Cloning and sequence analysis of cDNA for the canine neurotensin/neuromedin N precursor

(Received 14 December 1986; revised 9 March 1987. Accepted 27 March 1987)

PAUL R. DOBNER†, DIANE L. BARBER†, LYDIA VILLA-KOMAROFF§, AND COLLEEN MCKIERNAN*†

Departments of *Molecular Genetics and Microbiology, and †Physiology, University of Massachusetts Medical Center, 55 Lake Avenue North, Worcester, MA 01605; and §Department of Neuroscience, The Children’s Hospital, 300 Longwood Avenue, Boston, MA 02115

Communicated by Sanford L. Palay, February 5, 1987

ABSTRACT  Cloned cDNAs encoding neurotensin were isolated from a cDNA library derived from primary cultures of canine enteric mucosal cells. Nucleotide sequence analysis has revealed the primary structure of a 170-amino acid precursor protein that encodes both neurotensin and the neurotensin-like peptide neuromedin N. The peptide-coding domains are located in tandem near the carboxyl terminus of the precursor and are bounded and separated by the paired, basic amino acid residues Lys-Arg. An additional coding domain, resembling neuromedin N, occurs immediately after an Arg-Arg basic amino acid pair located in the central region of the precursor. Additional amino acid homologies suggest that tandem duplications have contributed to the structure of the gene. RNA blot analysis, using the cloned cDNA probe, has revealed several mRNA species ranging in size from 500 to 980 nucleotides in the canine enteric mucosa. In contrast, a single RNA species of 1500 nucleotides was detected in bovine hypothalamus poly-(A)* RNA. The ability of the canine probe to cross-hybridize with bovine mRNA suggests that this probe can be used to isolate neurotensin/neuromedin N genes from other mammalian species.

The peptide neurotensin was first isolated from extracts of bovine hypothalamus (1). In addition to a wide distribution throughout the central nervous system where neurotensin is thought to act as a neurotransmitter (2), high concentrations of neurotensin are also found in the mucosa of the distal small intestine (3). In mammals, cells containing neurotensin-like immunoreactivity are scattered in the epithelium of the jejunum–ileum and contain electron-dense secretory granules primarily in the basal portion of the cell; a narrow apical process contributes to brush border and contacts the lumen (4). Neurotensin levels increase in hepatic-portal blood after perfusion of the rat small intestine with fatty acid or alcohol (5), and infusion of neurotensin into the rat superior mesenteric artery results in increased translocation of labeled oleic acid from the lumen of the small intestine into the lymph (6). Thus, neurotensin may play an endocrine or paracrine role in the regulation of fat metabolism.

Bovine hypothalamus and intestine neurotensin are identical (7), and an essentially identical substance has been isolated from the human small intestine (8). Distinct, but structurally related, peptides have been isolated from both mammalian and nonmammalian tissues (9–12). The mammalian neurotensin-like peptide neuromedin N was isolated during an exhaustive search for biologically active peptides in extracts of porcine spinal cord (12). Like neurotensin, neuromedin N causes contraction of the guinea pig ileum and a hypertensive response in anesthetized rats although it is less potent than neurotensin (12). The family of neurotensin-related peptides possess common carboxyl-terminal sequences but differ at the amino terminus.

A method for obtaining primary cultures enriched in neurotensin-containing cells from the canine enteric mucosa has been described (13). We have used these enriched cells to isolate cDNA clones encoding neurotensin. Nucleotide sequence analysis of the cloned cDNA has revealed the primary structure of a precursor protein that encodes both neurotensin and neuromedin N.

MATERIALS AND METHODS

Materials. The following items were obtained from the indicated sources: 32P-labeled nucleotides, Amersham; nucleotides, P-L Biochemicals; restriction enzymes, Boehringer Mannheim and New England Biolabs; reverse transcriptase, Life Sciences (St. Petersburg, FL), cDNA Library Construction. Enteric mucosa cells from adult canine ileum were collected and enriched for those cells displaying neurotensin-like immunoreactivity as described (13). RNA was extracted by the guanidine thiocyanate procedure (14) and poly(A)* RNA was selected by passage over oligo(dT)-cellulose (15).

Double-stranded cDNA was synthesized from 5 µg of poly(A)* RNA, digested with S1 nuclease, and modified by EcoRI methylase as described (16). After addition of EcoRI linkers and digestion with EcoRI, the cDNA was size-selected on 1% low-melting-temperature agarose gel. Fragments migrating more slowly than a 580-base-pair marker were recovered from the agarose by extraction with butanol containing hexadecyltrimethylammonium bromide (17) and further purified by spermine precipitation (18). The final cDNA preparation was inserted into the EcoRI site of the λgt11 bacteriophage (19). DNA was packaged in vitro using a commercial extract (Stratagene), and portions of the reaction were used to infect the Escherichia coli strain Y1088 (19).

Screening Procedures. Bacteriophage plaques (20,000 per 15-cm plate) were transferred to nitrocellulose filters and lysed in situ (20). The filters were baked for 2 hr at 80°C and prehybridized in 6x NaCl/Cit (1 x NaCl/Cit = 0.15 M NaCl/0.015 M sodium citrate, pH 7.0), 50 mM sodium phosphate (pH 7), 10x Denhardt’s solution (1 x Denhardt’s solution = bovine serum albumin at 0.2 mg/ml, polyvinylpyrrolidone at 0.2 mg/ml, Ficoll at 0.2 mg/ml), yeast RNA at 100 µg/ml, and 0.5% NaDodSO4 for 3–16 hr at 42°C. Hybridization with oligonucleotide probes was in the same solution except that dextran sulfate was added to a final concentration of 10% (wt/vol). Oligonucleotides were synthesized on an Applied Biosystems DNA synthesizer and end-labeled using T4 polynucleotide kinase and [γ-32P]ATP (21). Labeled oligonucleotides were not separated from unincorporated label but added directly to the hybridization mixture (at 7 ng/ml). After hybridization for 12 hr at room temperature.
temperature, the filters were washed using a high-stringency tetramethylammonium chloride salt wash (22). The optimal wash temperature was determined to be 44°C for the tetradecamer oligonucleotide mix used.

DNA Sequencing. Positive bacteriophage were purified by several rounds of plating. DNA was prepared from small volumes of liquid culture (21) and digested with EcoRI. The EcoRI inserts were purified by electrophoresis on 1% low-melting-temperature agarose gels and ligated into either M13mp11 or pUC12 (23) directly from the gel slice. Various defined fragments were sequenced by the dideoxy method (24).

Single-Stranded Probe Preparation and RNA Blotting. A 32P-labeled antisense single-stranded probe was synthesized from an M13mp11 subclone containing the EcoRI fragment derived from NT-8 (25). The single-stranded fragment was purified by electrophoresis on a 5% acrylamide/8 M urea gel. The labeled fragment was excised from the gel, crushed, and eluted directly into RNA blot hybridization buffer (50% [vol/vol] formamide, 2x Denhardt’s solution, 20 mM sodium phosphate (pH 7), 0.1% NaDodSO4, 10% [wt/vol] dextran sulfate) minus dextran sulfate, which was added after removal of the gel fragments by centrifugation. RNA blotting from formaldehyde/agarose gels was performed as described (26).

RESULTS

Isolation of cDNA Clones. To identify cDNA clones encoding neurotensin, an oligonucleotide mixture comprising all possible complementary sequences corresponding to the amino acid residues 3–7 (Fig. 1) was chemically synthesized for use as a hybridization probe. Initial attempts to identify neurotensin cDNA clones in a bovine hypothalamus cDNA library consisting of several million independent recombinants were not successful. The levels of neurotensin-immunoreactive material in the distal small intestine are approximately the same as in the hypothalamus (3). A method for enriching the scattered cells of the canine enteric mucosa that react with antibodies against neurotensin has been developed (13). These partially purified cells represented a potentially enriched source of neurotensin mRNA and were used as starting material for the preparation of a cDNA library.

Poly(A)+ RNA was isolated from 8 × 107 partially purified canine mucosa cells and used for the construction of a λgt11 cDNA library. Since bovine hypothalamus and intestine neurotensin are identical and a variety of studies have indicated that the sequence is conserved among mammals (7, 8), the oligonucleotide probe predicted from the bovine amino acid sequence was used to screen the canine cDNA library. From the initial plating of 170,000 independent recombinants, 14 hybridization-positive clones were isolated, and three of these, designated NT-4, NT-5, and NT-8, were further characterized.

Bovine Neurotensin
(pyrOGl)
Gln-Leu-Tyr-Glu-Asn-Lys-Pro-Arg-Arg-Pro-Tyr-Ile-Leu-Oh
5'-CAG-CTT-TAT-GAA-AAT-AAA-CCT-CGT-CCT-TAT-ATT-CTT-3'
A T C C G C G C A C A C C C C T C
A A A A A A A G G G G G
Oligonucleotide Probe
3'-ATA-CTT-TTT-GG-5'
G C G C

Fig. 1. Synthetic oligodeoxynucleotides used for hybridization probes. All possible coding sequences for bovine neurotensin and the sequence of the tetradecamer hybridization probe are shown.

A Common Precursor for Neurotensin and Neuromedin N Revealed by Nucleotide Sequence Analysis. Nucleotide sequences were determined by the dideoxy method using the strategy depicted in Fig. 2A. The clone containing the largest EcoRI insert (NT-8, 729 base pairs) was fully sequenced on both strands. The complete sequence of the NT-5 EcoRI insert and part of the sequence of NT-4 were also determined, and the composite sequence is presented in Fig. 2B. An open reading frame extends from the beginning of the sequence to nucleotide 544 followed by a 3'-untranslated region of 197 bases. A sequence identical to that of bovine neurotensin, except for the amino-terminal glutamic residue, is located near the carboxyl terminus of the open reading frame. The difference in the amino-terminal residue is likely to be the result of cyclization of the glutamic residue to form a pyroglutamyl residue during the acid-extraction procedure used to isolate neurotensin (1). A sequence identical to the porcine neurotensin-like peptide neuromedin N (12) is located just prior to the neurotensin sequence, separated only by the paired, basic amino acid residues Lys-Arg.

We have assigned the first methionine codon encountered in the open reading frame as the first residue of the predicted precursor (Fig. 2B). The assignment is tentative since there is no in-phase stop codon upstream of this methionine codon and we are not certain that the composite sequence is full length. However, the 5' ends of the three cDNAs analyzed are closely spaced (within 12 base pairs of each other), suggesting that either a strong block to reverse transcription or the 5' end of the mRNA template occurs in this region. The latter possibility is more likely since, allowing for a poly(A) tail of 150–200 nucleotides, the size of the composite sequence is approximately the same as the size of the largest mRNA species detected by RNA blot analysis (see below; Fig. 3, lane A). Translation from the selected initiation codon would result in the production of a 170-amino acid protein with a predicted molecular weight of 19,864. The amino-terminal portion of the protein contains a stretch of hydrophobic amino acid residues typical of the signal peptides that precede most secreted proteins (27).

RNA Blot Analysis. To determine the size of neurotensin mRNA, a 32P-labeled probe, corresponding to the antisense strand of the NT-8 EcoRI insert, was used for RNA blot analysis (Fig. 3). Several mRNA species are detected in the RNA extracted from canine mucosa cells enriched for neurotensin immunoreactive cells (lane A). The largest of these is 980 nucleotides, and less clearly resolved smaller species between 500 and 700 nucleotides are also apparent. The broad size distribution of these mRNA species is not likely due to RNA degradation since rehybridizing the blot with a β-tubulin probe revealed a single nondegraded 1.8-kilobase mRNA (data not shown).

To examine the possibility that our failure to isolate neurotensin cDNAs from a bovine hypothalamus library was due to the relative low abundance of the corresponding mRNA in this tissue, we examined the poly(A)+ RNA
A composite sequence determined from the NT-4, NT-5, and NT-8 cDNA inserts indicate by downward pointing arrowheads. The neuronotensin and neuromedin N coding regions are boxed by solid lines, and a region similar to neuromedin N is boxed by dashed lines. A single-nucleotide difference in NT-5 and the resulting amino acid difference are indicated (nucleotide 202). A potential polyadenylation signal is underlined. The protein sequence begins at the first methionine codon encountered in the open reading frame.

preparation used to create this cDNA library on the same RNA blot (Fig. 3, lane B). The canine probe clearly identifies a single mRNA species of 1500 nucleotides. Although the difference in mRNA size could represent a species difference, it is possible that different mRNAs are produced in neuronal and gastrointestinal tissues. The levels of neurotensin mRNA in the bovine hypothalamus appear to be much lower than in the enriched canine intestine cells since a much longer exposure was required to visualize the band in lane B than the bands in lane A. Although this is consistent with our results from screening cDNA libraries prepared from these two RNA preparations, the signals obtained cannot be directly compared since the canine probe may not hybridize efficiently to the bovine RNA.

**DISCUSSION**

Many peptides that subserve endocrine and/or neural functions are synthesized as large polypeptide precursors that are subsequently processed to yield the biologically active peptides (28). Generally, the peptides are flanked within the precursor by paired, basic amino acid residues that provide cleavage signals for trypsin-like proteases. The neurotensin/neuromedin N precursor (drawn schematically in Fig. 4) has this same general structure since both of these peptides are bounded by paired, basic amino acid residues. The Arg-Arg pair that occurs within neurotensin must somehow be protected from cleavage or not be recognized by the processing enzyme. An additional set of basic residues (amino acids 86-88, Fig. 2B) could also be cleaved. Interestingly, the amino acid sequence in this region resembles the neuromedin N region. However, the neuromedin N-like peptide is bounded by paired, basic amino acids only after the amino terminus; cleavage at the next basic pair, just prior to the neuromedin N sequence, would result in a 53-amino acid peptide.

The tandem positioning of the neurotensin and neuromedin N coding regions on the same precursor suggests that individual cells may produce both peptides. Neuromedin N was isolated from extracts of porcine spinal cord; however, the distribution of the peptide in other tissues has not been examined. Neuromedin N-immunoreactive material has been detected in extracts of the enriched neurotensin-containing canine mucosa cells used here (D.L.B. and R.E. Carraway, unpublished observations). However, further immunohistochemical experiments are required to determine whether or not individual cells in this population produce both neurotensin and neuromedin N as is suggested by the structure of the precursor.

Avian counterparts of neurotensin and neuromedin N have been isolated from the chicken gastrointestinal tract. Chicken neurotensin is a triapeptide that varies from the bovine sequence at three amino-terminal amino acid residues (10). A 6-amino acid peptide, LANT-6, differs at only one amino acid...
residue with porcine neuromedin N (11). By using specific antisera that distinguish between avian neurotensin and LANT-6 (29), the two peptides have been colocalized in the same endocrine cells in the gastrointestinal tract of the quail, although cells that react with one or the other antibody are also observed (30). Similar results were obtained in reptiles, amphibians, and bonyfish, but, to our knowledge, these findings have not yet been extended to mammalian species.

The establishment of cell populations that differentially express two peptides that are encoded by the same gene could be accomplished in two ways. Like preproopiomelanocorticotropin (28), the precursor protein could be alternatively processed in a tissue-specific fashion, or, as is the case for the preproterphanin gene (31), distinct mRNAs encoding different sets of peptides could be spliced from the same primary transcript. We have completely sequenced two independent isolates (NT-5 and NT-8) and both encode neurotensin and neuromedin N. However, RNA blot analysis of the RNA preparation used to construct the canine cDNA library has revealed several species of mRNA that cross-hybridize with the NT-8 EcoR1 insert probe (Fig. 3). The NT-8 clone EcoR1 insert was most likely derived from the largest species of mRNA detected (980 bases). The smaller RNA species, which are not well resolved, could represent differentially spliced mRNAs potentially encoding different precursor proteins analogous to the preproterphanin example. However, they could also result from polyadenylation at different sites or be transcribed from distinct but related genes.

The identity of the carboxyl-terminal regions of neurotensin and neuromedin N and their tandem arrangement within the precursor suggests that the two peptide-coding domains arose by tandem duplication of an ancestral sequence. The existence of homologous peptides in chickens suggests that this duplication predated the divergence of mammals and birds. In addition, a much larger tandem duplication appears to have doubled the size of the coding region giving rise to the neuromedin N-like coding domain (Fig. 4). Thus, the region of the precursor between residue 94 and the last residue of the neuromedin N coding domains bears a striking resemblance to the region between residue 32 and the last residue of the neuromedin N-like domain (Fig. 5). The neurotensin coding domain either was not included in this duplication or arose afterwards.

The primordial coding domain that has apparently been duplicated may be quite ancient. The peptide xenopsin, isolated from the skin of Xenopus laevis, shares four of five carboxyl-terminal amino acid residues with neurotensin (9). A precursor for xenopsin has been postulated from the nucleotide sequence of a cDNA clone (32). The xenopsin precursor bears little resemblance to the neurotensin/neuromedin N precursor, however, direct comparison between the two amino acid sequences has revealed a four-amino acid identity in addition to the carboxyl-terminal resemblance between xenopsin and neurotensin. The sequence Glu-Ala-Met-Leu occurs just after a Lys-Arg pair in the xenopsin precursor (residues 64–67, ref. 32) and just downstream from the neuromedin N-like sequence (residues 99–102, Fig. 2B) in the neurotensin/neuromedin N precursor. Furthermore, this amino acid sequence lies within a region that constitutes the most conserved portion of the large tandem duplication (compare residues 98–103 and residues 33–39, see Fig. 5) we have postulated for the canine neurotensin gene. The significance of this homology is uncertain; however, it does suggest that the xenopsin and neurotensin genes arose from a common ancestral sequence.

The amino acid sequences of both neurotensin and neuromedin N, predicted from the nucleotide sequence of the canine cDNA, are identical to the sequences determined by direct protein sequencing of bovine neurotensin (33) and porcine neuromedin N (12) assuming that, in the case of bovine neurotensin, the amino-terminal pyroglutamate residue is the result of cyclization of a glutamine residue. The canine probe we have isolated cross-hybridizes with a bovine hypothalamus mRNA using standard high-stringency conditions (Fig. 3, lane B) indicating that regions of the mRNA
outside the peptide coding domains are also conserved. The canine probe should, therefore, enable the isolation of the corresponding gene from other mammalian species.

We thank Dr. Susan E. Leeman for advice and enthusiastic support and Dr. Andrew H. Soll in whose laboratory the canine cells were collected (supported by Grant AM3044 from the National Institutes of Health). We also thank Drs. Janet Stavnezer and Duane Jenness for helpful comments on the manuscript. This work was supported by a grant from the National Institutes of Health (HL33070) to P.R.D. Additional support was provided to P.R.D. from BRSG grant S07 RR5712 awarded by the Biomedical Research Support Grant Program, Division of Research Resources, National Institutes of Health. D.L.B. is the recipient of a National Institutes of Health postdoctoral fellowship (AM07662).