MOLECULAR AND CELLULAR BIOLOGY, Dec. 1995, p. 7135–7142 Vol. 15, No. 12
0270-7306/95/$04.00+0

An Intron Enhancer Containing a 5′ Splice Site Sequence in the Human Calcitonin/Calcitonin Gene-Related Peptide Gene

HUA LOU,[1] YUN YANG,[1] GILBERT J. COTE,[1] SUSAN M. BERGET,[2] AND ROBERT F. GAGEL[1]*

Section of Endocrinology, University of Texas M. D. Anderson Cancer Center,[1] and Verna and Marrs McClean Department of Biochemistry, Baylor College of Medicine,[2] Houston, Texas 77030

Received 16 June 1995/Returned for modification 11 August 1995/Accepted 20 September 1995

Regulation of calcitonin (CT)/calcitonin gene-related peptide (CGRP) RNA processing involves the use of alternative 3′ terminal exons. In most tissues and cell lines, the CT terminal exon is recognized. In an attempt to define regulatory sequences involved in the utilization of the CT-specific terminal exon, we performed deletion and mutation analyses of a mini-gene construct that contains the CT terminal exon and mimics the CT processing choice in vivo. These studies identified a 127-nucleotide intron enhancer located approximately 150 nucleotides downstream of the CT exon poly(A) cleavage site that is required for recognition of the exon. The enhancer contains an essential and conserved 5′ splice site sequence. Mutation of the splice site resulted in diminished utilization of the CT-specific terminal exon and increased skipping of the CT exon in both the mini-gene and in the natural CT/CGRP gene. Other components of the intron enhancer modified utilization of the CT-specific terminal exon and were necessary to prevent utilization of the 5′ splice site within the intron enhancer as an actual splice site directing cryptic splicing. Conservation of the intron enhancer in three mammalian species suggests an important role for this intron element in the regulation of CT/CGRP processing and an expanded role for intronic 5′ splice site sequences in the regulation of RNA processing.

RNA processing of the calcitonin/calcitonin gene-related peptide (CT/CGRP) gene pre-mRNA is regulated in a tissue-specific manner (4, 23). The key regulatory event in this processing choice is the inclusion or exclusion of an alternative 3′-terminal exon, exon 4. Inclusion of exon 4 results in the utilization of the poly(A) signal within exon 4; exclusion of exon 4 results in the utilization of the poly(A) signal within exon 6 (see Fig. 1A) (4, 23). Exon 4 inclusion occurs in most tissues and cell types in transgenic mice expressing the CT/CGRP gene (11). Skipping of exon 4 is restricted to only a few cell types. This pattern of processing suggests that the processing factors responsible for inclusion of exon 4 are widely distributed and that the sequences to which they bind are sequences commonly associated with pre-mRNAs.

The regulatory mechanisms involved in the alternative processing of CT/CGRP are poorly understood, but several features have been identified. The CT exon 4 possesses a weak 3′ splice site with a noncanonical branch point, and there is general agreement that this feature is important in the regulatory process (2, 3). Substitution of the noncanonical branch point nucleotide with the consensus branch point results in the constitutive inclusion of exon 4 and the use of the exon 4 poly(A) signal in all cell types (3, 5, 9, 26). In addition, three exon 4 elements which are of importance for exon 4 inclusion have been identified in the human gene (10, 24). These sequences, however, are poorly conserved, and studies have failed to demonstrate a role for these sequences in the regulation of the rat CT/CGRP gene.

This lack of consensus led us to reexamine the minimal sequence requirements for human CT exon 4 inclusion. These studies differ from previous studies from this laboratory and others in that we extended the search for regulatory sequences to include intronic sequences. A mini-gene construct in which a fragment containing CT exon 4 with 53 nucleotides of the upstream intron sequence and 571 nucleotides of the downstream intron sequence was developed and inserted into the first intron of the human metallothionein 2A (MT2A) (16) gene (Fig. 1B). CT exon 4 is included in the final mRNA, and the exon 4 poly(A) site is used when this construct is expressed in HeLa cells, a cell line which has historically been used to model recognition of exon 4 (1–3, 5, 9, 10, 13, 17, 22, 26) and which is useful because of the ability to extend in vivo observations to in vitro analysis. CT exon 4 is excluded from the final mRNA when the construct is expressed in T98G cells, a cell line derived from transformed glial tissue and used to model neuronal processing of CT/CGRP in which exon 4 is skipped (17).

Deletion and mutation analyses of this construct led to the identification of a previously unidentified intron regulatory element beginning 150 nucleotides downstream of the CT exon 4 poly(A) signal and containing a putative 5′ splice site. Extensive mutagenesis of this 5′ splice site-containing regulatory sequence provides evidence for the importance of this sequence in the recognition and inclusion of CT exon 4 and hints at the existence of a new class of intronic regulatory elements containing splice sites.

MATERIALS AND METHODS

Cell culture. HeLa cells were maintained on 100-mm-diameter plates in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum as previously described (17).

Plasmids. (i) Deletion mutants. pRSV/MT/CT4.1 contains 1,122 nucleotides of the human CT gene sequence, including exon 4 and its flanking intron sequences, inserted into the first intron of the human MT2A gene; transcription of this chimeric gene is driven by the Rous sarcoma virus promoter (see Fig. 1B) (17). A series of deletion and substitution constructs were created, and they are diagrammed in the figures describing the experiments in which they were employed. Deletion 1 (Del.1) of pRSV/MT/CT4.1 was created by an *Nhe*I-*Nco*I deletion (see Fig. 2). Deletion 2 (Del.2) of pRSV/MT/CT4.1 contains an *Nhe*I-to-*Eco*NI deletion which results in a 150-nucleotide deletion of the intron sequences downstream of CT exon 4 (see Fig. 3). Deletion 3 (Del.3) of pRSV/

* Corresponding author. Mailing address: Section of Endocrinology, Box 15, Department of Medical Specialties, The University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Blvd., Houston, TX 77030. Phone: (713) 792-2840. Fax: (713) 794-4065. Electronic mail address: bob_gagel@isqm.mda.uth.tmc.edu.

MT/CT4.1 was created by an *Eco*NI deletion which results in a 35-nucleotide deletion of the intron sequences downstream of CT exon 4 (see Fig. 3).

(ii) Substitution mutants. Substitution 1 (Sub.1) of pRSV/MT/CT4.1 was created by insertion of a blunt-ended *Spe*I-*Sac*I fragment from intron 3 of the human CT gene downstream into Del.1 (see Fig. 2). Substitution 2 (Sub.2) of pRSV/MT/CT4.1 was created by insertion of a blunt-ended *Hin*cII-*Eco*RI fragment from intron 2 of the hamster adenosine phosphoribosyltransferase (APRT) gene (12) into Del.1 (see Fig. 2).

The remaining insertion constructs (see Fig. 7 and 8) were created by insertion of one of six fragments into one of three restriction sites (*Nhe*I, *Sph*I, or *Avr*II) in the APRT intron 2 sequence of the Sub.2 construct. These fragments (designated A to F; see Fig. 7 and 8) were created by PCR amplification of the CT intron 4 sequence with oligonucleotides complementary to the sequences outside the inserts, which was followed by restriction digestion to generate the appropriately sized insert. The fragments contain portions of a DNA sequence (see Fig. 4 and 8). The sequences for each fragment are as follows (the numbers denote fragment size and location, with number 1 designating the first nucleotide of intron 4): A, 85 to 273; B, 146 to 273; C, 171 to 273; D, 171 to 238; E, 171 to 217; and F, 203 to 238. Insertion of a 5′ splice site consensus sequence, known as the G fragment, was accomplished by mutating sequence CAG/GACTGTA located 19 nucleotides downstream of the *Nhe*I site in Sub.2 to CAG/GTA AGTA.

Mutations were introduced into two additional constructs, pCTG6 (10) and pCT9 (8), to analyze the effect of CT intron 4 mutations in the natural gene (see Fig. 6). The pCTG6 construct contains the first exon from the adenovirus major late gene fused to the human CT gene sequence from a point 153 nucleotides upstream of exon 4 continuing to a point 1.5 kb downstream of exon 6. The pCT9 construct contains the *Sac*I-*Eco*RI enhancer fragment of the Rous sarcoma virus promoter oriented in an antisense orientation and fused to the blunt-ended *Xba*I site of a full-length human CT genomic clone. Point mutations within intron 4 (M1 to M5) (see Fig. 5) were introduced by PCR-directed mutagenesis. All constructs were sequenced to confirm the presence of point mutations and to exclude the introduction of unintended mutation by PCR amplification.

Transfections. Plasmids were introduced into cell lines by DEAE-dextran transfection as previously described (17). Total RNA was isolated 48 h after transfection by RNAzol B extraction (Cinna/Biotecx Laboratory, The Woodlands, Tex.). The results shown are representative of at least three transfections for each plasmid.

RT-PCR. Splicing phenotypes were assayed by reverse transcription PCR (RT-PCR). The low number of PCR cycles (15 to 17) used was determined to fairly represent the relative abundances of individual RNA species. To monitor the use of exon 4 as a 3′-terminal exon, a 3′ primer that is complementary to sequences in the last half of exon 4 and downstream of a cryptic 5′ splice site within exon 4 was used (see Fig. 2B). Likewise, to monitor the skipping of exon 4 and use of MT2A exon 3 as a 3′-terminal exon, a 3′ primer complementary to the sequence in MT2A exon 3 was used (see Fig. 2B). For all RT-PCRs (except for those in which only the CT exon 4 primer was used; see Fig. 2C), both 3′ primers were used so that the ratio of exon 4 inclusion to exon 4 skipping could be examined.

DNase-treated total RNA (5 μg) was used for each RT-PCR amplification. Primer annealing was performed in a 10-μl reaction volume containing 1× PCR buffer (20 mM Tris-HCl, pH 8.4; 50 mM KCl; and 1.5 mM $MgCl_2$) with one or two downstream primers (5 mM each) for 30 min at 55°C. Reverse transcription was performed by adjusting the annealing reaction mixture to 30 μl containing 1× PCR buffer, 10 mM dithiothreitol, 0.4 mM deoxynucleoside triphosphate (dNTP), 10 U of RNasin (Promega Corporation, Madison, Wis.), and 200 U of Moloney murine leukemia virus reverse transcriptase (Gibco/BRL, Gaithersburg, Md.) and incubating it at 37°C for 1 h. Finally, the reverse transcription mixture was adjusted to 100 μl containing 1× PCR buffer, 1 mg of RNase A, 0.4 mM dNTP, one or two downstream primers and the ^{32}P-labeled upstream primer (1 mM each), and 2.5 U of *Taq* DNA polymerase (Perkin-Elmer, Norwalk, Conn.). Fifteen to 17 cycles of PCR amplification were performed, with denaturation at 94°C for 1 min annealing at 55°C for 2 min, extension at 72°C for 2 min, and a final extension at 72°C for 7 min. Studies demonstrated that 15 to 17 cycles of PCR were in the linear range of amplification (data not shown). PCR products were analyzed by polyacrylamide gel electrophoresis and visualized by autoradiography. The PCR products obtained with the MT/CT model construct were cloned into a pCR II vector (Invitrogen, San Diego, Calif.) and sequenced to confirm the identity of each band. The following primers were used for PCR amplification: DS8, 5′-TTGACCATTCACCACATTGGTGTGC-3′; 4697, 5′-GAGTTTAGTTGGCATTCTGG-3′; HMT3, 5′-ATCTGGGAGCGGGGCT GT-3′; GN, 5′-CTGCTCAGGCTTGAAGGTCC-3′; and CT3, 5′-CTGCACTG GTGCAGGACTATGTGC-3′.

Sequence analysis of rat and mouse CT/CGRP genes. Rat and mouse genomic DNAs were isolated from the PC12 cell line and the C57B mouse strain, respectively, with a Genomix DNA extraction kit (Washington Biotechnology, Inc., Bethesda, Md.). Intron 4 sequences from the rat and mouse CT/CGRP genes were PCR amplified with species-specific primers complementary to exons 4 and 5, cloned into the pCR II vector, and sequenced.

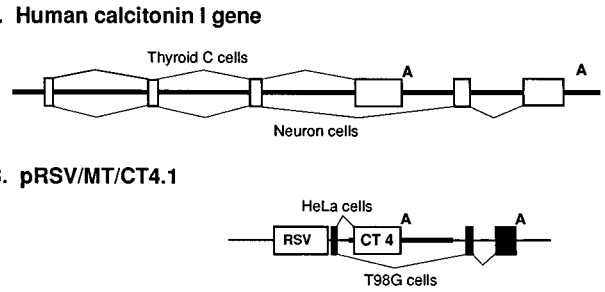

FIG. 1. Schematic representation of alternative splicings of the human CT/CGRP (A) and pRSV/MT/CT4.1 (B) pre-mRNAs. Sequences representing human CT/CGRP exons are shown as open rectangles; sequences derived from human MT2A exons are shown as solid rectangles. Sequences derived from the human CT/CGRP intron 4 are shown as thick lines; sequences derived from the MT2A intron are shown as thin lines. pRSV/MT/CT4.1 was constructed by insertion of the entire CT exon 4 (498 nucleotides), 53 nucleotides of the upstream intron, and 371 nucleotides of the downstream intron into intron 1 of the human MT2A gene. RSV, Rous sarcoma virus promoter; A, poly(A) site.

RESULTS

A role for CT intron 4 sequences in the recognition of CT exon 4. To examine the potential role of the intron 4 sequence in the regulation of CT/CGRP processing, we utilized a previously described mini-gene construct in which CT exon 4 with 53 nucleotides of the 5′ flanking intronic sequence and 571 nucleotides of the 3′ flanking intronic sequence was inserted into the first intron of the three-exon human MT2A gene (17). The resulting construct (Fig. 1B) has four exons and an organization similar to that of the natural CT/CGRP gene in that CT exon 4 with its poly(A) site is followed by an internal exon and a terminal exon with a poly(A) site. When CT exon 4 is recognized, exon 1 is spliced to this exon, resulting in the use of the CT exon 4 poly(A) site, which thereby makes CT exon 4 the terminal exon in this product RNA. When CT exon 4 is skipped, the final MT exon is recognized and polyadenylated, which thereby makes it the terminal exon.

Prior studies with this construct demonstrated a cell-specific processing of transcripts analogous to the processing of the natural CT/CGRP transcript: exon 4 is included in HeLa cells and excluded in T98G cells, a human glioblastoma cell line (17). Other mini-genes containing more of the natural gene also show this processing pattern in HeLa or T98G cells (see Fig. 6). The inclusion pattern seen with this construct in HeLa cells is reflective of the pattern of processing observed in most cultured cell lines for CT/CGRP genes. Only a few cell types direct skipping of exon 4 and thereby produce a processing pattern similar to that occurring naturally in neuronal cells.

An example of the RNA processing analysis of the mini-gene construct expressed in HeLa cells is shown in Fig. 2. Product RNAs were detected by RT-PCR amplification. Figure 2B includes a diagrammatic representation of the RNA products produced from this mini-gene following the transfection of HeLa cells. Use of a primer for reverse transcription that is derived from exon 4 sequences indicated efficient production of mRNA, including CT exon 4 (Fig. 2C, right panel). Because product RNAs containing exon 4 do not contain downstream exons from the mini-gene and vice versa, simultaneous detection of exon 4 inclusion and skipping requires the use of amplification reactions containing two different primers for reverse transcription (utilized primers are diagrammed in Fig. 2B). A complicated processing pattern was observed when a 3′ primer complementary to the downstream MT2A exon 3 in addition to the exon 4 primer was included in the reaction

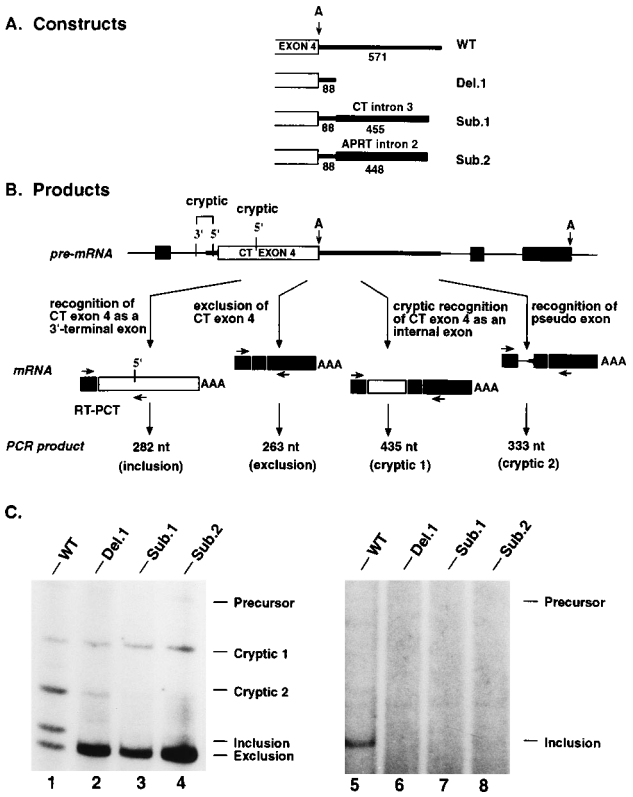

FIG. 2. Effect of deletions and substitutions within intron 4 on CT exon 4 inclusion. (A) Depiction of deletion and substitution constructs showing the nature of the intron 4 sequences. WT, wild type (pRSV/MT/CT4.1). (B) pRSV/MT/CT4.1 pre-mRNA is schematically depicted. Shown within this section are all identified RNA products derived from the expression of this construct in HeLa cells, as determined by the sequencing of RT-PCR amplification products (see Materials and Methods). The relative positions of PCR primers and product sizes are shown. nt, nucleotide; A, poly(A) site; AAA, poly(A) tail. (C) RT-PCR of total cell RNA was performed as described in Materials and Methods. An upstream primer specific for MT2A exon 1 and two downstream primers specific for CT exon 4 and MT2A exon 3 were used in the reactions shown in lanes 1 to 4. The same upstream primer and a single downstream CT exon 4 primer were used in lanes 5 to 8. The identities of the RT-PCR products are indicated on the right side of each panel.

mixture (Fig. 2C, lane 1). Characterization of each of these products by cloning and sequencing identified multiple processing pathways for the mini-gene. Three major and one minor processing products were observed for the wild-type construct resulting from CT exon 4 inclusion, exon 4 exclusion or skipping, and the utilization of two cryptic splice sites (Fig. 2C).

In addition to that resulting from exon 4 inclusion, a major product resulted from the skipping of exon 4. In fact, inclusion and skipping of exon 4 were routinely observed at equal frequencies with this mini-gene construct (Fig. 2C, lane 1). Partial skipping of differential exons when transferred to a heterologous gene is commonly observed. This partial inclusion phenotype provides an ideal background in which to assess the requirement of individual sequence elements for both inclusion and skipping. Two cryptic splicing events also occurred. One minor cryptic RNA (designated cryptic 1 in Fig. 2C) resulted from the recognition of exon 4 as an internal exon by the activation of a 5′ splice site within exon 4 (Fig. 2B). RNAs produced with this cryptic site within exon 4 are seen at a low abundance in transfections with the wild-type CT/CGRP gene. A second cryptic RNA (designated cryptic 2 in Fig. 2C) resulted from the inclusion of a pseudoexon located upstream of exon 4. This pseudoexon was created by the fusion of CT intron and MT intron sequences (Fig. 2B). Production of the first cryptic RNA did not respond to the enhancer characterized in this study, is not known to be relevant to the natural regulation of CT/CGRP RNA processing, and is not considered further in this report. Production of the second cryptic RNA did respond to the enhancer characterized in this study (Fig. 2C, left panel). Because production of this cryptic RNA may reflect the ability of the documented enhancer to activate the inclusion of internal exons as well as 3′-terminal exons, this RNA is discussed in this report. Therefore, in the following discussion, we distinguish between the enhancement of all exon inclusion events (including the production of cryptic 2) and the enhancement of the utilization of exon 4 as a 3′-terminal exon.

As a first step in the analysis of intron 4 sequences required for the inclusion of exon 4, three constructs were created. One construct (Del.1) deleted most of the intron (Fig. 2). In the other two constructs (Sub.1 and Sub.2), the deleted sequence was replaced by one of two heterologous intron sequences (Fig. 2). Deletion of nucleotides 89 to 571 in intron 4 (Del.1) (Fig. 2A) eliminated utilization of CT exon 4 as a 3′-terminal exon (Fig. 2C, lanes 2 and 6) and depressed recognition of the pseudoexon (cryptic 2) (lane 2). The major product produced in cells transfected with the deletion was an RNA (exclusion) resulting from the skipping of all exon 4 sequences (Fig. 2C, lane 2). Replacement of the deleted sequence with two different vertebrate intron sequences, one (Sub.1) from human CT intron 3 and the other (Sub.2) from hamster APRT intron 2 (Fig. 2A), did not restore CT terminal exon inclusion (Fig. 2C, lanes 3, 4, 7, and 8). Therefore, a sequence(s) located in intron 4 enhanced the inclusion of exon 4 as a terminal exon. In addition, the same sequence also enhanced the inclusion of the pseudoexon (Fig. 2C; compare lane 1 with lanes 3 and 4), suggesting that the intron sequence is capable of stimulating the recognition of both terminal and internal exons.

The same constructs were used to transfect T98G cells to see if the deleted or replaced sequences were necessary for the skipping of exon 4 in neuronal cells. All constructs demonstrated skipping of exon 4 and no activation of cryptic splicing (data not shown). Exon skipping is the result expected for transfection of T98G cells if the deleted element is required for exon inclusion. The observation of skipping does not rule out, however, the possibility that the deleted intron element is relevant to the exon exclusion phenotype of neuronal cells, because neuronal exon skipping could result from the absence or lowered concentration of factors that recognize the deleted intron sequences or from the production of factors that recognize sequences within the intron element and negate its enhancement properties (see Discussion).

A conserved 35-nucleotide intron element is required for CT exon inclusion. To define the region within intron 4 necessary for the inclusion of the CT exon 4 as a terminal exon, internal deletions of 150 (Del.2) or 35 (Del.3) nucleotides within intron 4 were introduced into pRSV/MT/CT4.1 (Fig. 3). Expression of these constructs in HeLa cells resulted in the loss of both exon 4 and cryptic exon inclusion (Fig. 3). These results suggested that the sequences within the deleted 35 nucleotides were responsible for the observed exon inclusion events.

To search for possible consensus sequences within the deleted sequences, we compared the mouse, human, and rat CT/CGRP intron 4 sequences. These genes have identical exon and intron architectures and differential splicing patterns. The intron 4 sequences are 63% identical in a 276-nucleotide region extending 3′ from the AATAAA hexanucleotide of exon

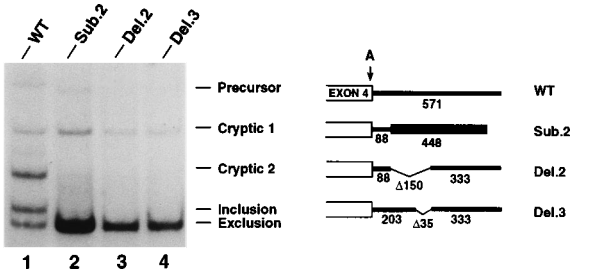

FIG. 3. Effect of small CT intron 4 deletions on CT exon 4 inclusion. The four constructs diagrammed on the right half of the figure were expressed in HeLa cells. The wild-type (WT) and Sub.2 constructs are the same as those shown in Fig. 2. The Del.2 and Del.3 constructs are internal deletions of the wild-type construct as diagrammed. RT-PCR of total cell RNA was performed with an upstream primer specific for MT2A exon 1 and two downstream primers (one primer specific for CT exon 4 and a second primer specific for MT2A exon 3) in the same reaction. The precursor and product bands are identical to those shown in Figure 2 and are indicated on the right side of the left panel. A, poly(A) site.

4 (Fig. 4). The sequences abruptly diverge after that point. The deleted sequences in pRSV/MT/CT4.1 Del.2 and Del.3 lie within this region of homology, suggesting that a functionally important element exists within intron 4 downstream of the poly(A) site.

Intron 4 contains a 5′ splice site sequence necessary for CT exon 4 inclusion. Examination of the 35-nucleotide sequence removed in Del.3 (Fig. 4) identified a 10-nucleotide sequence (box 3) with 100% homology across the three species (Fig. 4) and a high degree of homology to the 5′ splice site consensus sequence, CAG/GTAAGTA. An additional, but less well conserved, 5′ splice site sequence (box 1) is located upstream of the 35-nucleotide deletion (Fig. 4). Two additional sequences

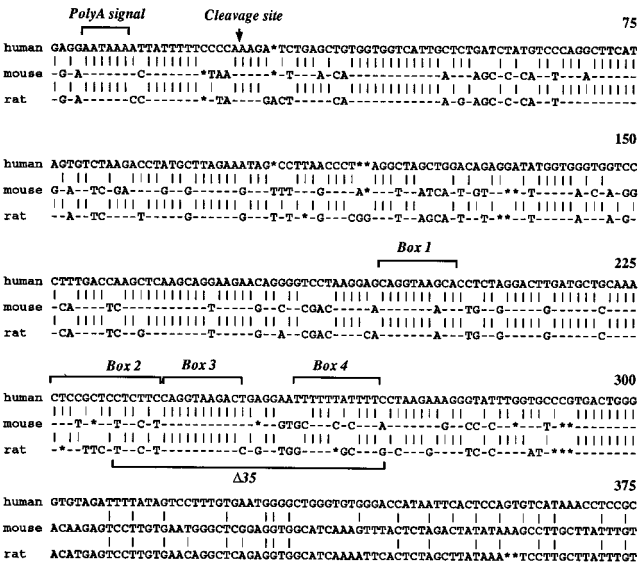

FIG. 4. Comparison of intron 4 sequences from human, mouse, and rat CT/CGRP genes. A comparison of the intronic sequences shows a high degree of homology over a 300-nucleotide region beginning with the poly(A) hexanucleotide (AATAAA). The poly(A) hexanucleotide and cleavage sites are shown. Boxes 1, 2, 3, and 4 (Fig. 5) are delineated. The Δ35 nucleotide deletion (Del.3) (Fig. 3) is shown. Note the abrupt diversity of the sequences after 300 nucleotides. Dashes, nucleotides that are the same as those in the human sequence; asterisks, nucleotides missing compared with the human sequence; vertical bars, identical nucleotides among three species (thick bars) or between either human and mouse or human and rat sequences (thin bars).

(boxes 2 and 4) flanking the 5′ splice site in box 3 were highlighted because both are pyrimidine rich (Fig. 4), with the sequence in box 4 consisting almost entirely of thymidines.

The identification of a perfect 5′ splice site sequence in this region led us to focus our analysis by performing point mutations of pRSV/MT/CT4.1. A 3-base mutation of the 5′ splice site CAG/GTAAGAC to CAG/CATAGAC in box 3 (Fig. 5, lane 3) and a second single-base mutation at position +1 of the 5′ splice site from CAG/GTAAGAC to CAG/CTAAGAC (Fig. 5, lane 6), mutations which normally inactivate a 5′ splice site (20, 27), resulted in both exon 4 and pseudoexon exclusion. These results not only provide compelling evidence for the importance of this intron sequence but point to a specific role for an intronic 5′ splice site sequence in controlling exon 4 recognition and inclusion.

Mutations were introduced into other regions of the intron sequence and revealed complexity in the properties of the enhancer in that recognition of exon 4 as a 3′-terminal exon and recognition of the cryptic internal exon required different sequences within the element in addition to their mutual requirement for the 5′ splice site sequence in box 3. Such complexity was revealed by mutation of box 1, 2, or 4 (Fig. 5). Mutation of a sequence within the pyrimidine-rich regions of box 2 or 4 reduced the relative proportions of exon 4 inclusion (Fig. 5, lanes 4 and 5) but had minimal impact on the inclusion of the cryptic exon (Fig. 5, lanes 4 and 5). These results point to a modifying role of the pyrimidine tracts and suggest that their presence is important in the recognition and polyadenylation of exon 4 as a terminal exon. The nucleotide sequence in box 1 resembles a 5′ splice site (CAG/GTAAGCA). A 3-nucleotide mutation of box 1 to CAG/CATAGCA reproducibly increased the proportion of exon 4 inclusion relative to exon 4 exclusion (Fig. 5, lane 2) and eliminated cryptic splicing. These results suggest that although the intron element enhances exon inclusion of both internal and 3′-terminal exons, the mechanisms of these two events are slightly different.

The mutant mini-genes described in Fig. 5 were also transfected into T98G cells to see if any of the mutations affected the ability of neuronal cells to skip exon 4. All constructs demonstrated normal exon skipping phenotypes (data not shown), with no trace of exon 4 inclusion or activation of cryptic splicing. Because the mutation of the 5′ splice site within box 3 causes exon skipping in cell lines that normally include exon 4, the observation of exon skipping in neuronal cells does not rule out the possibility that this 5′ splice site is the target of negatively acting neuronal factors. For those mutations that have partial inclusion phenotypes in HeLa cells, however, these results suggest that the mutated sequences are not the target of neuronal factors that could bind to the intron enhancer and negate its enhancement property.

Introduction of 5′ splice site mutations into the natural human CT/CGRP gene depresses recognition of exon 4. To confirm the relevance of the 5′ splice site in box 3, sequence CAG/GTAAGAC within the box was mutated to CAG/CATAGAC in two additional constructs. The first construct contains a fragment of the natural human CT/CGRP gene stretching from the middle of intron 3 to the 3′ end of the gene fused to a first exon and half of an intron from the major late transcription unit of adenovirus (Fig. 6, lane 1). The second construct contains the entire natural human CT/CGRP gene (Fig. 6, lane 4). The 3-nucleotide mutation of the 5′ splice site in box 3 resulted in a reduction of CT exon 4 inclusion in both pCTG6 (Fig. 6, lane 3) and the natural gene, pCT9 (Fig. 6, lane 6). These results provide additional and independent corroboration of the role of this 5′ splice site in the recognition of CT

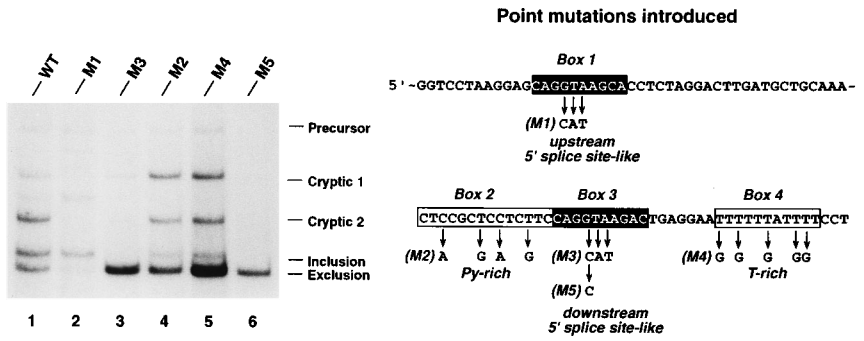

FIG. 5. Mutational analysis of the CT intron 4 element. Point mutations (M1 to M5, right panel) were created in pRSV/MT/CT4.1 as outlined in Materials and Methods. Each construct was transfected individually into HeLa cells; the resultant RNA products were analyzed by RT-PCR of total RNA with an upstream primer for MT2A exon 1 and two downstream primers (a CT exon 4- and an MT2A exon 3-specific primer) in each reaction mixture. RT-PCR products for the wild-type (WT) construct and each mutant construct (M1 to M5) are shown in the left panel. The identity of each band is indicated to the right of the panel.

exon 4. Mutation of the 5′ splice site in box 1 had no effect on RNA splicing in these two constructs (Fig. 6, lanes 2 and 5).

A 127-nucleotide fragment of intron 4 functions in a heterologous context. The previous experiments demonstrated the importance of the conserved region of intron 4 in the enhancement of exon 4 inclusion. To ascertain if all important sequences were located within the region containing boxes 1 to 4, a gain of function experiment was performed. In this experiment, two intron fragments with sizes of 127 and 188 nucleotides were inserted into the Sub.2 construct in which the CT intron sequences had been replaced with intron sequences from APRT (Fig. 2 and 7). The inserts, A and B, were introduced at one of three intronic locations designated I, II, and III (Fig. 7). Insertion of either the 188-nucleotide (insert A) (Fig. 7, lane 3) or 127-nucleotide (insert B) (Fig. 7, lane 5) fragment into position I, a location roughly equivalent to its natural position in CT intron 4, resulted in enhanced exon 4 inclusion. Insertion of these same sequences in an antisense direction had no effect (Fig. 7, lanes 4 and 6). Placement of insert B farther downstream at position II (Fig. 7, lane 7) or III (Fig. 7,

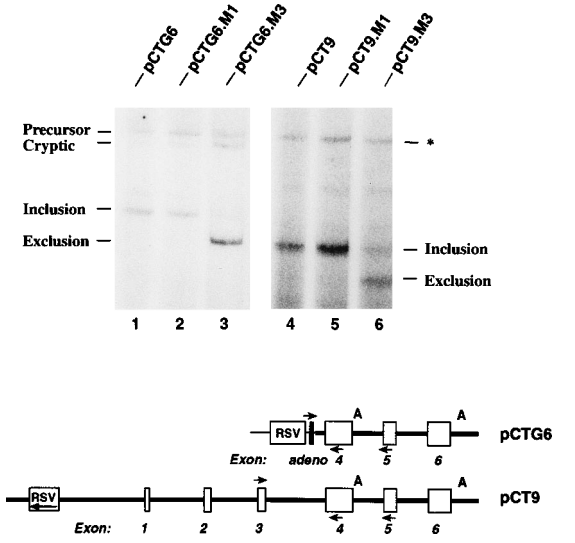

FIG. 6. Mutational analysis of the CT intron enhancer 5′ splice site sequences in the natural human CT/CGRP gene. The M1 and M3 mutations of the 5′ splice site shown in Fig. 5 were inserted into two constructs as described in Materials and Methods. pCTG6 fuses the terminal three exons of the CT/CGRP gene to adenovirus exon 1, and pCT9 is the full-length natural human CT/CGRP gene fused to the enhancer region of the Rous sarcoma virus (RSV) promoter in a reverse orientation to enhance the level of expression (diagrams). The primers used in the RT-PCRs are shown above or below each construct. Each RT-PCR contained a single upstream primer and two downstream primers. Lanes 1 to 3 show the RT-PCR products (designated at the left of the panel) for pCTG6 (lane 1) and pCTG6 with mutations M1 (lane 2) and M3 (lane 3) transfected into HeLa cells (the M1 and M3 mutations are shown in Fig. 4). The precursor band is 481 nucleotides, the cryptic band is 452 nucleotides, the inclusion band is 319 nucleotides, and the exclusion band is 280 nucleotides. Lanes 4 to 6 show the RT-PCR products (designated at the right of the panel) for pCT9 (lane 4) and pCT9 with mutations M1 (lane 5) and M3 (lane 6). The inclusion band is 276 nucleotides, and the exclusion band is 237 nucleotides. The PCR product indicated by the asterisk represents a nonspecific product. A, poly(A) site.

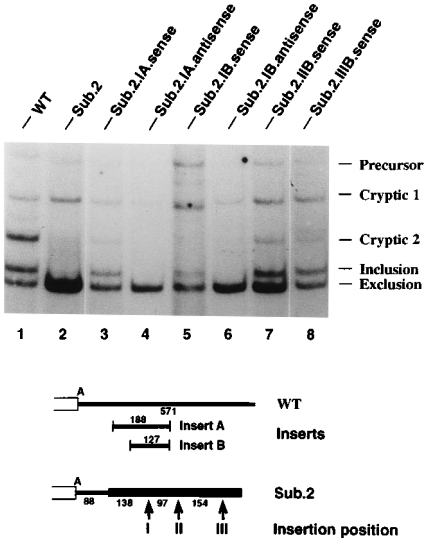

FIG. 7. The CT intron element can function in a heterologous sequence and in a position-independent manner. Inserts A and B were constructed as described in Materials and Methods and inserted into positions I (227 nucleotides downstream of CT exon 4), II (323 nucleotides downstream of CT exon 4), and III (477 nucleotides downstream of CT exon 4) of APRT intron 2 as shown in the bottom portion of this figure. Each construct (as designated above the lanes in the upper panel) was transfected into HeLa cells and analyzed by RT-PCR with a single upstream primer complementary to MT2A exon 1 and two downstream primers (one complementary to the sequence in CT exon 4 and the other to MT2A exon 3). The RT-PCR products are designated at the right of the upper panel. Abbreviations: WT, wild type (pRSV/MT/CT4.1); Sub.2IA.sense, Sub.2 with insert A at position I in a sense orientation; Sub.2IA.antisense, Sub.2 with insert A at position I in an antisense orientation; Sub.2IB.sense, Sub.2 with insert B at position I in a sense orientation; Sub.2IB.antisense, Sub.2 with insert B at position I in an antisense orientation; Sub.2.IIB.sense, Sub.2 with insert B at position II in a sense orientation; Sub.2.IIIB.sense, Sub.2 with insert B at position III in a sense orientation. A, poly(A) site.

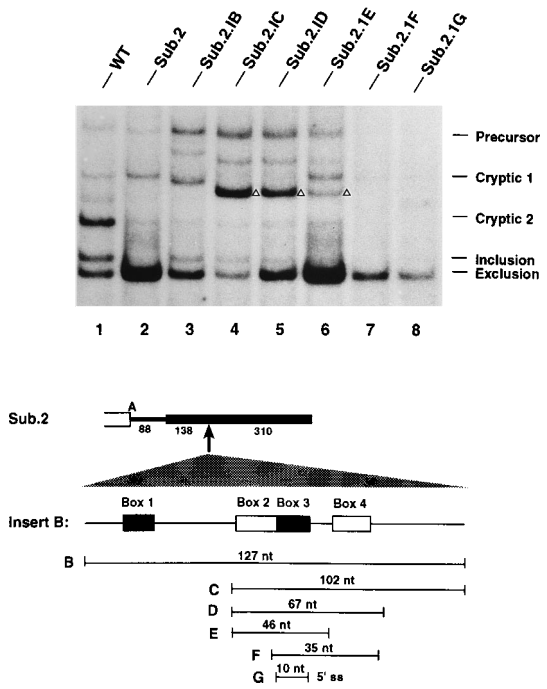

FIG. 8. Deletion analysis of a 127-nucleotide fragment containing the intron element. Fragments of a 127-nucleotide sequence containing boxes 1 to 4 and flanking sequences (insert B) were inserted into construct Sub.2 at position I (shown schematically). Each construct (as designated above the lanes in the upper panel) was transfected into HeLa cells and analyzed by RT-PCR with a single upstream primer complementary to MT2A exon 1 and two downstream primers (one complementary to the sequence in CT exon 4 and the other to MT2A exon 3). The upper panel shows the RT-PCR products for each of these constructs. Specific products are designated at the right of the upper panel. The bands indicated by the triangles represent a new cryptic splice product with a size of 401 nucleotides, as discussed in the text. Abbreviations: nt, nucleotide; WT, wild type (pRSV/MT/CT4.1); Sub.2.IB, Sub.2 with insert B at position I; Sub.2.IC, Sub.2 with insert C at position I; Sub.2.ID, Sub.2 with insert D at position I; Sub.2.1E, Sub.2 with insert E at position I; Sub.2.1F, Sub.2 with insert F at position I; Sub.2.1G, Sub.2 with insert G at position I; A, poly(A) site; 5′ ss, 5′ splice site sequence. Position I is defined in the legend to Fig. 7.

lane 8) had identical effects, indicating the position-independent function of this intron element. The ability of the intron element to function independently of position in a heterologous sequence context led us to term the element an intron enhancer.

Evidence that all four conserved intron 4 domains (boxes 1 to 4) are required for enhancement of CT exon 4 inclusion. To further analyze the individual contributions of the motifs within the 127-nucleotide intron enhancer, a series of smaller fragments were inserted into position I of the Sub.2 construct (described in Materials and Methods). The resulting constructs (Sub.2 with inserts B to G) and analysis of the RNA products produced after transfection of HeLa cells with the mutants are shown in Fig. 8. Exon 4 inclusion was enhanced relative to that with Sub.2 (Fig. 8, lane 2) by all constructs that contained boxes 2, 3, and 4 (fragments B, C, and D) (lanes 3 to 5). Exclusion of boxes 1 and 4 (Fig. 8, lane 6) or boxes 1 and 2 (lane 7) resulted in exon 4 exclusion. Finally, insertion of a 10-nucleotide sequence (CAG/GTAAGTA [fragment G]) identical to a minimal 5′ splice site consensus sequence (20) at position I did not cause exon 4 inclusion (Fig. 8, lane 8). These results suggest the presence of a complex element in which the 5′ splice site sequence in box 3 is necessary but not sufficient for optimal enhancer function.

Expression of the constructs containing inserts C, D, and E (Fig. 8, lanes 4 to 6) also resulted in a new RNA species resulting from cryptic splicing (marked with triangles in Fig. 8). Sequence analysis of this band demonstrated that it was a spliced RNA in which the 5′ splice site in box 3 of the enhancer was utilized to create a pseudoexon by activation of a cryptic 3′ splice site within the APRT intron sequence. Recognition of the 5′ splice site sequence within box 3 as a valid 5′ splice site when other elements of the intron enhancer have been deleted suggests both that the sequence can be recognized as a 5′ splice site and that other elements within the enhancer domain prevent the use of this sequence as an actual 5′ splice site in the natural gene.

DISCUSSION

The regulatory mechanisms controlling alternative RNA processing of the CT/CGRP gene are poorly understood. Although several different sequence elements in both the rat and human CT/CGRP genes that seem to influence the observed processing pattern have been identified (1, 3, 5, 7, 8, 10, 13, 22, 24, 26), no single important conserved regulatory sequence or model for regulation has yet emerged. This problem led us to consider the possibility of additional regulatory sequences that had not yet been identified. Here, we report the identification of a 35-nucleotide sequence located within the intron downstream of the alternative exon that functions as an enhancer for exon inclusion. This element is highly conserved between humans and rodents, and important nucleotides defined by mutation in this study are identical in mouse, rat, and human genes.

The intron enhancer was determined to be an essential RNA element for exon 4 recognition by several experimental approaches. Perhaps most importantly, alteration of three nucleotides within the enhancer region severely depressed exon 4 recognition when this mutation was introduced into the natural CT/CGRP gene. The mutation so introduced altered a sequence within the enhancer that is identical to a consensus 5′ splice site. Given the ability of exon 4 to be recognized in most cell types, it is perhaps not surprising that the sequence element identified as important within the enhancer is a common one that should be recognizable by multiple processing factors within a wide variety of cell types.

The CT/CGRP intron enhancer has several properties that support its definition as a processing enhancer. First, it is located some distance downstream of its target exon (153 nucleotides) and can function at even greater distances (477 nucleotides). At these distances, the element cannot be considered as part of the normal poly(A) site terminating exon 4. Exhibiting another property befitting the term enhancer, the element can make exon inclusion possible when it is transferred into heterologous intron sequences artificially introduced into the CT/CGRP gene. Even more interestingly, in preliminary experiments we have seen that the CT/CGRP enhancer can facilitate exon inclusion in a totally heterologous gene (16a). Some evidence for strong enhancement of heterologous processing events by the enhancer was obtained in this study as well. In addition to enhancement of normal exon 4 utilization, the enhancer also stimulated inclusion of an upstream pseudoexon created entirely out of intron sequences. Inclusion of the pseudoexon was dependent on the 5′ splice site within the enhancer, supporting the observation that the 5′ splice site within the enhancer is an important part of the enhancer.

The intron enhancer has several properties that distinguish the element from other known RNA processing enhancers.

First, the enhancer is the first intron element associated with an alternative 3′-terminal exon. Secondly, to our knowledge, it is the first described intronic enhancer that contains splice site-like sequences. 5′ splice sites are thought to be the targets for two known splicing regulators, ASF/SF2 and heterogeneous nuclear ribonucleoprotein A1 (19, 28), although to date, these proteins are known to regulate only 5′ splice sites that are actually used to direct a splicing event. The 5′ splice site within the CT/CGRP enhancer is not used to produce the two major mRNAs made from this gene. Adema (1) and Bovenberg et al. (7), however, have observed that one of the 5′ splice site-like sequences within the enhancer is infrequently utilized as an alternative to the exon 4 poly(A) signal to produce an unusual RNA species with an elongated exon 4. We also observed activation of the enhancer-located 5′ splice site for use as a 5′ splice site of a cryptically spliced internal exon when other elements of the enhancer were deleted. This result suggests that although the enhancer-located 5′ splice site is recognized during enhancer function, this recognition differs from normal splice site recognition during splicing and that the difference is enforced by other sequence elements within the enhancer.

The mechanism by which the enhancer functions to stimulate exon inclusion is unknown. One possibility involves the participation of U1 small nuclear ribonucleoproteins and their ability to bind to 5′ splice sites. U1 small nuclear ribonucleoprotein A has been implicated in both positive (18) and negative (6, 15) regulation of polyadenylation. Similarly, 5′ splice sites placed near poly(A) sites have been implicated in both positive (25) and negative (14, 21) regulation of polyadenylation. Therefore, the binding of U1 small nuclear ribonucleoproteins to the 5′ splice site within the enhancer might be predicted to influence polyadenylation of exon 4. Indeed, in recent experiments, it has been observed that the intron enhancer activates in vitro polyadenylation (16a).

Any model for enhancement must also include an explanation for the ability of the enhancer to stimulate the inclusion of upstream internal exons. A second sequence feature of this intron enhancer is a pyrimidine tract upstream of the important 5′ splice site. Our results indicate that the presence of a 5′ splice site consensus sequence alone is inadequate to cause exon 4 inclusion. However, the presence of a sequence containing only the pseudo-5′ splice site and flanking pyrimidine tract was sufficient for exon 4 inclusion. Our limited mutational analysis of the pyrimidine tract indicates that introduction of purines into the tract lowers exon 4 inclusion but has little impact on the enhancement of the inclusion of the internal pseudoexon, suggesting differences in the mechanisms by which the enhancer stimulates inclusion of internal and terminal exons.

Is it possible to develop a model for CT/CGRP alternative splicing which includes the intron enhancer? The only model for regulation of CT/CGRP processing which has gained any acceptance is that in which splicing signals adjacent to exon 4, including the exon 4 branch point and pyrimidine tract, are weak (3, 5, 9, 26). Therefore, it might be anticipated that CT exon 4 recognition and polyadenylation require strengthening by multiple regulatory sequences and factors. We propose that widely distributed processing factors that bind 5′ splice sites and polypyrimidine tracts bind to the enhancer in most cell types and activate exon 4 recognition. The mechanism of the skipping of exon 4 in restricted types of neuronal cells is less clear. None of our mutations that lowered exon 4 recognition in HeLa cells showed any phenotype in neuronal cells, i.e., all constructs demonstrated the normal process of exon skipping. Although this result is only informative concerning those mutations that retain partial inclusion phenotypes in HeLa cells (the pyrimidine tract mutations), it does suggest that these sequences are not the targets of negatively acting neuronal factors that bind to the enhancer and negate its enhancement. Observation of skipping in neuronal cells with the 5′ splice site mutant is not informative, because a wild-type splice site within the enhancer is absolutely required for exon 4 inclusion in HeLa cells. Therefore, it is possible that the 5′ splice site within the enhancer is the target for negatively acting neuronal factors. Alternatively, it is also possible that there is no negative factor in neuronal cells and that skipping occurs because of the alteration or lowered amounts of a factor required for enhancer recognition within these cells.

ACKNOWLEDGMENTS

We thank Suchen E. Huang for maintaining HeLa and T98G cells and Pam Thomas for the mouse calcitonin genomic sequence. We acknowledge the helpful advice of members of the Berget and Cooper laboratories.

This work was supported by NIH grants R01-DK38146 (R.F.G.) and 2P30-CA16672 (M. D. Anderson Cancer Center) and American Cancer Society award NP-695 (S.M.B.).

REFERENCES

1. **Adema, G. J.** 1991. Ph.D. thesis. Institute of Molecular Biology and Medical Biotechnology, Utrecht, The Netherlands.
2. **Adema, G. J., R. A. L. Bovenberg, H. S. Jansz, and P. D. Baas.** 1988. Unusual branch point selection involved in splicing of the alternatively processed calcitonin/CGRP-I pre-mRNA. Nucleic Acids Res. **16:**9513–9526.
3. **Adema, G. J., R. A. L. van Hulst, and P. D. Baas.** 1990. Uridine branch acceptor is a cis-acting element involved in regulation of the alternative processing of calcitonin/CGRP-I pre-mRNA. Nucleic Acids Res. **18:**5365–5372.
4. **Amara, S. G., V. Jonas, and M. G. Rosenfeld.** 1982. Alternative RNA processing in calcitonin gene expression generates mRNAs encoding different polypeptide products. Nature (London) **298:**240–244.
5. **Benett, M. M., and S. G. Amara.** 1993. Molecular mechanisms of cell-specific and regulated expression of the calcitonin/a-CGRP and β-CGRP genes. Ann. N. Y. Acad. Sci. **257:**36–49.
6. **Boelens, W. C., E. J. Jansen, W. J. van Venrooij, R. Stripecke, I. W. Mattaj, and S. I. Gunderson.** 1993. The human U1 snRNP-specific U1A protein inhibits polyadenylation of its own pre-mRNA. Cell **26:**881–892.
7. **Bovenberg, R. A. L., T. C. Moen, H. S. Jansz, and P. D. Baas.** 1989. In vitro splicing analysis of mini-gene constructs of the alternatively processed human calcitonin/CGRP-I pre-mRNA. Biochim. Biophys. Acta **1008:**223–233.
8. **Cote, G. J., R. V. Abrussese, C. J. Lips, and R. F. Gagel.** 1990. Transfection of calcitonin gene regulatory elements into a cell culture model of the C cell. J. Bone Miner. Res. **5:**165–171.
9. **Cote, G. J., N. Nguyen, C. J. M. Lips, S. M. Berget, and R. F. Gagel.** 1991. Validation of an in vitro RNA processing system for CT/CGRP precursor mRNA. Nucleic Acids Res. **19:**3601–3606.
10. **Cote, G. J., D. T. Stolow, S. Peleg, S. M. Berget, and R. F. Gagel.** 1992. Identification of exon sequences and an exon binding protein involved in alternative RNA splicing of calcitonin/CGRP. Nucleic Acids Res. **20:**2361–2366.
11. **Crenshaw, E. B., A. F. Russo, L. W. Swanson, and M. G. Rosenfeld.** 1987. Neuron-specific alternative RNA processing in transgenic mice expressing a metallothionein-calcitonin fusion gene. Cell **49:**389–398.
12. **de Boer, J. G., E. A. Drobetsky, A. J. Grosovsky, M. Mazur, and B. W. Glickman.** 1989. The Chinese hamster aprt gene as a mutational target. Its sequence and an analysis of direct and inverted repeats. Mutat. Res. **226:**239–244.
13. **Emeson, R. B., F. Hedjran, J. M. Yeakley, J. W. Guise, and M. J. Rosenfeld.** 1989. Alternative production of calcitonin and CGRP mRNA is regulated at the calcitonin-specific splice acceptor. Nature (London) **341:**76–80.
14. **Furth, P. A., W.-T. Choe, J. H. Rex, J. C. Byrne, and C. C. Baker.** 1994. Sequences homologous to 5′ splice sites are required for the inhibitory activity of papillomavirus late 3′ untranslated regions. Mol. Cell. Biol. **14:**5278–5289.
15. **Gunderson, S. I., K. Beyer, G. Martin, W. Keller, W. C. Boelens, and L. W. Mattaj.** 1994. The human U1A snRNP protein regulates polyadenylation via a direct interaction with poly(A) polymerase. Cell **11:**531–541.
16. **Karin, M., and R. I. Richards.** 1982. Human metallothionein genes—primary structure of the metallothionein-II gene and a related processed gene. Nature (London) **299:**797–802.

16a. **Lou, H.** Unpublished data.

17. **Lou, H., G. J. Cote, and R. F. Gagel.** 1994. The calcitonin exon and its flanking intronic sequences are sufficient for the regulation of human calcitonin/calcitonin gene-related peptide alternative splicing. Mol. Endocrinol. **8:**1618–1626.
18. **Lutz, C. S., and J. C. Alwine.** 1994. Direct interaction of the U1 snRNP-A protein with the upstream efficiency element of the SV40 late polyadenylation signal. Genes Dev. **8:**576–586.
19. **Mayeda, A., D. M. Helfman, and A. R. Krainer.** 1993. Modulation of exon skipping and inclusion by heterogeneous nuclear ribonucleoprotein A1 and pre-mRNA splicing factor SF2/ASF. Mol. Cell. Biol. **13:**2993–3001.
20. **Moore, M. J., C. C. Query, and P. A. Sharp.** 1993. Splicing of precursors to mRNA by the spliceosomes, p. 303–358. *In* R. F. Gesteland and J. F. Atkins (ed.), The RNA world. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
21. **Niwa, M., C. C. MacDoald, and S. M. Berget.** 1992. Are vertebrate exons scanned during splice-site selection? Nature (London) **360:**277–280.
22. **Roesser, R. J., K. Littschwager, and S. E. Leff.** 1993. Regulation of tissue-specific splicing of the calcitonin/calcitonin gene-related peptide gene by RNA-binding proteins. J. Biol. Chem. **268:**8366–8375.
23. **Rosenfeld, M. G., S. G. Amara, and R. M. Evans.** 1984. Alternative RNA processing: determining neuronal phenotype. Science **225:**1315–1320.
24. **van Oers, C. C. M., G. J. Adema, H. Zandberg, T. C. Moen, and P. D. Baas.** 1994. Two different sequence elements within exon 4 are necessary for calcitonin-specific splicing of the human calcitonin/calcitonin gene-related peptide I pre-mRNA. Mol. Cell. Biol. **14:**951–960.
25. **Wassarman, K. M., and J. A. Steitz.** 1993. Association with terminal exons in pre-mRNAs: a new role for the U1 snRNP? Genes Dev. **7:**647–659.
26. **Yeakley, J. M., F. Hedjran, J.-P. Morfin, N. Merillat, M. G. Rosenfeld, and R. B. Emeson.** 1993. Control of calcitonin/calcitonin gene-related peptide pre-mRNA processing by constitutive intron and exon elements. Mol. Cell. Biol. **13:**5999–6011.
27. **Zhuang, Y., and A. M. Weiner.** 1986. A compensatory base change in U1 snRNA suppresses a 5′ splice site mutation. Cell **46:**827–835.
28. **Zuo, P., and J. L. Manley.** 1994. The human splicing factor ASF/SF2 can specifically recognize pre-mRNA 5′ splice sites. Proc. Natl. Acad. Sci. USA **91:**3363–3367.