 02114

Cryptdin is a Paneth cell corticostatin/defensin in the mouse small bowel. To help define the intestinal role of cryptdin, cryptdin-related sequence (CSD) mRNAs have been characterized with respect to developmental regulation, sequence homology, putative coding function, and occurrence in myeloid cells. Cryptdin, CRS1C, and CRS4C mRNAs are transcribed from separate genes, occur at equivalent abundance in small intestine, and appear in the small bowel in concert during the 2nd and 3rd weeks postpartum. Cryptdin and CRS1C mRNAs are not detectable in adult mouse bone marrow, but probes specific for the 5′- or the 3′-untranslated regions of CRS4C mRNA hybridize to a moderately abundant 1.05-kilobase marrow mRNA in contrast to a highly abundant 0.75-kilobase mRNA in small intestine. Nucleotide sequences corresponding to the deduced prepro-coding regions of cryptdin, CRS1C, and CRS4C mRNAs contain a highly conserved 200-base pair region of 92% sequence similarity (CSE.2), but the mRNAs are not homologous otherwise. The deduced CRS1C and CRS4C polypeptides are apparent precursors of secreted, cationic, proline- and cysteine-rich peptides that contain Cys-Pro-X repeats. Unlike cryptdin, however, the proposed CRS1C and CRS4C mature peptide regions lack the structural motif characteristic of defensins. Attempts to find homologies between the putative CRS peptides and existing protein sequences have been unsuccessful, leading us to speculate that CRS1C and CRS4C represent a new family of nondefensin antimicrobial peptides in the mouse small bowel.

The cryptdin gene codes for a precursor to a 3-kDa cationic, cysteine-rich peptide that contains an arrangement of cysteine and arginine residues characteristic of the corticostatin and defensin (CSD)† peptide family (Fig. 1; Ouellette et al., 1989a). CSDs are abundant peptides of mammalian leukocyte granules (Gabay et al., 1989; Selsted et al., 1985a, 1985b; Wilde et al., 1989) that exhibit broad spectrum antiviral and microbicidal activities in vitro (Ganz et al., 1985), antagonize ACTH responses in cultured rat adrenal cells (Singh et al., 1988; Zhu et al., 1988), and are chemotactic for monocytes (Territo et al., 1989). The presence of the defensin motif in cryptdin (Fig. 1), the conserved linkage homology between the mouse cryptdin locus and the human defensin genes (Ouellette et al., 1989b; Sparkes et al., 1989), the existence of defensin resistance genes in S. typhimurium (Fields et al., 1989; Miller et al., 1989), and the abundance of cryptdin mRNA in Paneth cells (Ouellette et al. 1989a) suggest that cryptdin functions as an antimicrobial peptide, perhaps to restrict microbial colonization of the intestinal mucosa.

Before cryptdin mRNA was found in Paneth cells of the mouse small bowel, defensins and defensin mRNAs had been isolated only from cells of myeloid origin. Human and rabbit leukocyte defensins and their corresponding mRNAs are highly abundant in spleen and bone marrow, but leukocyte defensin cDNA probes do not detect the sequences in small intestine of the same species (Daher et al., 1988; Ganz et al., 1989). Since cryptdin mRNA was cloned from the small intestine (Ouellette and Cordell, 1988; Ouellette et al., 1989a), we speculated that cryptdin and related sequences (CRS) gene family could contain members expressed exclusively in the intestine or in myeloid cells as shown for the mouse M (macrophage) and P (Paneth cell) lysozyme genes (Cross et al., 1988).

Cryptdin mRNA is one of many highly abundant mRNAs of low molecular weight that appear in postnatal intestinal development (Birkenmeier and Gordon, 1986; Ouellette and Cordell, 1988). In rodents, critical events in small bowel development and maturation occur in the suckling animal (Henning, 1985), but the mechanisms that modulate those events remain uncharacterized (Gordon, 1989). In this regard, we are studying cryptdin and cryptdin-related sequence (CRS) mRNAs as markers of Paneth cell differentiation. Here, we report that cryptdin mRNA and the related sequences CRS1C and CRS4C contain regions of extensive DNA sequence similarity, that the three genes are expressed differentially in cells of myeloid origin, and that the apparent CRS1C and CRS4C mature peptide regions lack the CSD motif, identifying the sequences as members of a new gene family.

EXPERIMENTAL PROCEDURES

Animals—Outbred Swiss mice ((C57BL/6J*(ICR)F1)BR), 45-day-old males (30-35 g) or timed pregnant dams, were obtained from Charles River Breeding Laboratories, Inc., North Wilmington, MA. In studies of neonatal and juvenile mice, litters were culled to eight pups within

---

* This work was supported by Grant 15888 from the Shriners Hospitals for Crippled Children. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† The abbreviations used are: CSD, corticostatin/defensin; CRS, cryptdin-related sequence; kb, kilobase; bp, base pair; CSE.2, 0.2-kb cryptdin-related sequence element; ACTH, adrenocorticotropic hormone.

---

9831
**Intestinal Cationic Peptide cDNAs**

**Results**

Isolation of Cryptdin-related cDNA Clones—CRS cDNA clones were detected in a mouse small intestinal cDNA library by cross-hybridization with asb/134 UNA, a 315-bp cryptdin cDNA probe (see “Experimental Procedures”). Since rabbit neutrophils contain six defensins (Solsted et al., 1985a) and adult mouse small bowel contains an estimated 20–25 abundant mRNAs that code for 6–10-kDa polypeptides (Ouellette and Cordell, 1988), isolation of several CRS clones was expected. CRS cDNAs with distinctive PstI restriction patterns were tested for hybridization to cryptdin cDNA and the probe hybridized strongly to CRS1C and CRS4C. CRS1C contains no internal PstI sites; on the other hand, the 728-bp CRS4C clone contains two internal PstI sites (see Fig. 8), but only the central 395-bp PstI fragment reacted with the cryptdin probe. The absence of hybridization to the 115- and 218-bp fragments showed that cryptdin and CRS4C contained regions both of similarity and divergence.

Extensive Sequence Similarity between Cryptdin, CRS1C, and CRS4C cDNAs—Cryptdin, CRS1C, and CRS4C mRNAs contain regions of extensive sequence similarity. The three clones contain initiating methionine codons preceded by canonical CAGGCC translational start sites (Kozak, 1984), putative polyadenylation start sites near their 3′ termini, and termination codons that occur 100–105 residues from the apparent polyadenylation start sites (Fig. 2). CRS4C contains a 250-nucleotide long 5′-untranslated region that is not found in cryptdin or in CRS1C mRNAs (Fig. 2, A–C, and Fig. 8). Open reading frames shown in Fig. 2 were the only ones consistent with the appropriate orientation of the sequences in relation to those functional markers (Fig. 2, A–C).

Dot matrix DNA sequence comparisons using the program DIAGON (Staden, 1986) showed that the 5′-terminal 200 residues in cryptdin and CRS1C cDNAs and nucleotides 265–465 of CRS4C cDNA are nearly identical (Fig. 3, A–C). The diagonals in Fig. 3, B and C, are shifted rightward because of the additional 250 bp in the 5′ end of CRS4C cDNA. At the nucleotide level, these regions of highly conserved sequence are 92% similar (Fig. 4), and they correspond to proposed preproregions of the deduced mRNA translation products (Figs. 2 and 8). We refer to this conserved sequence as the 0.2-kb cryptdin-related sequence element (CSE-2).

**Developmental Regulation of CRS mRNAs—Cryptdin, CRS1C, and CRS4C mRNAs accumulate together during postnatal intestinal development.** Experiments using full-length cDNA probes had shown that CRS1C and CRS4C mRNAs are not present in intestinal RNA from 1-day-old mice (not shown). However, since CSE-2 is a common feature of these sequences, their developmental appearance was reexamined individually using sequence-specific 3′-regional probes (see “Experimental Procedures”). Hybridization experiments with specific probes showed that cryptdin, CRS1C, and CRS4C mRNAs accumulate in the small bowel to equivalent levels and with the same kinetics of appearance (Fig. 5, 1–3). In mice younger than 20 days, only Paneth cells in the intestine hybridize to a cryptdin cDNA probe that contains

---

3 Supported by National Institutes of Health Grant NO1-GM-7-2110.
Intestinal Cationic Peptide cDNAs

FIG. 2. DNA and polypeptide sequences of cryptdin, CRS1C, and CRS4C. Both strands of cDNA clones were sequenced (see "Experimental Procedures"). The apparent coding regions of cDNA sequences are indicated by upper case notation; deduced translational initiation start sites and polyadenylation start sites are shown in bold, underlined typeface. Arrows above lines of peptide sequence denote possibly conserved Arg residues at position 60. A, cryptdin; B, CRS1C; C, CRS4C.

100 bp of CSE.2 (Ouellette et al., 1989a). In those experiments, cells containing the CRS1C and CRS4C mRNAs are likely to have been detected by the same probe, and we infer that CRS1C and CRS4C mRNAs also accumulate in Paneth cells concomitantly with cryptdin mRNA; expression of CRS1C and CRS4C genes in other intestinal cell types cannot be excluded at later times in development. The abundance of Cvs, Pro, and basic residues in CRSs (Fig. 2) and their coordinate regulation with cryptdin, an apparent defensin, suggest that CRSs may be new microbicidal peptides.

Differential Accumulation of Cryptdin and CRS mRNAs in Cells of Myeloid Origin—Because human and rabbit leukocyte defensin cDNA probes do not hybridize to small intestinal RNA from those species (Daher et al., 1988; Ganz et al., 1989), we tested for the presence of cryptdin and CRS mRNAs in myeloid cells by Northern blot analysis of RNA from adult mouse bone marrow and small intestine (Fig. 6A). Sequence-specific probes for cryptdin and CRS1C hybridized strongly to RNA from small bowel but not to RNA from bone marrow (Fig. 6A, panels 2 and 3). In contrast, CRS4C mRNA, which is as abundant in small bowel as cryptdin and CRS1C mRNAs, also occurs in bone marrow at 10–20% the intestinal concentration but at higher molecular weight (Fig. 6A, panel 1). In small intestine, the CRS4C mRNA is much more abundant than the 1.05-kb form and it frequently obscures detection of the longer sequence. Probes specific to the 5′ (Fig. 6A, panel 1) and 3′ (not shown) ends of CRS4C mRNA both hybridize to 1.05-kb femur marrow mRNA, suggesting that the same mRNA sequence is assayed by the two probes. As shown in Fig. 6B (upper panel, lane b), the quantity of 1.05-kb CRS4C mRNA in spleen is equivalent to that in bone marrow. Therefore, unlike the cryptdin and CRS1C genes, the CRS4C gene(s) is expressed in cells of both epithelial and myeloid lineage, and the mRNA is present both in a population of myeloid cells in bone marrow and probably in circulating leukocytes in spleen.

Within the gastrointestinal tract, cryptdin mRNA is specific to the jejunum and ileum in that it has not been detected in Northern blots containing 20 µg of poly(A)+ RNA from stomach, colon, cecum, liver, or submandibular gland (Ouellette and Cordell, 1988). Similarly, CRS1C mRNA is not detectable in 10 µg of total RNA from colon (not shown), stomach, liver, or kidney (Fig. 6C). On the other hand, CRS4C probes hybridize to 0.75-kb mRNA in colon (not shown) and to 1.05-kb mRNA in kidney, liver, and stomach (Fig. 6C). Within the limitations of these techniques it appears that the...
Intestinal Cationic Peptide cDNAs

Fig. 5. Developmental regulation of CRS mRNAs in mouse small intestine. Intestinal RNA (10 μg/lane) from mice 1, 9, 16, 25, and 40 days old (lanes A-E, respectively) was electrophoresed and blotted, and replicate blots were hybridized with 3′-specific probes (Fig. 5 and “Experimental Procedures”) for the following cDNAs: 1, cryptdin; 2, CRS IC; 3, CRS 4C; 4, F31 (Theodet et al., 1985). Arrows at left denote rRNA markers, and arrow at right in 4 denotes hybridization to P31 mRNA.

0.75-kb CRS 4C mRNA may be specific to the gastrointestinal tract, but this notion is speculative in the absence of in situ hybridization and DNA sequence data. Occurrence of 1.05-kb CRS 4C sequence in kidney, stomach, and liver (Fig. 6C) may reflect the presence of circulating leukocytes in those organs (Cross et al., 1988).

Cryptdin-related cDNAs Code for Nondefensin Cationic Proteins Rich in Cysteine and Proline—The deduced CRS IC and CRS 4C protein products contain 117 and 110 amino acids, respectively (Fig. 2, B and C), and they resemble cryptdin in being apparent precursors to secreted, cysteine-rich, and cat-ionic peptides. Alignment of cryptdin, CRS IC, and CRS 4C cDNAs with the program BESTFIT (Devereux et al., 1985) positions the initiating methionine codons of the three mRNAs in direct register (Fig. 4) and suggests that the proposed prepro-regions of the three deduced polypeptides may be nearly identical in length. Applying the (-3,-1) rule deduced by von Heijne (1984, 1985), the most likely signal peptide cleavage sites in the derived polypeptides are predicted by the program to produce the maximum number of possible matches.

Although CRS IC and CRS 4C have extensive DNA sequence similarity to cryptdin, a CSD, both derived peptide sequences lack the CSD motif. The first 48 N-terminal amino acids of the proposed cryptdin and CRS 4C polypeptide sequences are identical, except for a Gly to Ala substitution at position 14 (Fig. 7B). Curiously, deletion of a single 5′-terminal nucleotide in CRS IC mRNA produces a hypothetical frameshifted product that lacks an initiating methionine but is identical to residues 15–47 in putative prepro-CRS 4C and preprocryptdin.

Thus, the N-terminal regions of three gene products once may have been nearly identical. Other than being cationic and cysteine-rich, the deduced cryptdin and CRS 4C mature peptide regions show no resemblance from residues 49 through their C termini; similarly, the primary structures of the proposed cryptdin and CRS IC mature peptide regions are unrelated. Despite extensive DNA sequence similarity with cryptdin and known defensins, the deduced CRS IC and CRS 4C mature peptides both differ from cryptdin in that their Cys, Gly, and Arg residues are not arranged to form the CSD core motif (Figs. 1 and 7B). The deduced primary structures of the proposed mature CRS IC and CRS 4C peptide regions are related (Fig. 7C, boldface residues), and their higher order structures may be similar. For example, the apparent C-terminal region of CRS 4C contains five consecutive Cys-Pro/X repeats that can be aligned with four Cys-Pro/X triplets among the nine Cys-X-X repeats in the putative CRS IC peptide. Comparisons of secondary structure predictions (Chou and Fasman, 1974a, 1974b) for the presumptive CRS IC and CRS 4C mature peptide regions (Fig. 7C, boldface residues) showed that both deduced peptides contain predicted β-sheet structures over their 20 N-terminal residues, β turns followed by α helices over residues 30–45 in the central regions of the two molecules, and β sheet structures in their C-terminal regions (data not shown). These predictions differ markedly from those for the putative mature cryptdin peptide (not shown) and from
Intestinal Cationic Peptide cDNAs

FIG. 1. Alignment of cryptdin, CRS1C, and CRS4C polypeptide sequences. Deduced polypeptide sequences of cryptdin, CRS1C, and CRS4C are shown in single-letter notation as aligned using the program BESTFIT with a gap weight of 5.0 and a length weight of 0.30 (Devereux et al., 1985). Matched residues are indicated with bars between lines of sequence. Proposed mature peptide regions are indicated in boldface type. A, cryptdin versus CRS1C; B, cryptdin versus CRS4C; C, CRS1C versus CRS4C. In C, regions defined by arrow brackets denote the Cys-Pro-X and Cys-X-X triplets in CRS1C and CRS4C.

FIG. 6. Differential accumulation of cryptdin and CRS mRNAs in small bowel and bone marrow. A, RNA samples (10 μg/lane) from adult small intestine (lanes a) and bone marrow (lanes b) were electrophoresed and blotted, and replicate blots were hybridized with the following sequence-specific probes (see “Experimental Procedures”): 1, CRS4CM; 2, cryptdin (a54/134L); 3, CRS1CU. Arrows in 1 denote, in descending order, 28 S and 18 S rRNA markers, 1.05-kb CRS4C mRNA (lanes a and b), and 0.75-kb CRS4C mRNA (lane a only). B, samples (10 μg/lane) of bone marrow (lane a), spleen (lane b), small bowel (lane c), and kidney (lane d) RNA from adult mice were electrophoresed, blotted, and hybridized with CRS4CM (top) and CRS1CU (bottom) sequence-specific probes. In each panel, arrow at left denotes the position of 1.05-kb CRS4C mRNA, and arrow at right denotes the position of 0.75-kb CRS4C mRNA. C, RNAs from small intestine (lane a), kidney (lane b), liver (lane c), and stomach (lane d) of adult mice were blotted as in A and hybridized with probes specific for cryptdin (top), CRS1C (middle), and CRS4C (bottom) mRNAs.

the known structures of rabbit and human defensins (Stanfield et al., 1988; Westbrook et al., 1984). Because the primary structures of the actual mature peptides are not known definitively, and because of acknowledged uncertainties in the accuracy of secondary structure predictions by these algorithms (Fasman, 1989), particularly without known Cys-Cys bonding patterns, secondary structures should be determined on the native intestinal peptides.

Comparisons of CRS1C and CRS4C sequences with those in DNA and peptide sequence data bases were unsuccessful in identifying related peptides. Searches in the Swiss-Prot data base were conducted on GenBank On-line Service (IntelliGenetics Inc., Mountain View, CA) with the program FASTA (Pearson and Lipman, 1988) with default search parameters (ktup = 1 for proteins, ktup = 4 for DNA). Except for similarities in signal peptides from diverse species, searches with the deduced CRS1C and CRS4C full-length and proposed mature peptide sequences found no significant matches with known peptides, providing little insight into
their identities. Cys and Pro residues in the proposed CRS peptides matched some residues in Cys- and Pro-rich proteins such as gamma heavy chain disease protein (Alexander et al., 1982), Anabaena sp. ferredoxin-like protein (Mulligan et al., 1988), putative structural sperm protein (Kuhn et al., 1988), metallathioneins, keratins, and collagenogens, but none of the matches proved consistent with the chain-length or source of the CSH mRNAs. Searches in the GenBank and EMBL data bases with the full-length cryptdin, CRS1C, and CRS4C cDNAs and with CSE.2 showed similarities only with sequences coding for prepro regions of human and rabbit defensins. The failure to identify related peptide and DNA sequences indicates that cryptdin, CRS1C, and CRS4C are members of a novel gene family.

**DISCUSSION**

CRS1C and CRS4C are cryptdin-related genes that appear to code for novel, cationic, non-CSD peptides. The extensive similarity of CRS1C and CRS4C mRNAs to defensin-like cryptdin mRNA and the high Cys and Pro contents of most cationic antimicrobial peptides from human leukocytes (Barker et al., 1989; Gabay et al., 1989; Ooi et al., 1987; Selsted et al., 1985; Shafer et al., 1984; Tobias et al., 1988; Wilde et al., 1989), dipterans (Lambert et al., 1989; Matsuyama and Natori, 1988), and plants (Ohanti et al., 1977) support the general notion that members of this gene family may have microbiocidal activities.

Cryptidin and these CRS sequences are related both at the genomic and protein levels (Figs. 3, 4, 7, and 8). Since the cDNA library was constructed by homopolymeric tailing (Ohante and Cordell, 1988) cloning artifacts potentially introduced during ligation of unrelated sequences would have been circumvented. The presence of CCAGCC translational start sites (Kozak, 1984), polyadenylation start sites, and poly(A) tracts in the cDNAs confirm the orientation of the CSE.2 sequences and identified the sense strands and open reading frames of the three sequences (Fig. 2). Despite extensive homology, single nucleotide differences within the respective CSE.2 regions of cryptdin, CRS1C, and CRS4C mRNAs show that they are transcribed separately. Furthermore, the CRS1C and cryptdin (Defer, Ouellette et al., 1989b) loci are closely linked but distinct, since 1, perhaps 2, discordancies occur between strain distribution patterns of CRS1C and Defer markers in DNAs from 60 recombinant inbred mouse strains. Pending confirmation, we propose tentatively that the CRS1C locus is within 0.02-4 centimorgans of Defer and between Defer and the xenotropic proviral locus Xmu-26 (Frankel et al., 1989) in the proximal region of mouse chromosome 8.

The conserved mouse CSE.2 regions have apparent analogues in rabbit and human defensin genes. Overall, cryptdin, CRS1C, and CRS4C mRNAs show comparatively little homology to those of human (Fig. 3, D) or rabbit (not shown) defensins (see “Results”). On the other hand, the CSE.2 sequence exhibits 65 and 65% identity with corresponding regions of human and rabbit defensin cDNAs, respectively. The 5′-translated regions of rabbit and human defensin mRNAs contain a 110-bp span of 86% identity, and, perhaps more importantly, the prepro coding regions of rabbit MCP-1 and MCP-2 mRNAs, coded by the second exons of the respective genes, differ by only a single nucleotide (Ganz et al., 1989). Thus, rabbit neutrophil defensin mRNAs contain a region of almost perfectly conserved sequence that is similar in length and corresponds to the position of CSE.2 in the mouse cryptdin and CRS mRNAs. Furthermore, the CSE.2-like elements in the rabbit genes appear to be organized into conserved functional domains (Ganz et al., 1989). We speculate that members of the mouse cryptdin/CRS gene family contain similarly conserved exons of CSE.2 sequence that may have been important in their evolution.

Members of the cryptdin/CRS gene family exhibit a pattern of expression similar to that described for the mouse M and P lysozyme genes (Cross et al., 1988), which are expressed specifically in myeloid and Paneth cells, respectively. Like M lysozyme mRNA, rabbit and human neutrophil defensin mRNAs are highly abundant in bone marrow, spleen, and other myeloid cell lines, but they are not detectable in intestinal RNA (Daher et al., 1988; Ganz et al., 1989). Conversely, CRS1C mRNA and defensin-related cryptdin mRNA resemble the P lysozyme expression pattern in that they are abundant in small bowel, but, in contrast to CRS4C mRNA, neither is found in spleen or in the marrow of the femur (Fig. 6, A and B, Ouellette and Cordell, 1988). In situ hybridization experiments show that cryptdin mRNA exists in apparent intestinal leukocytes as well as in Paneth cells of mice older than 20 days (Ouellette et al., 1990a). These data suggest that cryptdin/CRS-positive intestinal leukocytes may be activated by mediators that function specifically within the small bowel. Functional analysis of the promoters of cryptdin and CRS genes should provide insights into the mechanism(s) that control differential expression of members of the gene family.

---

and reveal informative comparisons with the regulatory elements of the M and P lysozyme genes. Knowledge of these mechanisms should be helpful in identifying the mesenchymal factors that initiate and maintain the differentiated state of the small intestinal epithelium.

Acknowledgments—We thank Dana Frederick, Robin Huffman, Julia Katz, and Hong Yang for excellent technical assistance and Michael Y. Lin for performing some of the Northern blot hybridizations. We also thank Drs. Tomas Ganz and Forrest H. Fuller for sharing unpublished results and Drs. Marek Kloczewiak and Ronald A. Malt for helpful discussions and suggestions.

REFERENCES
Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J., and Rutter, W. J. (1979) Biochemistry 18, 5294-5299

Intestinal Cationic Peptide cDNAs
A novel mouse gene family coding for cationic, cysteine-rich peptides. Regulation in small intestine and cells of myeloid origin.
A J Ouellette and J C Lualdi


Access the most updated version of this article at http://www.jbc.org/content/265/17/9831

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 0 references, 0 of which can be accessed free at http://www.jbc.org/content/265/17/9831.full.html#ref-list-1