Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 9085–9089, October 1993
Biochemistry

The human gastrin/cholecystokinin type B receptor gene: Alternative splice donor site in exon 4 generates two variant mRNAs

(gastrointestinal hormone/G protein-coupled receptor/gastric function/alternative splice junction/chromosome 11)

IL SONG*, DAVID R. BROWN*†, RODNEY N. WILTSHIRE‡§, IRA GANTZ¶, JEFFREY M. TRENT‡§, AND TADATAKA YAMADA*‖

Departments of *Internal Medicine, ‡Radiation Oncology, §Human Genetics, and ¶Surgery, The University of Michigan Medical Center, Ann Arbor, MI 48109

Communicated by Horace W. Davenport, July 12, 1993

ABSTRACT **Gastrin and its carboxyl-terminal homolog cholecystokinin (CCK) exert a variety of biological actions in the brain and gastrointestinal tract that are mediated in part through one or more G protein-coupled receptors which exhibit similar affinity for both peptides. Genomic clones encoding a human gastrin/CCK_B receptor were isolated by screening a human EMBL phage library with a partial-length DNA fragment which was based on the nucleotide sequence of the canine gastrin receptor. The gene contained a 1356-bp open reading frame consisting of five exons interrupted by 4 introns and was assigned to human chromosome 11p15.4. A region of exon 4, which encodes a portion of the putative third intracellular loop, appears to be alternatively spliced to yield two different mRNAs, one containing (452 amino acids; long isoform) and the other lacking (447 amino acids; short isoform) the pentapeptide sequence Gly-Gly-Ala-Gly-Pro. The two receptor isoforms may contribute to functional differences in gastrin- and CCK-mediated signal transduction.**

The gastrointestinal hormone gastrin is synthesized and released from G cells in the gastric antrum. Its putative physiological actions include stimulation of gastric acid production (1), induction of gastric smooth muscle contraction (2), and promotion of growth and differentiation of the stomach mucosa (3). Cholecystokinin (CCK), a peptide possessing an identical C-terminal pentapeptide sequence to gastrin, mimics the actions of gastrin in the stomach. On the basis of pharmacological and radioligand binding studies, the diverse effects of CCK in peripheral organs and the central nervous system are thought to be mediated by two receptor types (4, 5). The CCK_A receptor has been localized to the gallbladder, pancreas, vagal afferent neurons, spinal cord, and certain mesencephalic nuclei. It is selectively blocked by the asperlicin analog devazepide and exhibits a relative affinity for sulfated CCK that is 3 orders of magnitude higher than for gastrin. The cDNAs encoding this receptor have been cloned from rat pancreas (6) and human gallbladder (7). The CCK_B receptor is more ubiquitously distributed in the brain than the CCK_A receptor, being chiefly found in hypothalamic nuclei, the limbic system, and the frontal cortex. Unlike the CCK_A receptor, the CCK_B receptor exhibits nearly equal affinity for gastrin and both sulfated and nonsulfated forms of CCK. Moreover, CCK_B antagonists which are currently available can discriminate between CCK_A and CCK_B receptors with up to 1000-fold selectivity (8). A cDNA encoding the putative CCK_B receptor has been cloned from AR42-J cells (a rat pancreatic acinar cell line) and rat cerebral cortex. It encodes a 452-aa protein manifesting 48% amino acid identity with the CCK_A receptor (9). Recently, an homologous CCK_B receptor cDNA has been cloned from human cerebral cortex (10, 11).

Receptors for gastrin that have been characterized on parietal cells and gastric myocytes have a pharmacological profile which is similar to that of the CCK_B receptor. Indeed competitive CCK_B antagonists such as L-365,260 and PD 134308, which readily discriminate between CCK receptor types, are incapable of selectively differentiating CCK_B and gastrin receptors (8). Recently, gastrin receptor cDNAs have been isolated from cDNA libraries derived from canine parietal cells and rodent (*Mastomys*) enterochromaffin-like tumor cells (12, 13). They encode homologous proteins of similar length (453 and 450 aa, respectively) with seven deduced transmembrane regions, comparable to other G protein-coupled receptors. The recombinant receptors, when transfected into COS-7 cells, manifest high (nanomolar) affinities for both gastrin and CCK and are coupled to an increase in intracellular calcium (12, 13). Structurally, the canine gastrin receptor displays nearly 90% amino acid identity with the putative CCK_B receptors from human and rat (9–11).

The sequence and organization of the full-length genes encoding members of the gastrin/CCK receptor family have not been identified. Such information is essential for analysis of their transcriptional regulation and structural relationships to other receptor genes. We have cloned and sequenced a gastrin/CCK_B receptor gene from a human genomic library**. The gene appears to be alternatively spliced to yield two different receptor isoforms which differ by 5 aa in the putative third intracellular loop. The gastrin/CCK_B receptor gene was assigned to human chromosome 11p15.4 by fluorescence *in situ* hybridization.

METHODS

Genomic Library Screening and DNA Sequencing. A human genomic library in the bacteriophage λEMBL-3 (Clontech) was screened with a ^{32}P-labeled cDNA fragment generated by the polymerase chain reaction (PCR) with primers based on the cDNA sequence of the canine gastrin receptor (see below). Two positive clones were digested with *Sst* I and the resulting restriction fragments of length 0.7–3.2 kb (Fig. 1) were subcloned into phage M13mp18 or -mp19 and sequenced in both directions by the dideoxy method (14). In

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.

Abbreviation: CCK, cholecystokinin.

†Present address: Department of Veterinary PathoBiology (Pharmacology Section), University of Minnesota, 1988 Fitch Avenue, St. Paul, MN 55108.
‖To whom reprint requests should be addressed at: Department of Internal Medicine, 3101 Taubman Center, University of Michigan Medical Center, Ann Arbor, MI 48109-0368.

**The sequence reported in this paper has been deposited in the GenBank data base (accession no. L10822).

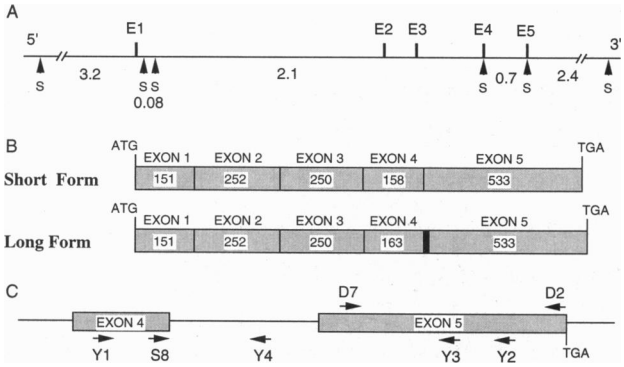

FIG. 1. Structure of the human gastrin/CCK_B receptor gene. (*A*) Restriction map and exon location. *Sst* I restriction sites (underscored by S with arrowhead) within the gene were located in intron 1 and exons 4 and 5. The *Sst* I restriction fragments were subcloned into the M13 vector and sequenced. (*B*) Two variant RNAs encode the gastrin/CCK_B receptor. The sizes of exons (bp) appear within rectangles. The RNA variants differ only in the size of exon 4, in which a sequence of 15 nt (black bar in long form) is either absent (short form) or present (long form). The ATG in exon 1 and TGA in exon 5 indicate sites of translation initiation and termination, respectively. (*C*) Oligonucleotide primers used to amplify DNA fragments by PCR (see also Table 1). Primers D7 and D2 were used to generate the 390-bp DNA fragment used as a probe to screen canine and human genomic libraries. Y2 was used as a primer for reverse transcription of human fundic RNA to cDNA, and primers Y1 and Y3 were employed in PCR amplification of the exon/intron boundary between exons 4 and 5. Primers Y1 and Y4 were used to amplify the sequence between exon 4 and intron 4 in human genomic DNA to confirm the existence of a single gene. Primer S8 was used together with primer Y3 to amplify the long isoform from human stomach RNA.

addition, an 81-bp DNA fragment was determined to span two *Sst* I sites in the first intron as determined by sequencing an *Sph* I restriction fragment (restriction site in exon 2) in the 3'-to-5' direction. A canine genomic library (Clontech) was also screened for this gene. Oligonucleotide primers for sequencing or PCR were synthesized on a DNA synthesizer (model 380B; Applied Biosystems). Computer analyses of nucleotide sequences were performed with the Genetics Computer Group (GCG) program (University of Wisconsin Biotechnology Center, Madison, WI). Nucleotide sequences were translated to amino acid sequences by using the IUB code in the GCG program (15).

PCR Cloning. RNA was extracted (16) from canine and human gastric fundic mucosa. Total RNA (10–20 μg) was reverse-transcribed to cDNA (17), and this cDNA served as the template for subsequent PCRs. PCR primers D2 and D7, which contained *Bam*HI restriction sites at their 5' ends (Fig. 1 and Table 1) were used to generate a 390-bp probe for screening human and canine genomic libraries. These sequences were based on the canine gastrin receptor cDNA sequence (12) and correspond to nt 2810–2839 (primer D7) and 3167–3199 (primer D2) of the human receptor gene (Fig. 2). PCR was performed with *Taq* DNA polymerase (Promega) through 30 cycles of denaturation (1 min at 94°C), annealing (2 min at 50°C), and extension (3 min at 72°C), with a final extension period of 10 min at 72°C. PCR products were purified by phenol/chloroform extraction and digested with *Bam*HI for ligation into M13mp18 and sequence confirmation. The 390-bp cDNA fragment derived from canine fundic RNA was labeled with ^{32}P by random priming and used as a screening probe. To analyze human stomach gastrin/CCK_B receptor cDNAs, either oligonucleotide Y2 or a random hexamer was used as a primer for reverse transcription of human fundic total RNA to cDNA. Oligonucleotides Y1 and Y3 were used to amplify the exon/intron splicing junction occurring between exons 4 and 5. Primer Y1 contained internal *Sst* I restriction sites, and primer Y3 possessed an *Sst* I site at its 5' end. The resulting PCR product (646 bp) was sequenced as described above. Primer Y4 was constructed as a 20-mer based on a sequence in intron 4. Primer S8 was used in combination with primer Y3 to verify the long receptor RNA isoform (see Fig. 1 for primer locations and Table 1 for sequence positions).

Chromosomal Mapping. To identify its chromosomal location, 1 μg of DNA from a clone of the human gastrin/CCK_B receptor gene (≈10 kb in length) was labeled with biotin and hybridized to human metaphase chromosomes (18). The chromosomes identified by fluorescence *in situ* hybridization were previously G-banded (using trypsin/Giemsa stain) and photographed with a Zeiss Axiophot microscope equipped with a dual-bandpass fluorescein/rhodamine filter to allow direct comparisons of the results.

RESULTS AND DISCUSSION

Organization of the Gastrin/CCK_B Receptor Gene. A random-primed ^{32}P-labeled canine gastrin receptor cDNA probe gave positive hybridization signals with 2 clones of 5×10^5 plaques from a human genomic library. One of the human genomic clones was incomplete (exon 1 absent) and the other clone, H50-1, contained the full-length gene encoding the gastrin/CCK_B receptor. An incomplete gastrin/CCK_B receptor gene lacking exon 1 was also isolated from a canine genomic library. The human receptor gene exceeded 8 kb in length and contained a 1356-bp open reading frame which was interrupted by four introns of 164–1177 bp (Figs. 1 and 2). Its deduced amino acid sequence delineates it as a member of the G protein-coupled, heptahelical receptor superfamily. Exon 1 encodes the putative extracellular amino terminus of the receptor. Exons 2 and 3 encode transmembrane regions I–IV, and exon 4 encodes the fifth transmembrane region and an initial portion of the third intracellular loop. Exon 5 encodes the remainder of this intracellular loop, the remaining transmembrane regions VI and VII, the intracellular carboxyl terminus, and the 3' untranslated region (Fig. 3). The 5' region immediately upstream from the ATG initiation codon to nucleotide −293 manifests a high G+C content (76%). A polyadenylylation site (AATAAA) occurs 558 bp downstream from the stop (TGA) codon (Fig. 2).

Table 1. Sequence of oligonucleotide primers used for PCR

Primer	Sequence (5' → 3')	Orientation	Location/exon*
D2†	TCAGCCAGGCCCTAGCGTGCTGATGGTGGTGTA	Antisense	3199–3167/5
D7†	CGGCCCTACCAGGCCAAGCTGTTGGCCAAG	Sense	2810–2839/5
Y1	GTGGCCTACGGGCTTATCTCTCGCGAGCTCTACTTA	Sense	2354–2389/4
Y2	CATCGGGAAGAGCCCTGGGGCGAGCTCGTGGAGGCC	Antisense	3122–3087/5
Y3	GAGGCGTAGCTCAGCAAGTGAATGAAGGAGAT	Antisense	3000–2969/5
Y4	tcccagctcaactccatttc	Antisense	2589–2570/intron 4
S8	CAGGTGGGGCTGGAC	Sense	2457–2471/4

*Based on the human gastrin/CCK_B receptor gene sequence shown in Fig. 2.
†The sequence of primers D2 and D7 is from the canine gastrin receptor cDNA.

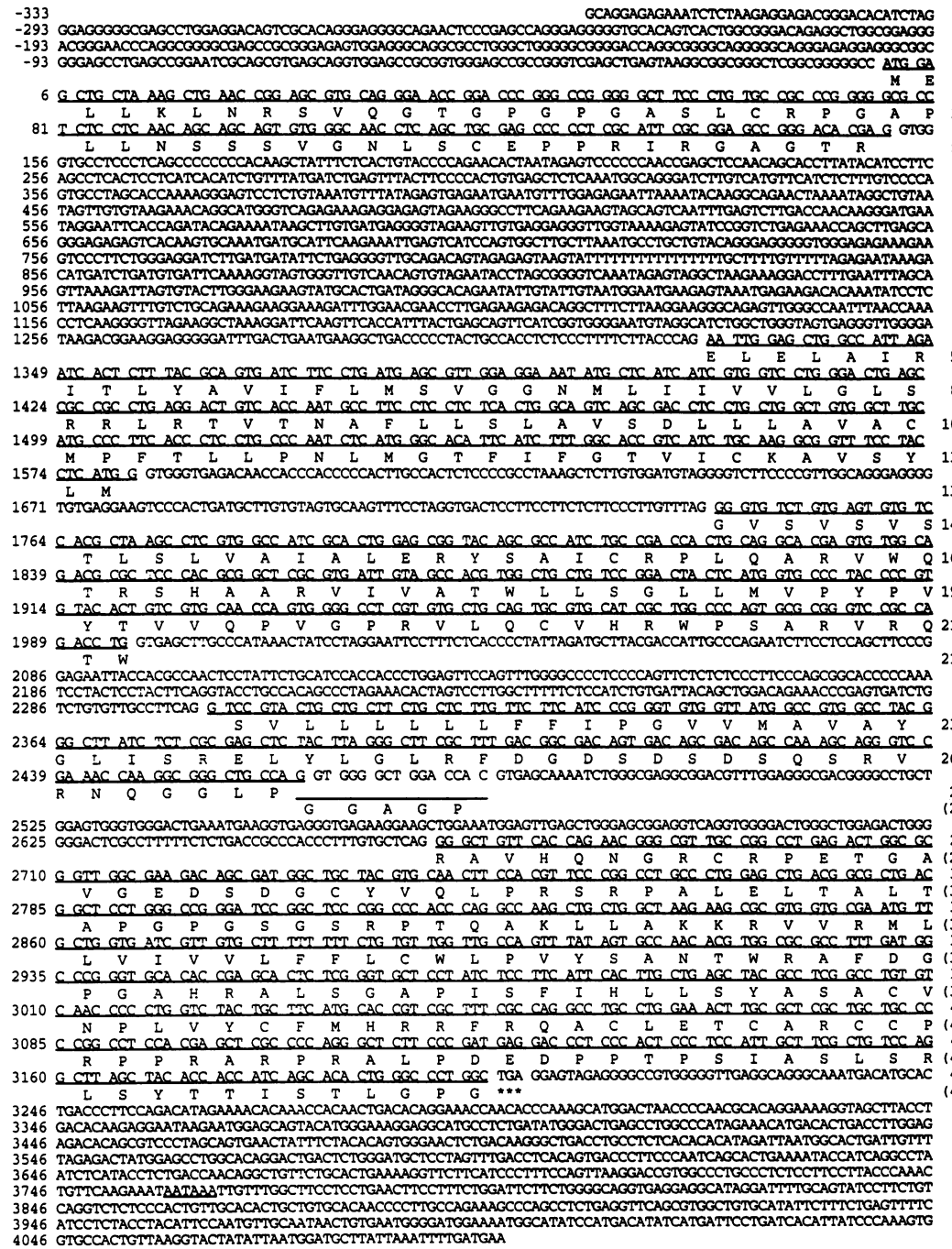

FIG. 2. Nucleotide and deduced amino acid sequence of the human gastrin/CCK_B receptor gene. The five exons are underlined. The nucleotide sequence is numbered on the left relative to the first nucleotide of the translation initiation codon; amino acids are numbered on the right and are depicted in the single-letter symbols below each triplet codon. The putative stop codon (TGA) at nt 3197–3199 is underscored by asterisks, and a polyadenylylation consensus motif (AATAAA) is underlined at nt 3758–3763. In exon 4, alternative splicing results in the expression of the pentapeptide cassette Gly-Gly-Ala-Gly-Pro (GGAGP) at nt 2460–2474 (overlined amino acids).

The interrupted open reading frame of the gene exhibited 88% and 100% nucleotide identity with canine gastrin receptor and human CCK_B receptor cDNA sequences, respectively (10–12). In contrast, the protein-coding region of the gene manifests ≈50% amino acid identity with rat and human CCK_A receptor cDNAs (6, 7). Many genes encoding G protein-coupled receptors, such as those encoding most types of adrenergic and muscarinic cholinergic receptors, possess an intronless open reading frame. Nevertheless, several exceptions are known to exist. Genes encoding three receptor subtypes mediating the actions of another class of gut peptides, the tachykinins, are among these intron-containing genes. The nucleotide sequences of the tachykinin receptor genes display a relatively high homology to that of the gastrin/CCK_B receptor and also contain four introns (19).

Alternative Splicing of Exon 4. The human gastrin/CCK_B receptor gene contained exon/intron boundaries at aa 51, 135, 218, and 271 (or 276) that conformed to known splice junction consensus sequences (20). In a region of exon 4, however, two of these consensus sequences were found in close proximity (Fig. 3). This region of exon 4 encodes a portion of the third intracellular loop, an important site for effector coupling and regulatory phosphorylation in most G protein-coupled receptors. Using PCR with strategically designed primers (Fig. 1 and Table 1), we determined that the human gastrin/CCK_B receptor gene was alternatively spliced

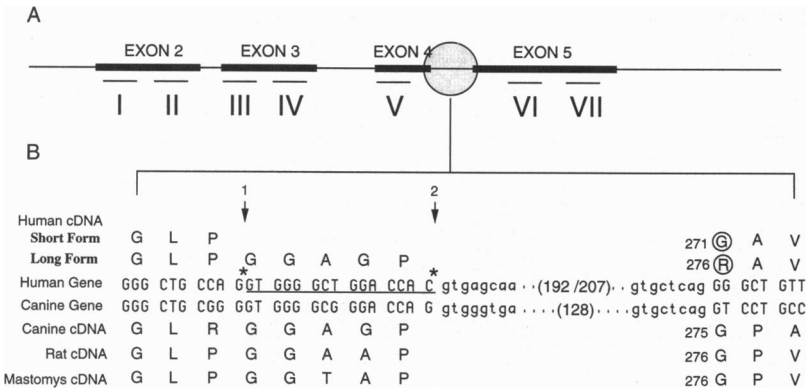

FIG. 3. Alternative splicing of the gastrin/CCK_B receptor gene. (*A*) Relationship of exons to putative membrane-spanning regions I–VII. Circled area highlights region of alternative splicing. (*B*) Alternative RNA splicing in exon 4. The pentapeptide cassette GGAGP encoded by 15 nt (underlined in the human gene) is located at aa 271–275 in the third intracellular loop of the long receptor isoform. Asterisks indicate splice sites. RNA splicing at arrow 1 (G/gt) yields a glycine (G) residue at position 271 of the short isoform, and at arrow 2 (C/gt) an arginine (R) residue at position 276 is translated in the long isoform.

in exon 4. Two mRNA variants were found to exist in human stomach: one, termed the long isoform, contained a block of 15 nt in exon 4 that encoded the pentapeptide cassette GGAGP. A second, short isoform lacked this sequence. Both isoforms were identified to coexist in each of three separate human gastric specimens; thus, the variant mRNAs do not appear to arise from differences between individuals. When the Y1 and Y3 primers were utilized for PCR, only the short isoform was identified at first; the long isoform was identified with the S8 and Y3 primers. However, the long isoform of the gastrin/CCK_B receptor was also identified when the PCR fragments obtained using the Y1 and Y3 primers were used as templates for a second PCR reaction with the S8 and Y3 primers.

The short isoform of the human gastrin/CCK_B receptor is spliced at a G/gt consensus site (Fig. 4, arrow 1); the 15 nt after the splice junction are not included in the mRNA encoding this isoform. In contrast, the long isoform, spliced at C/gt (Fig. 4, arrow 2) encodes the pentapeptide sequence in its mRNA. The pentapeptide cassette (with slight modifications) is also encoded in the gastrin receptor cDNAs from canine parietal cells (12) and *Mastomys* enterochromaffin-like tumors (13) and in the CCK_B receptor cDNA from rat brain (9). However, it is apparently absent in the gastrin/CCK_B receptor cDNA from human brain (10, 11). These species differences may arise from an important difference in genomic sequence at the splice junction in the fourth exon. G/gt is the most common motif for RNA splicing at the 3′ end of an exon (20). In the canine gene, the splice junction in question is G/gt (Fig. 3). In the human gene, however, the comparable site is C/gt; thus a G/gt site 15 bp upstream may become the preferred splice junction, giving rise to the short isoform of the receptor.

We hypothesize that these two mRNA isoforms arise from a single gene. A report by Lee *et al.* (11) indicates that a single gene encodes the human gastrin/CCK_B receptor as determined by Southern blotting of human genomic DNA. Further, when we amplified DNA from (*i*) two different samples of human genomic DNA by PCR using primers Y1 and Y3 or (*ii*) genomic DNA obtained as a contaminant during RNA extraction from the human stomach specimens using the primers Y1 and Y4, only a single gene sequence corresponding to that shown in Fig. 2 was obtained.

Alternative splicing occurs in the third intracellular loop of certain other G protein-coupled receptors, such as the human D_2 and D_3 dopamine receptors (21–23). In the case of the D_2 dopamine receptor, alternative splicing may affect the coupling of the receptor to second-messenger pathways (24, 25).

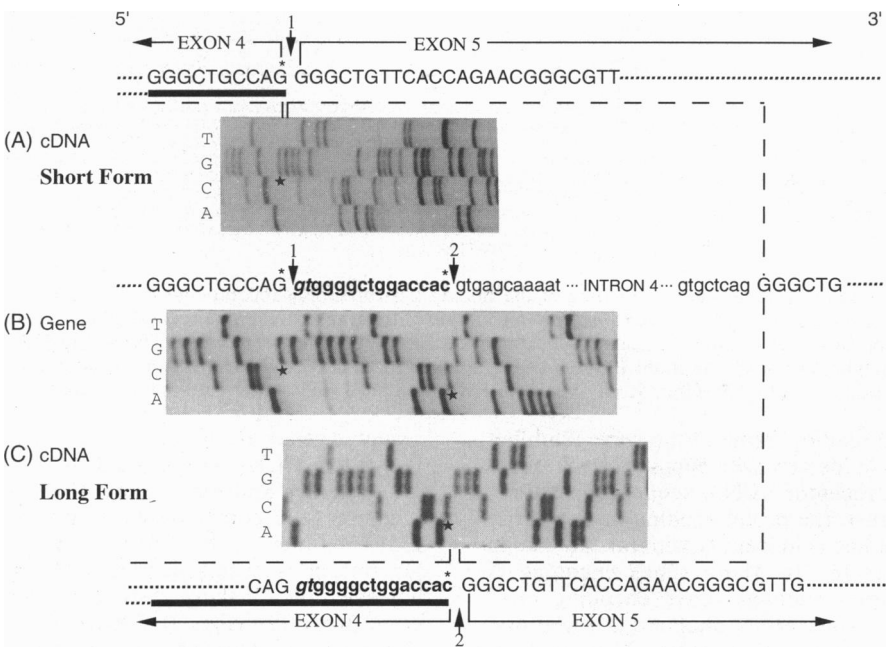

FIG. 4. Sequence analysis of gastrin/CCK_B receptor isoforms observed by PCR amplification of RNA from the fundic mucosa of a single human stomach. In *A*, a cDNA (short form) which lacks a tract of 15 nt appears to arise from a splice junction at G/gt (arrow 1) in the gene (*B*). In *C*, a longer cDNA isoform contains the 15-nt sequence encoding the pentapeptide GGAGP and is spliced at C/gt (arrow 2) in the gene. Asterisks indicate splice sites.

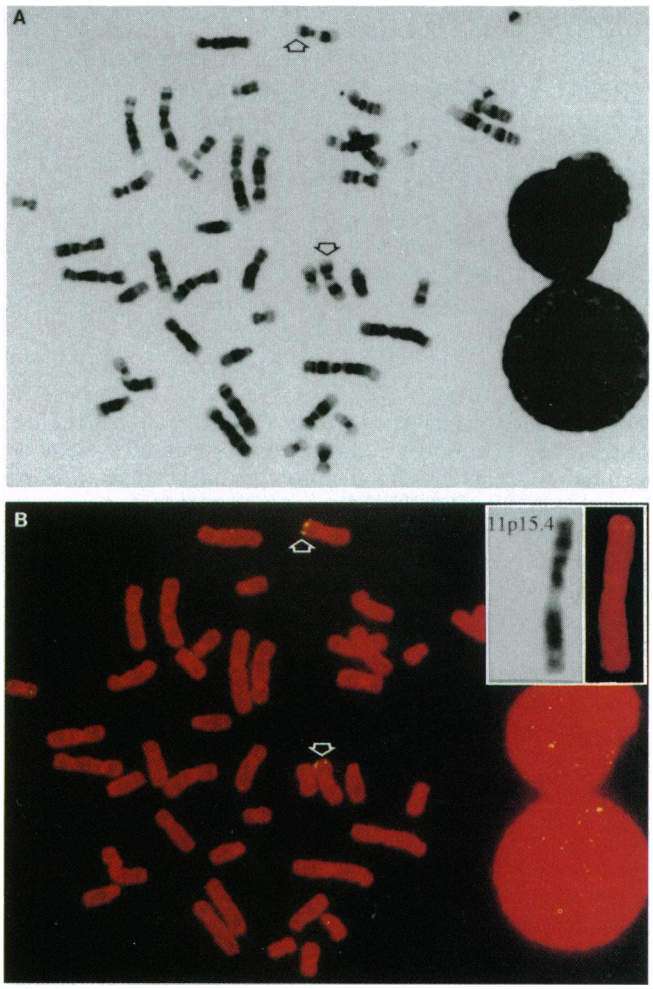

FIG. 5. Localization of the gene encoding a gastrin/CCK_B receptor to human chromosome 11. (*A*) G-banded metaphase chromosome spread. (*B*) Identical metaphase chromosome spread after fluorescence *in situ* hybridization with a biotin-labeled probe of the gastrin/CCK_B receptor gene. Arrows indicate chromosome 11. (*Inset*) Localization of the fluorescent signal to 11p15.4 under a higher resolution of chromosome banding.

Thus, the pentapeptide cassette in the gastrin/CCK_B receptor may be of potential functional importance. If so, aberrations in alternative RNA splicing within the internal exon may contribute to disease states involving the gastrointestinal tract and possibly the central nervous system.

Chromosomal Mapping. Twenty-seven metaphase cells were examined and 24 cells exhibited positive fluorescent signals on the terminal short arm of chromosome 11. Fifty-nine percent (16/27) of the cells manifested specific signals on both homologs, whereas 30% (8/27) displayed a specific signal on one homolog. The results demonstrated the localization of the gastrin/CCK_B receptor gene to chromosome 11p15.4 (Fig. 5). Genes encoding other members of the G protein-coupled receptor superfamily on this chromosome include the D_4 dopamine receptor gene, which maps near the telomere at 11p15.5 (26), and the m_4 muscarinic cholinergic receptor gene, which is localized near the centromere at 11p11–p12 (27). Deletions in the short arm of chromosome 11 have been linked to astrocytic tumors in adults and primitive neuroectodermal tumors of children (28). Determination of the role, if any, of altered gastrin/CCK_B receptor gene expression in these disorders may be worthy of further exploration.

We thank Dr. M. K. Abbas for technical assistance. This work was supported in part by National Institutes of Health Grants RO1-DK34306 and P30-CA46592 and funds from the University of Michigan Gastrointestinal Peptide Research Center (National Institutes of Health Grant P30-DK34933). D.R.B. was a Visiting Research Investigator supported in part by National Research Service Award F33-DK08686.

1. Dockray, G. J. & Gregory, R. A. (1990) in *Handbook of Physiology*, eds. Schultz, S. G., Makhlouf, G. M. & Rauner, B. B. (Am. Physiol. Soc., Baltimore, MD), Vol. 2, Sect. 6, pp. 311–336.
2. Grider, J. R. & Makhlouf, G. M. (1990) *Am. J. Physiol.* **259,** G184–G190.
3. Zachary, I. & Rozengurt, E. (1993) *Handb. Exp. Pharmacol.* **106,** 343–362.
4. Innis, R. B. & Snyder, S. H. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 6917–6921.
5. Moran, T. H., Robinson, P. H., Goldrich, M. S. & McHugh, P. R. (1986) *Brain Res.* **362,** 175–179.
6. Wank, S. A., Harkins, R., Jensen, R. T., Shapira, H., de Weerth, A. & Slattery, T. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 3125–3129.
7. Ulrich, C. D., Ferber, I., Holicky, E., Hadac, E., Buell, G. & Miller, L. J. (1993) *Biochem. Biophys. Res. Commun.* **193,** 204–211.
8. Presti, M. E. & Gardner, J. D. (1993) *Am. J. Physiol.* **264,** G399–G406.
9. Wank, S. A., Pisegna, J. R. & de Weerth, A. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 8691–8695.
10. Pisegna, J. R., de Weerth, A., Huppi, K. & Wank, S. A. (1992) *Biochem. Biophys. Res. Commun.* **189,** 296–303.
11. Lee, Y. M., Beinborn, M., McBride, E. W., Lu, M., Kolakowski, L. F., Jr., & Kopin, A. S. (1993) *J. Biol. Chem.* **268,** 8164–8169.
12. Kopin, A. S., Lee, Y.-M., McBride, E. W., Miller, L. J., Lu, M., Lin, H. Y., Kolakowski, L. F., Jr., & Beinborn, M. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 3605–3609.
13. Nakata, H., Matsui, T., Ito, M., Taniguchi, T., Naribayashi, Y., Arima, N., Nakamura, A., Kinoshita, Y., Chihara, K., Hosoda, S. & Chiba, T. (1992) *Biochem. Biophys. Res. Commun.* **187,** 1151–1157.
14. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
15. Nomenclature Committee (1985) *Eur. J. Biochem.* **150,** 1–5.
16. Chomczynski, P. & Sacchi, N. (1987) *Anal. Biochem.* **162,** 156–159.
17. Gubler, U. & Hoffman, B. (1983) *Gene* **25,** 263–269.
18. Meltzer, P. S., Guan, X.-Y., Burgess, A. & Trent, J. M. (1992) *Nat. Genet.* **1,** 24–28.
19. Gerard, N. P., Bao, L., Xiao-Ping, H. & Gerard, C. (1993) *Regul. Pept.* **43,** 21–35.
20. Mount, S. M. (1982) *Nucleic Acids Res.* **10,** 459–472.
21. Giros, B., Sokoloff, P., Martres, M.-P., Riou, J.-F., Emorine, L. J. & Schwartz, J.-C. (1989) *Nature (London)* **342,** 923–926.
22. Monsma, F. J., Jr., McVittie, L. D., Gerfen, C. R., Mahan, L. C. & Sibley, D. R. (1989) *Nature (London)* **342,** 926–929.
23. Fishburn, C. S., Belleli, D., David, C., Carmon, S. & Fuchs, S. (1993) *J. Biol. Chem.* **268,** 5872–5878.
24. Montmayeur, J.-P. & Borrelli, E. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 3135–3139.
25. Liu, Y. F., Civelli, O., Grandy, D. K. & Albert, P. R. (1992) *J. Neurochem.* **59,** 2311–2317.
26. Gelernter, J., Kennedy, J. L., van Tol, H. H. M., Civelli, O. & Kidd, K. K. (1992) *Genomics* **13,** 208–210.
27. Grewal, R. P., Martinez, M., Hoehe, M., Bonner, T. I., Gershon, E. S. & Detera-Wadleigh, S. (1992) *Genomics* **13,** 239–240.
28. Fults, D., Petronio, J., Noblett, B. D. & Pedone, C. A. (1992) *Genomics* **14,** 799–801.