# RNA Secondary Structure: An Important *cis*-Element in Rat Calcitonin/CGRP Pre-Messenger RNA Splicing†

Timothy P. Coleman and James R. Roesser*

*Department of Biochemistry, Virginia Commonwealth University, 1101 E. Marshall St., Richmond, Virginia 23227*

*Received April 10, 1998; Revised Manuscript Received August 17, 1998*

ABSTRACT: The calcitonin/CGRP gene pre-mRNA is alternatively spliced in a tissue-specific manner resulting in the formation of calcitonin mRNA in thyroid C-cells and CGRP mRNA in neurons. Computer analysis of the RNA containing the 3′ splice acceptor of the calcitonin-specific exon 4 predicts that this region has the potential to form a thermodynamically stable stem-loop. Data from CD spectroscopy and solution phase structure probing with single-strand specific and double-strand specific RNases indicates that RNA in this region is substantially double stranded. In vitro splicing of chimeric human $\beta$-globin/calcitonin transcripts in HeLa nuclear extract was inhibited by base changes predicted to destabilize the stem, while compensatory base changes resulted in splicing at 50% of wild-type levels. Changing the residue opposite the AG dinucleotide adenosine in the stem from G to U, allowing the formation of an A−U basepair, abolished usage of this splice acceptor *in vitro*. These results indicate that a thermodynamically stable RNA stem-loop forms in vitro at the 3′ splice acceptor of exon 4 of the calcitonin/CGRP gene transcript. This RNA secondary structure acts as a novel cis-element involved in proper splice site selection.

Alternative pre-mRNA splicing is an important form of genetic regulation necessary to create the genetic diversity found in eukaryotes (*1*, *2*). The selection of splice sites in pre-mRNAs is accomplished through the coordination of several cis- and trans-acting elements (*3*). While many of the elements involved with constitutive splicing have been defined, it has been more difficult to elucidate the factors which control the highly specific inclusion or exclusion of exons during developmental, sexual, and tissue-specific alternative splicing events.

There are several examples of sequences other than those located in the consensus splice signals that affect splice site selection (*4*−*7*). However, the role of RNA secondary structure in splice site selection has been difficult to ascertain. It has been proposed that there is a "window" of opportunity of about 100 nucleotides that exists behind the transcribing polymerase in which the nascent RNA is free to fold (*8*). This would give ample opportunity for thermodynamically stable hairpin loops to form under physiological conditions. It has been shown that the presence of RNA structures, either artificial or natural (*4*, *9*−*12*) can influence splice site selection and splicing efficiency. Sequestration of a splice site in an RNA duplex decreases its utilization both in vitro and in vivo (*12*, *13*). Conversely, RNA structures have been identified which can enhance splice site recognition by reducing the spacing between splice signals (*14*) or by facilitating the presentation of splice sites (*4*).

Blanchette and Chabot (*11*) recently reported an RNA secondary structure that modulates the alternative splicing of the mouse hnRNP A1 gene transcript. The structure appears to be evolutionarily conserved in humans. The 5′ splice donor of the alternatively spliced exon 7B is sequestered in a duplex which forms in vitro and in vivo. The duplex involves nucleotides from exon 7B and intron 8 and promotes skipping of exon 7B by inhibiting assembly of a U1 snRNP-dependent complex. Other pre-mRNAs such as the adenovirus type 2 transcript (*15*) and the rat $\beta$-tropomyosin transcript (*10*, *16*) have been shown to possess RNA secondary structures that influence splice site selection both in vivo and in vitro. Although the data supporting local RNA secondary structure as an influencing factor on splice site selection is limited, it is apparent that some pre-mRNA splice choices are modulated by RNA secondary structure.

The processing of the calcitonin/calcitonin gene-related peptide (CGRP[1]) pre-mRNA was one of the first tissue-specific alternative processing events described in mammals (Figure 1) (*17*). In thyroid C cells, the first three exons of the calcitonin/CGRP gene are spliced to the calcitonin-specific exon 4 and a polyadenylation site at the end of this exon is used to generate a mature calcitonin mRNA. Neurons produce the mature CGRP mRNA by splicing the first three common exons to exons 5 and 6 and utilizing the polyadenylation site located at the end of exon 6. It is generally accepted (*18*−*20*) that the 3′ splice acceptor of exon 4 is a poor match to consensus splice site sequences and that the intrinsic weakness of the splice acceptor is important for calcitonin splice regulation. The nonconserved

† This work was supported by grants from the A. D. Williams Foundation and the Jeffress Foundation.

* Corresponding author. Fax: (804)-828-1473 E-mail: Jroesser@HSC.VCU.EDU.

[1] Abbreviations: CGRP, calcitonin gene related peptide; CIP, calf intestinal alkaline phosphatase; depc, diethyl pyrocarbonate; dNTP, deoxyribonucleotide triphosphate; NTP, ribonucleotide triphosphate; PAGE, polyacrylamide gel electrophoresis; PCR, polymerase chain reaction; RT, reverse transcriptase.

10.1021/bi9808058 CCC: $15.00 
Published on Web 10/17/1998

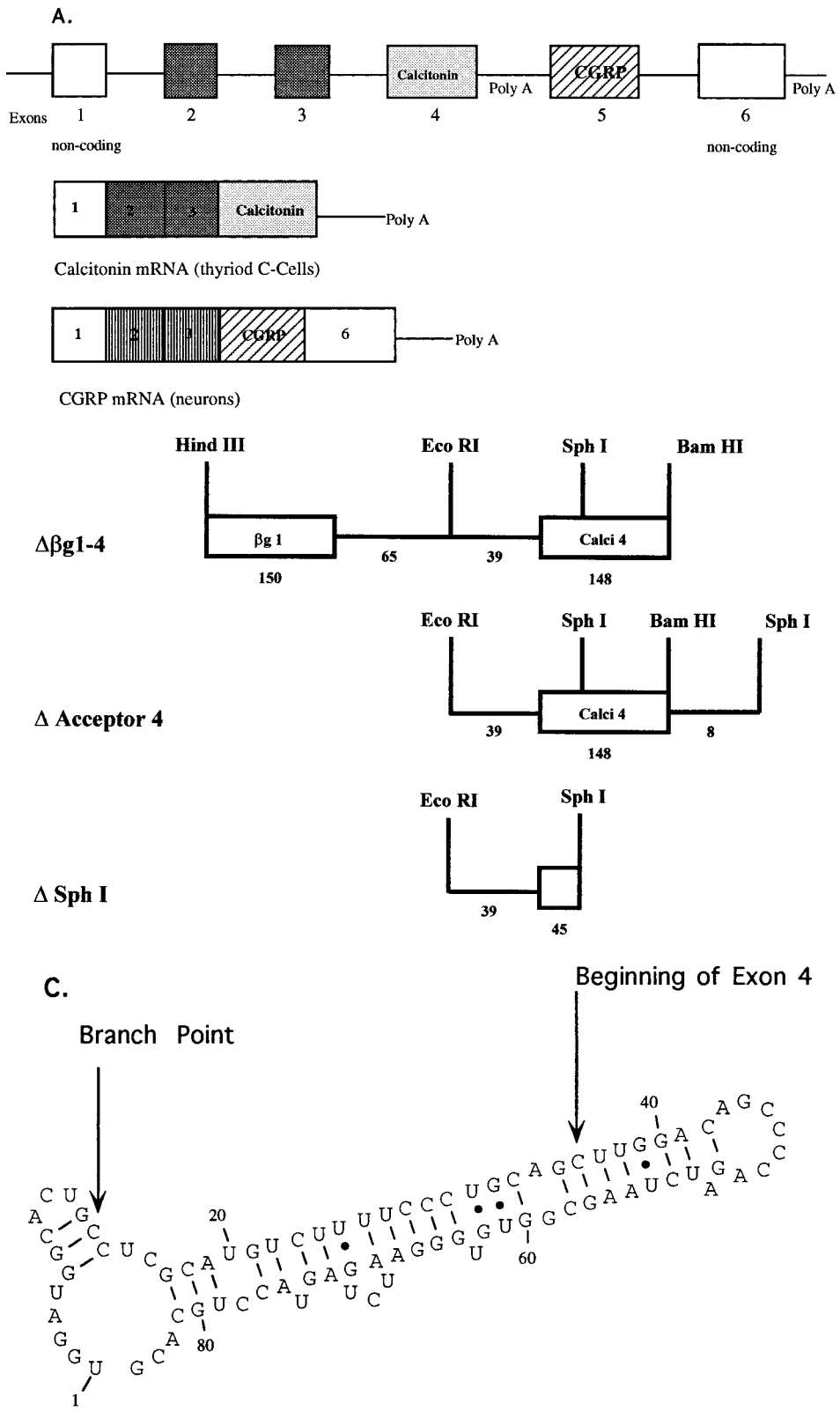



FIGURE 1: The calcitonin/CGRP gene and minigene constructs. (Panel A) The structure of the rat calcitonin/CGRP gene. *Thin lines* and *boxes* represent introns and exons, respectively. *Open boxes* are noncoding exons (boxes 1 and 6), and stippled *boxes* are the coding regions common to both calcitonin and CGRP (boxes 2 and 3). The calcitonin-specific exon (box 4) is labeled *calcitonin*, and the CGRP-specific coding region (box 5) is labeled *CGRP*. The mature mRNAs encoding calcitonin or CGRP are shown below. (Panel B) Minigenes constructed to examine the secondary structure and structure-dependent calcitonin-specific acceptor splice usage in vitro. The $\Delta\beta 1-4$ is a hybrid minigene which consists of the human $\beta$-globin first exon, 65 nucleotides of the first intron fused to the 3′ end of the calcitonin third intron (39 nucleotides), and exon 4 (148 nucleotides). (Panel C) Predicted RNA stem loop structure containing the 3′ splice acceptor at the 5′ end of the calcitonin-specific exon 4.

splice signals surrounding this splice acceptor include a noncanonical branchpoint, that is, a cytosine instead of an adenosine residue, and a polypyrimidine tract diluted by the presence of purine residues. Despite these intrinsically weak

splice signals, exon 4 is chosen constitutively in most tissues (*21*). Previous studies with the human calcitonin gene have shown that a deletion of nucleotides 18−45 within exon 4 caused exclusion of this exon in vitro (*22*). Further investigation of the human calcitonin-specific exon by mutational analysis showed that changing as few as 5 nucleotides in the first 30 nucleotides of the exon could cause skipping of the exon both in vitro and in vivo (*23*). The sequences of the rat and human genes are highly conserved in this region, suggesting that any cis element located in this region may be conserved.

Computer analysis of the 84 nucleotides surrounding the rat calcitonin exon 4 3′ splice acceptor indicates that RNA in this region is capable of forming a thermodynamically stable stem-loop structure under physiological conditions. This potential RNA secondary structure is evolutionarily conserved. To investigate whether the potential RNA stem loop structure is a cis element involved in splice regulation, we first examined whether the predicted structure exists in solution by circular dichroism and structure-specific RNase cleavage. Once the structure was defined, in vitro splicing assays were used to assess its biological function. Base changes predicted to disrupt the stem were introduced into calcitonin minigenes, and transcripts synthesized in vitro from these minigenes were spliced in HeLa nuclear extract to determine the effect of the base changes on splicing. Our results indicate that this secondary structure exists in solution and that it plays a critical role in the incorporation of exon 4 into the mature calcitonin mRNA transcript. Our data supports the hypothesis that this stem loop structure is a novel cis element involved in the splice regulation of the rat calcitonin/CGRP gene transcript.

## MATERIALS AND METHODS

*Plasmid Construction.* To examine the cis regulatory elements of the calcitonin-specific exon, exon 4, a series of plasmids were constructed for structure-specific sequencing, biophysical analysis, and splicing assays. Using site-directed mutagenesis utilizing PCR, an *Eco*RI site was engineered into intron 3 at position 1637 which is 39 nucleotides upstream from the start of exon 4. The primer 5′-CCAGGAATTCTGCATGGCACTGCCTCGC-3′ was used to construct this site. This position was chosen because it represents the border of the most minimal part of the intron needed for proper regulation of calcitonin splicing as determined by deletion analysis (*18*) and our lab. The reverse primer was the T3 primer 5′-AATTAACCCTCACTAAAGGG-3′ located at the end of the multiple cloning site of the parent plasmid pBS$^+$ Acceptor 4 (*19*). Following amplification by PCR, the fragment was cut with *Eco*RI and *Bam*HI and ligated into pBS$^+$ (Stratagene) cut with *Eco*RI and *Bam*HI, the resulting plasmid is pBS$^+$ ΔAcceptor 4. PCR mutagenesis was used to make the various base changes in the stem that we have investigated in this paper. Base changes in the 5′ part of the stem were made using the primer 5′-CTTAGATCTGGGGCTGTCC**T**A**C**CT**C**CAG**C**-**C**AAAAGACATGCGAGGCAG-3′; the altered nucleotides are underlined and in bold text. The second primer was 5′-TAATACGACTCACTATAGGG-3′. Following PCR, the product was cut with *Bgl*II and *Eco*RI. This fragment was then isolated and subcloned into pBS$^+$ ΔAcceptor 4 digested with the same enzymes. To make the 3′ stem base changes, the primer 5′-CCCAGATCT**T**A**C**CG**C**TGTG**CC**AATCT-GAGTACCTGCATG-3′ was used, with T3 as the second primer 5′-AATTAACCCTCACTAAAGGG-3′. Following PCR amplification, the product was digested with *Bgl*II and *Bam*HI and subcloned into pBS$^+$ ΔAcceptor 4 digested with the same enzymes. A minigene with both the 3′ and the 5′ stem changes was made by subcloning the *Bgl*II−*Bam*HI fragment from the 3′ stem mutant into pBS$^+$ ΔAcceptor 4 (5′ mutant) vector also cut with *Bgl*II−*Bam*HI. Minigenes containing base changes at position 59 were made by PCR mutagenesis using the primers 5′-CCCAGATCTAAGC**X**GT-GTGGGAATCTGAGT-3′ and 5′-ATTAACCCTCACTAAA-GGG-3′ as the reverse primer. **X** represents an A, C, or U in that particular position. The PCR product was digested with *Bgl*II−*Bam*HI and subcloned into pBS$^+$ ΔAcceptor 4 vector cut with the same enzymes. Plasmid DNA was prepared and subjected to DNA sequence analysis (*24*). All three mutants were isolated. For in vitro splice analysis, the *Eco*RI−*Bam*HI fragments from all of the various mutants were subcloned into the previously reported pBKS+ $\beta$g1-cal4, digested with *Eco*RI−*Bam*HI (*19*). These new plasmids are the Δ$\beta$1−4 series; they can be transcribed in vitro by T3 RNA polymerase and contain all the elements needed for in vitro analysis of tissue-specific splicing of this exon (*19*). For biophysical assays and enzymatic structure determination, 100-nucleotide *Sph*I deletions were made in ΔAcceptor 4 and its derivatives. This deletion left 39 nucleotides of intron 3 and 45 nucleotides of exon 4 remaining in the parent pBS$^+$. This plasmid can be linearized with *Hind*III and transcribed with T7 RNA polymerase.

*Circular Dichroism.* Circular dichroic spectra were measured with a Jasco-600 spectropolarimeter, equipped with a programmable temperature controller. Spectra were recorded from 300 to 190 nm and averaged over three scans. Twenty micrograms of RNA was dissolved in 200 $\mu$L of buffer consisting of 5 mM $Na_2HPO_4$, pH 7.0, 0.1 mM EDTA and 100 mM NaCl. Cylindrical cuvettes with a path length of 1 cm were used.

*RNA Secondary Structure Determination.* RNA transcripts were made in vitro using T7 RNA polymerase (New England Biolabs) and either Δ*Sph*I or mutant Δ*Sph*I plasmids linearized with *Hind*III. Transcripts were gel-purified and desorbed in 0.5 M ammonium acetate, 1 mM EDTA, and 0.1% SDS (elution buffer) overnight at 37 °C. Following extraction and precipitation with ethanol, transcripts were dephosphorylated using calf alkaline phosphatase (CIP). Mixtures were then incubated at 65 °C for 15 min to inactivate the CIP, extracted with phenol/chloroform, precipitated with ethanol, washed, and resuspended in 1X T4 polynucleotide kinase reaction buffer. Transcripts were then 5′ labeled with $^{32}$P[$\gamma$]-ATP and T4 polynucleotide kinase. Radiolabeled transcripts were gel-purified on a 6% sequencing gel and desorbed in elution buffer overnight at 37 °C. Following extraction with phenol/chloroform and ethanol precipitation, transcripts were redissolved in diethylpyrocarbonate-treated dd$H_2O$ at a specific activity of 50 000 Cerenkov counts/$\mu$L.

RNA used for structure determination was heated to 70 °C for one minute and allowed to cool to 37 °C for twenty minutes to ensure a similar population of thermodynamically stable structures. Statistical cleavage of the RNA (less than 1 cut/molecule) was carried out using RNases specific for

**A.**

**Wildtype**

a. UUUUCCCUGC*ag*CUUGGACAGCCCCAGAUCUAAGCGGUGUGGGA

**Loop mutations**

b. UUUUCCCUGC*ag*CUUGGACA**GGGG**AGAUCUAAGCGGUGUGGGA

**5′ Stem mutations**

c. UUUU**GG**CUG**G***ag***G**U**A**GGACACCCCAGAUCUAAGCGGUGUGGGA

**3′ Stem mutations**

d. UUUUCCCUGC*ag*CUUGGACACCCCAGAUCU**U**A**C**CG**C**UGUG**CC**A

**Double mutations**

e. UUUU**GG**CUG**G***ag***G**U**A**GGACACCCCAGAUCU**U**A**C**CG**C**UGUG**CC**A

**Bulge mutation G->C**

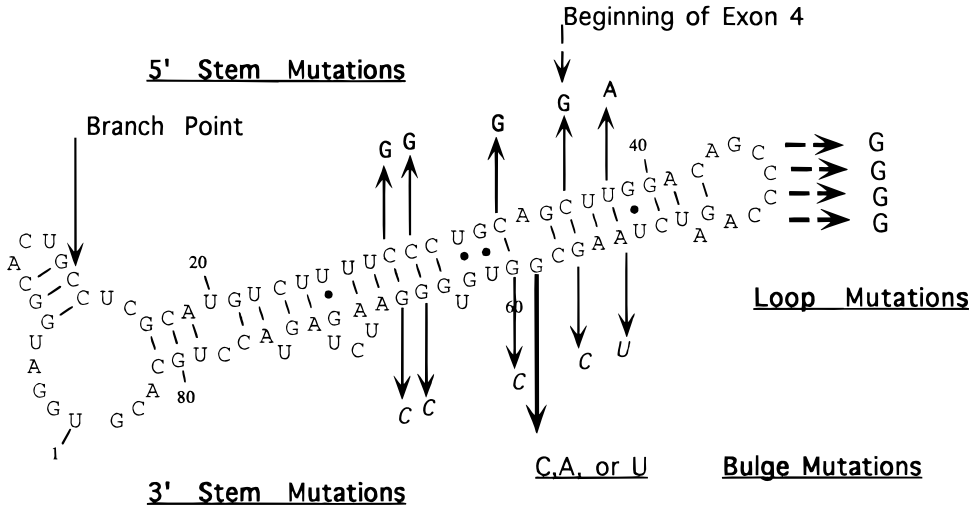
f. UUUUCCCUGC*ag*CUUGGACAGCCCCAGAUCUAAGC**C**GUGUGGGA

**Bulge mutation G->A**

g. UUUUCCCUGC*ag*CUUGGACAGCCCCAGAUCUAAGC**A**GUGUGGGA

**Bulge mutation G->U**

h. UUUUCCCUGC*ag*CUUGGACAGCCCCAGAUCUAAGC**U**GUGUGGGA




**C.**

```
MOUSE   agCTTGGACAGCCCCAGATCTAAGCGG  TGTGGGAATCTGAGTACCTGCATGC
RAT     agCTTGGACAGCCCCAGATCTAAGCGG  TGTGGGAATCTGAGTACCTGCATGC
HUMAN   agCCTGGACAGCCCCAGATCTAAGCGG  TGTGGTAATCTGAGTACTTGCATGC
CANINE  agCCTGGACAGCTCCAGAGCTAAGCGG  TGCAGTAATCTGAGTACCTGTGTGC
SHEEP   agCCTGGACAGCTCCAGAGCTAAGCGG  TGCAGTAATCTGAGTACCTGTGTGC
Hum II  agCTAGAGCAGTCCTAGATTTAAGTAGCATATAGTAATCTGAGTACCTGCTTGC
Rat β   agCTTGAGCAGTCCTAGATTTAAGTAGCATATAGTAATCTGAGTACCTGCTTGC
```

FIGURE 2: (Panel A) Sequences of the wild-type and mutant exon 4 3′ splice acceptors. The intron terminal dinucleotides (*ag*) are in *italics*. Mutations are underlined and in **bold** type. (Panel B) Graphical representation of base changes. 5′ mutations are in **bold**. 3′ mutations are in italics, and bulge mutations are underlined. (Panel C) Sequences of the mouse, rat, human, sheep, canine, human II, and rat $\beta$-CGRP exon 4 3′ splice acceptor. *Note that the mouse, rat, human, canine, and sheep acceptors are used as splice acceptors while the human II and the rat beta are not used under any conditions.* The intron terminal dinucleotide is in *italics*, and the conserved bulge nucleotide is in **bold.** Alignment was made by using the PRETTY program of the Genetics Computer Group (University of Wisconsin Biotechnology Center).

single strands $T_1$, $T_2$, and the double-strand-specific $CV_1$. Cleavage reactions were generally carried out as follows (*25*): 1 $\mu$L (50 000 Cerenkov cts) of 5′-labeled RNA was mixed with 2 $\mu$L of 2XTMK buffer (10 mM Tris-HCl, pH 7.5, 2 mM $MgCl_2$, 200 mM KCl). One microliter of the appropriate enzyme was added, and the reaction was allowed to incubate on ice for 5−10 min. The following amounts of RNases were added: $7 \times 10^{-3}$ units of $CV_1$, $1 \times 10^{-2}$ units of $T_1$, and 1 unit of $T_2$. Reactions were stopped by adding 5 uL of urea loading buffer, quickly heating to 70 °C for 30 s, then loading onto either a 6% or 10% sequencing gel. $T_1$, $U_2$, and hydroxide ladders were generated using Pharmacia's RNA sequencing protocol. Gels were dried and exposed to X-ray film overnight at −70 °C.

*In Vitro Splicing.* HeLa nuclear extract was isolated by the method of Dignam (*26*). In vitro splicing reactions were

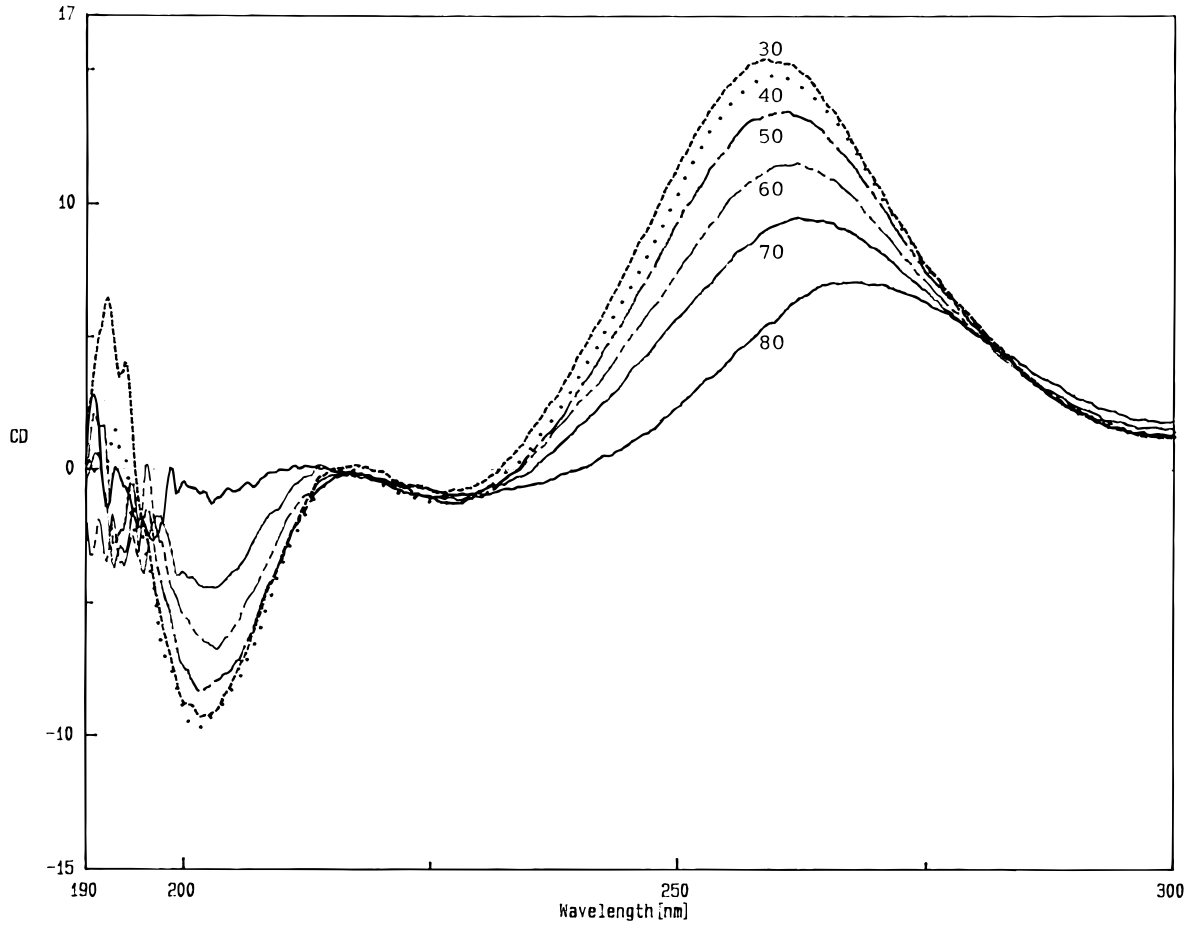



FIGURE 3: Circular dichroism spectra of Δ*Sph*I RNA as a function of temperature. At low temperatures, the CD spectra showed a highly structured RNA with a large amount of helix. The spectra were collected in 5 mM $Na_2HPO_4$, pH 7.0, 100 mM NaCl, and 0.1 mM EDTA in a 1-cm-path-length cell on a Jasco J-600 spectropolarimeter.

carried out essentially as described (*27*). Reactions were carried out at 30 °C in a total volume of 20 μL containing 60% HeLa nuclear extract in Dignam Buffer D (20 mM HEPES, pH 7.9, 20% (v/v) glycerol, 0.1 M KCl, 0.2 mM EDTA, 0.5 mM PMSF, and 0.5 mM DTT). Splicing products were analyzed by 6% denaturing polyacrylamide gel electrophoresis (PAGE). Putative splice RNAs were identified by excision from gels and elution in 0.5 M ammonium acetate, 10 mM EDTA, and 0.1% SDS overnight at 37 °C. The RNAs were then reverse transcribed using the primers 5′-CCCGCATGCAATTGGGGTTGGAG-3′ and 5′-ATGGTGCACCTGACTCCTGA-3′, amplified by PCR, and characterized by DNA sequence analysis (*24*).

## RESULTS

*RNA Model System.* To examine potential RNA secondary structure in the calcitonin exon 4 splice acceptor region, we utilized an RNA termed Δ*Sph*I that contained the proposed stem loop region (Figure 1C) for circular dichroism and RNase structure probing experiments. Δ*Sph*I and mutant Δ*Sph*I RNAs were synthesized in vitro and contained five additional nucleotides at the 5′ end (5′-GGGAG-3′) and one extra A at the 3′ end derived from cloning. This transcript maintains the sequence and structural elements necessary to form the proposed helical structure found in the native transcript. All mutations introduced into this system were based on the proposed structure shown in Figure 2B. Comparative analysis of the canine, sheep, rat, human, and mouse transcripts shows that the potential RNA structure is evolutionarily conserved (Figure 2C).

*Circular Dichroism (CD) Spectroscopic Analysis.* The conformation of the Δ*Sph*I RNA was first analyzed by CD spectroscopy to confirm that the molecule had a spectrum characteristic of an A-form RNA double-stranded helix. Characteristic features of A-form helical conformation include a positive peak at 262 nm, a small negative signal at approximately 235 nm, and a large negative signal at 209 nm (*28*−*31*). At low temperatures, the CD spectra indicated the presence of a highly ordered RNA conformation with a significant amount of A-form helix (Figure 3). With increasing temperature, significant changes were observed at the 262 nm peak and the 209 nm minimum. These changes reflect the melting of a structured molecule. At temperatures above 83 °C the CD spectrum was typical of a nonbase-pairing single-stranded RNA.

To determine if base changes in the proposed stem region (Figure 2A) destabilize the RNA duplex, melting curves were generated using CD spectroscopy. By measuring the decrease of ellipticity at ∼260 nm with increasing temperature, one can examine the melting of duplex RNA (*28*). The change in fractional percent helix versus temperature in the linear range of the melting curves (*28*, *29*) was plotted for each RNA (data not shown) (*32*). These data indicated that RNAs with base changes in the 5′ stem and 3′ stem regions

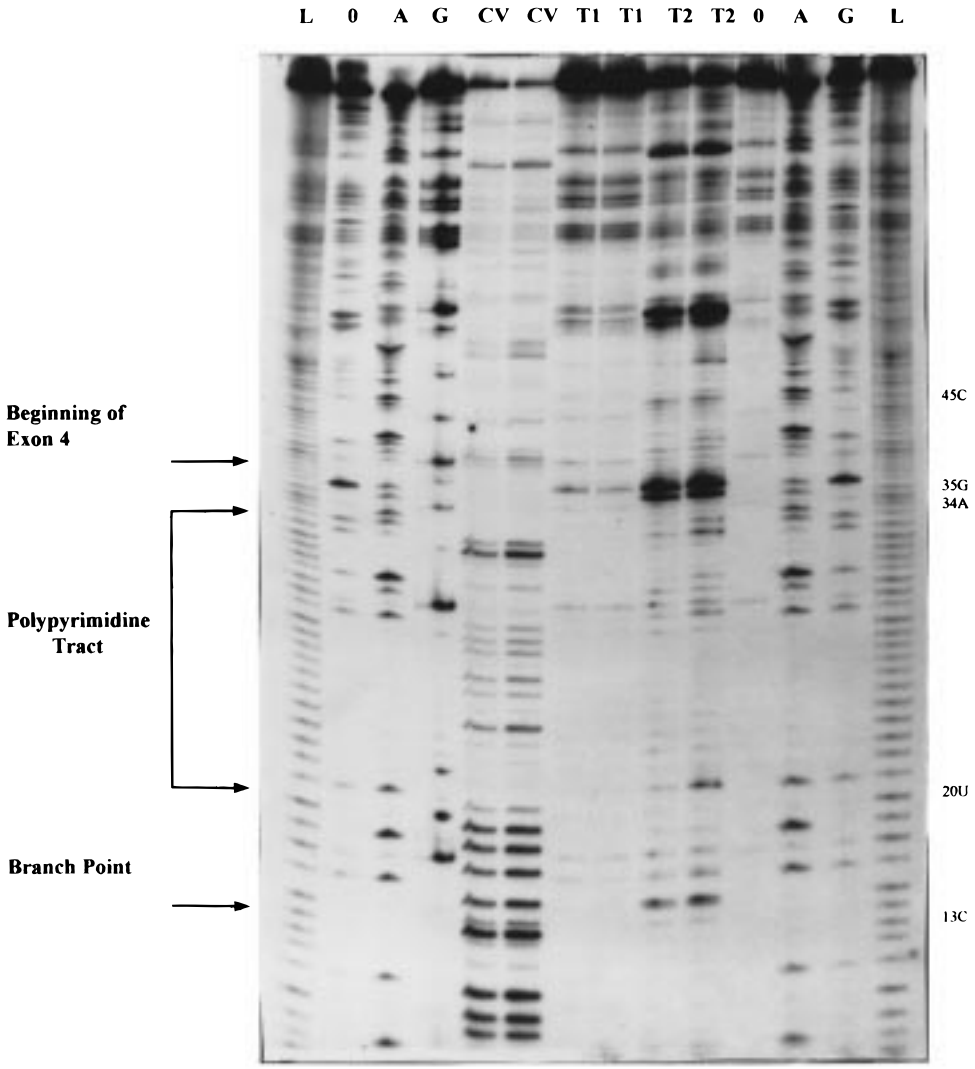



FIGURE 4: Example of nuclease mapping of the 3′ splice acceptor of rat calcitonin exon 4 model RNA, Δ*Sph*I. Autoradiogram of 10% polyacrylamide gel of cleavage products of 5′-labeled Δ*Sph*I RNA: lanes L, hydroxyl ladder; lanes 0, control RNA; lanes A, RNase $U_2$ hydrolysis ladder; lanes G, RNase $T_1$ hydrolysis ladder; lanes $CV_1$, RNase $CV_1$ mapping; lanes $T_1$, RNase $T_1$ mapping; and lanes $T_2$, RNase $T_2$ mapping. Experimental conditions for mapping were as follows: $7 \times 10^{-3}$ units of $CV_1$, $1 \times 10^{-2}$ units of RNase $T_1$, and 1 unit of $T_2$. All reactions were incubated for 10 min on ice.

had melting temperatures of approximately 52.5 °C while transcripts containing both 5′ and 3′ stem changes (double transcript, Figure 2A) and wild-type transcripts melted at approximately 58.5 °C. CD was used, instead of UV spectroscopy, to measure melting point, because of concerns that the mutant RNAs would be in several conformations, all having different melting transitions. However, as illustrated in the representative CD spectra (Figure 3), two isoelliptic points were seen at 225 and 275 nm. Isoelliptic points are indicative of a molecule that exists as a single conformation and is undergoing a single transition (*33*). We conclude from these experiments that the 3′ splice junction of exon 4 was in a double-stranded helical conformation. Base changes in the stem predicted to destabilize the stem lowered the melting temperature of the molecule. Compensatory base changes restored the melting temperature to the wild-type level, indicating that these compensatory changes restored the predicted stem.

*Ribonuclease Digestion Analysis.* To identify more precisely the nucleotides involved in the RNA double-stranded helix shown to exist by CD analysis, Δ*Sph*I transcript was probed with structure-specific ribonucleases. RNA transcribed in vitro was subjected to statistical cleavage (less than 1 cut/molecule) under native conditions. RNase $CV_1$, which is specific for double-stranded or stacked nucleotides, and the single-strand specific RNases $T_1$ or $T_2$ were used to probe the structure, as described in Materials and Methods.

The RNase probing results provide strong evidence for the existence of the stem loop structure. A representative gel is shown in Figure 4, and a summary of the RNase probing results is given in Figure 5. Strong CV1 cleavages were always observed near the noncanonical branchpoint, nucleotides 10−16. The polypyrimidine tract of the splice acceptor is interrupted by several purines, making it a poor match to the mammalian consensus sequence (*34*−*36*). Our results indicate that cleavages by the double-strand-specific RNase $CV_1$ were consistently strong within the region that makes up the polypyrimidine tract (Figure 4, nucleotides 19−30). These results indicate that the polypyrimidine tract could be recognized inefficiently due to two factors (*35*, *36*). First, this region contains several purines, which have been shown to reduce the affinity for $U2AF^{65}$, an important component in early spliceosome formation (*36*). Second, our results showed the polypyrimidine tract as part of a stable

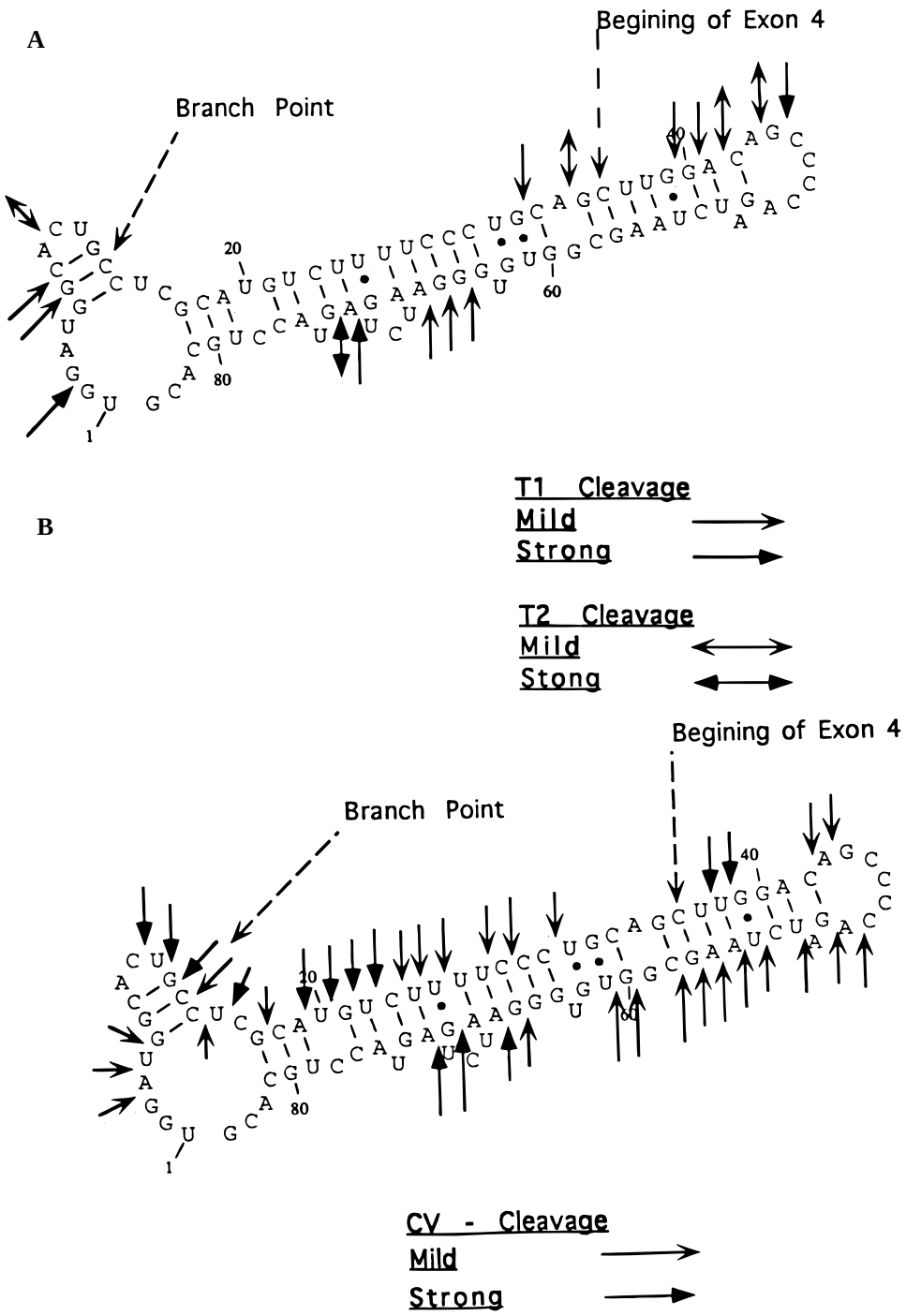



FIGURE 5: Composite of cleavages overlaid on predicted secondary structure. (A) Single-strand-specific cleavages from RNase $T_1$ and $T_2$. (B) Double-strand-specific cleavages from $CV_1$. Intensities of cuts are proportional to the darkness of the symbol.

RNA duplex which might have a reduced affinity for constitutive splice factors (*36*). However, the magnitude of cleavages suggests that these nucleotides were accessible to CV1. This implies that constitutive trans-acting splice factors, needed for spliceosome complex formation, may have reduced affinity for this region but still have access to this region in the native transcript. Therefore, it is possible that this structure and the presence of purines in this region slow the process of substrate recognition by spliceosomal components.

The lack of $CV_1$ cleavages surrounding the intron terminal A34 residue may indicate that this region of the RNA duplex is "breathing". $T_2$ and $T_1$ cleavages at residues A34 and G32 indicate that this portion of the stem was in an accessible tertiary conformation which had more single-stranded than double-stranded characteristics. Other nucleotides that were cleaved by these single-strand specific nucleases were in the predicted loop region. The loop region also had some strong CV1 cleavages. These cleavages were at the base of the predicted loop that would be consistent with the formation of a non-Watson−Crick hydrogen base pair (A−G) or stable stacking interactions (*25*).

The guanosine (G59) predicted to form a bulge in the stem region opposite A34 was cleaved weakly with both CV1 and $T_1$. This type of cleavage is indicative of stacked bases or Hoogstein base pairing (*25*, *37*). However, since the A34

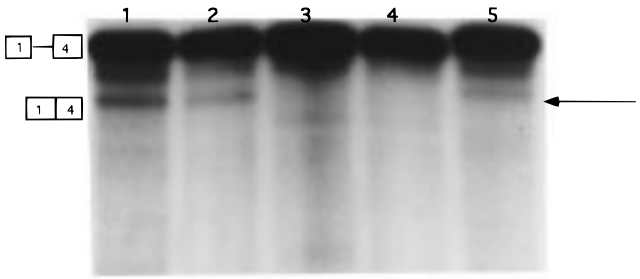



FIGURE 6: In vitro splicing of $\Delta\beta1-4$ and its mutant derivatives (5′, 3′, loop, and double (see Figure 2A,B)). Splicing was carried out in HeLa nuclear extract (*19*) using $^{32}$P-labeled transcripts synthesized in vitro. Splice products were separated by PAGE on 10% nondenaturing gels and products visualized by autoradiography: lane 1, wild-type acceptor; lane 2, loop mutant; lane 3, 5′ stem mutant; lane 4, 3′ stem mutant; and lane 5, double stem mutant. The band in lane 3 migrating slightly faster than spliced RNA was not present in all splicing experiments and was not ATP-dependent and so was not characterized further.

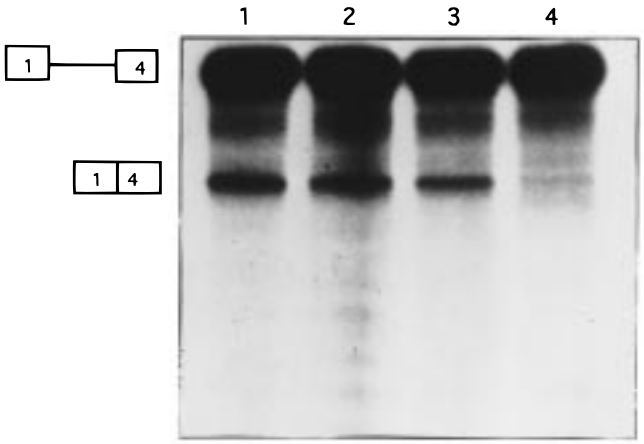



FIGURE 7: In vitro splicing of $\Delta\beta1-4$ and the bulge mutant transcripts (see Figure 2A,B) was carried out as in Figure 5: lane 1, wild-type; lane 2, G → A mutant; lane 3, G → C mutant; and lane 4, G → U mutant.

was not cut at all by CV1, G59 was most likely forming a thermodynamically stable coaxial stacking (*38*) interaction with the surrounding stem nucleotides instead of forming hydrogen bonds with A34. Overall, the computer predicted secondary structure is consistent with our enzymatic structural analysis.

*In Vitro Splicing Assays.* To test whether the RNA secondary structure at the exon 4 splice acceptor has an effect on splice site recognition, the effects of base changes in the stem (Figure 2B) were examined using in vitro splicing reactions with HeLa nuclear extract. $\Delta\beta1-4$, a chimeric minigene containing the first exon of the human $\beta$-globin gene and calcitonin exon 4, was previously shown to act as a splice substrate (*19*). $\Delta\beta1-4$ minigenes with base changes in the exon 4 splice acceptor region were created by replacing the *Eco*RI−*Sph*I of $\Delta\beta1-4$ minigene with an altered *Eco*RI−*Sph*I fragment. For in vitro splicing, $\Delta\beta1-4$, $\Delta\beta1-4^{loop}$, $\Delta\beta1-4^{5'}$, $\Delta\beta1-4^{3'}$, $\Delta\beta1-4^{bulge}$ RNAs, and $\Delta\beta1-4^{double}$, which contains the base changes in $\Delta\beta1-4^{3'}$ and $\Delta\beta1-4^{5'}$ (Figure 2B), were synthesized in vitro with T3 RNA polymerase, incubated with HeLa nuclear extract, and then examined for the formation of splice products. $\Delta\beta1-4$ and $\Delta\beta1-4^{loop}$ were spliced accurately in HeLa nuclear extracts (Figure 6, lanes 1 and 2). Transcripts with base changes in the 5′ or 3′ stem regions predicted to destabilize the RNA stem loop structure were not substrates for splicing (Figure 6, lanes 3 and 4). The $\Delta\beta1-4$ transcript containing changes in both of the 3′ and 5′ stem regions restoring possible base-pairing interactions (double mutant) is utilized as a splice substrate approximately 50% as well as wild-type $\Delta\beta1-4$ (Figure 6, lane 5). These results indicate that the base changes predicted to destabilize the predicted stem inhibit splicing, while compensatory changes restore splicing.

Our model suggests that the terminal A34 residue of the intron terminal AG dinucleotide is located in a bulge conformation opposite residue G59. To determine whether such a bulge has a function in the recognition of exon 4 during splicing, residue G59 was changed to either an A, C, or U residue (Figure 2B). Of these three changes, only the G to U change was predicted to allow formation of an A−U basepair with A34. The results of splicing assays, which test whether this bulge is critical for splicing in vitro, are shown in Figure 7. As can be seen in lanes 2 and 3, changing G59 to A or C, which are incapable of Watson−Crick basepairing with A34, did not affect splicing in vitro. However, when G59 was changed to U59, allowing the formation of an A−U basepair, the mutated transcript was not a substrate for in vitro splicing (lane 4). This suggests that a bulge in the RNA stem at this position was necessary for splice site recognition.

## DISCUSSION

Alternative splicing often involves the preferential inclusion or exclusion of an exon in the final messenger RNA. Exon 4 of the calcitonin/CGRP gene is alternatively spliced in a tissue-specific manner (*18*, *19*, *21*). It was shown previously that exon 4 is surrounded by weak splice signals (*19*, *39*). We are interested in understanding how this exon is constitutively chosen in most tissues (*19*, *21*) while remaining weak enough to be effectively switched off in neurons by trans-acting factors (*19*). Features which make the 3′ splice acceptor of exon 4 suboptimal (*18*) are a noncanonical branch point and a short polypyrimidine tract interspersed with purines (*23*). Increasingly, sequences located in exons are being shown to play a critical role in the selection of alternatively spliced exons. In this work we have identified a cis element that includes nucleotides from the intron and exon surrounding the 3′ splice acceptor of exon 4.

Computer-generated folding of the nucleotides surrounding the 3′ splice acceptor of calcitonin/CGRP gene exon 4 predicts that a stable RNA stem loop structure can form under physiological conditions. This potential RNA secondary structure is conserved in canine, human, rat, mouse, and sheep. Evidence for the possibility that this RNA stem loop is involved in splice site recognition comes from a second calcitonin/CGRP gene in mammalian cells (calcitonin/CGRP II). This gene transcript is processed only into CGRP mRNA and never calcitonin mRNA in all cell types, although it contains an exon 4-like sequence. The unutilized splice acceptor region of the cryptic exon 4 in calcitonin/CGRP II is 90% identical to the exon 4 splice acceptor in the calcitonin/CGRP gene, but does not have the potential to form a stable stem loop structure. Previously it had been shown that replacing the first 30 nucleotides of the human

calcitonin/CGRP I gene with the first 30 nucleotides of exon 4 of calcitonin/CGRP II gene inhibited usage of this exon in vivo. Similarly, when base changes were made in the repeat GCGGT, conserved in the rat gene, skipping of this exon also occurred (*23*). The base changes made in these previous studies targeted nucleotides involved in the RNA stem-loop structure, and these changes would be predicted to significantly destabilize the secondary structure (*23*). Taken together, these data indicate that the potential for forming a stable RNA stem-loop in the calcitonin/CGRP transcript has been evolutionarily conserved and suggest that this RNA secondary structure is involved in splice site selection.

The double-stranded helical nature of the exon 4 3′ splice acceptor RNA in solution was confirmed by CD spectroscopy. Melting curves generated for the RNAs containing base changes predicted to destabilize the helix had lower melting temperatures than the RNA with compensatory changes or wild-type RNA. This suggests that a double-stranded helix exists at the 3′ splice acceptor of exon 4, as predicted by the computer secondary structure-folding program. Large differences in melting points were not observed because thermodynamically stable structures are predicted for the 5′ and 3′ stem mutants by computer analysis. However, none of these structures is predicted to be as stable as the structures predicted for RNA with wild-type RNA sequences or RNA with compensatory changes that restore the wild-type stem loop structure. The isoelliptic points at 270 and 220 nm of the wild-type splice acceptor RNA indicate that the molecule essentially follows a two-state melting transition (Figure 3). These results indicate that a single thermodynamically stable structure does form at the 3′ splice acceptor of exon 4.

To identify the nucleotides involved in the RNA secondary structure, enzymatic secondary structural probing was performed. These data suggest that the computer-predicted RNA secondary structure exists in solution and sequence homology between species indicates that the structure appears to be evolutionarily conserved (Figure 2C). The branchpoint and the polypyrimidine tract sequences are double-stranded but remain accessible in the overall tertiary structure of the molecule. This interpretation is consistent with the high extent of the double-stranded cleavages present under native conditions. The exon nucleotides proposed to be involved in the helix are cleaved with $CV_1$ but not as effectively as nucleotides found at the branchpoint and polypyrimidine tract. These results indicate that the molecule is essentially double-stranded, and also that the overall tertiary structure may keep the splice signals located in the intron accessible. This is important because trans-acting molecules needed for spliceosome formation, such as $U2AF^{65}$/$U2AF^{35}$, polypyrimidine binding protein (*36*), and U2 snRNP, would still have access to their sites of interaction with the pre-mRNA.

To determine whether the RNA secondary structure we have identified is important in splice site selection of exon 4, we used RNAs with base changes predicted to destabilize the putative stem as potential splice substrates for in vitro splicing reactions. RNAs containing base changes that disrupt the RNA stem were not spliced in HeLa nuclear extract. However, an RNA containing compensatory base changes which restore the helical stem structure was spliced effectively under the same conditions. These results are in agreement with earlier in vivo data with the human gene (*23*) and suggest that the secondary structure present at the 3′ splice acceptor is necessary for constitutive usage of this splice acceptor.

The computer-predicted RNA secondary structure indicates that the adenosine residue of the intron terminal dinucleotide AG (A34) should be involved in a bulge in the helix. This was confirmed by cleavage at A34 by single-stranded specific RNase $T_2$. Both $CV_1$ and $T_1$ cut G59, the nucleotide predicted to be opposite A34 in the stem. Such cleavage patterns usually indicate that the region is in flux between single-stranded and stacked or helical form (*40*). This potential bulge is evolutionarily conserved as shown by sequence analysis (Figure 2C). Our in vitro splicing analysis showed that changing G59 to either C or A had no effect on splicing. Changing G59 to a U that would potentially base pair with A34 and prevent the formation of a bulge abolished in vitro splicing, suggesting strongly that a bulge at this position is necessary for calcitonin-specific splicing. The A34 residue was not changed because altering the intron terminal AG dinucleotide would be expected to prevent splicing to exon 4. Thus we cannot exclude the possibility that there may be a bias against a U at position 59; however this seems highly unlikely. The function of this bulge in splice site recognition has yet to be investigated but it may have one of the following effects: First, the addition of this base pair changes the calculated (*41*) stability of the helix from −26.1 to −30.1 kcal/mol at 30° C. This added stability might make the secondary structure less accessible to splice factors needed to form an active spliceosome. It may be necessary to melt the stem-loop structure during the process of splice site selection, and making the stem more thermodynamically stable may prevent the necessary disruption of the stem. Second, it may be important for the intron terminal dinucleotide AG not to be involved in base pairing during the second step of splicing. Third, bulges represent unique structures in the overall folding of RNA molecules. Cellular factors involved in the maturation of RNA molecules have been found that have a high affinity for bulges (*42*). It is possible that a bulge in this position is recognized by a specific trans-factor necessary for splice site recognition.

The presence of a stable RNA secondary structure has been observed at other alternatively used splice junctions (*16*, *43*). However most of these structures lead to reduced usage of the nearby splice site. In our system, the stem appears to be necessary for the efficient recognition of the weak splice signals at this acceptor site. Our data demonstrate that stable RNA structures can effectively participate in constitutive splicing. A role for RNA secondary structure, involved in the tissue-specific splicing of the calcitonin/CGRP gene, was first proposed by Leff et al. (*39*). They suggested that an RNA secondary structure is promoted by a neuronal regulatory factor at the exon 4 3′ splice acceptor to promote skipping of this exon in neurons. Our data amend this model to suggest that the stable structure exists in the native transcript and functions as a cis-acting element necessary for calcitonin-specific splicing. Our results are consistent with a model whereby sequences proximal to the 3′ splice junction of exon 4 are involved in a secondary structure that contributes to the regulation of this exon. We propose that this structure has two functions. First, the intron portion of the helix may act as a molecular scaffold to ensure proper

orientation of the suboptimal splice signals surrounding this exon for recognition by constitutive splicing machinery. Its second function may be similar to that of the secondary structures found in some large yeast introns. These structures reduce the distance between splice signals to ensure proper splicing. Van Oers et al. (*44*) have identified two elements in exon 4 needed for constitutive splicing of this exon in the human calcitonin/CGRP gene. One of these elements resembles a purine-rich exonic splice enhancer (*44*). These elements are conserved in most mammals and appear to function in a similar manner in the rat transcript. This structure could act to reduce the distance between the splice enhancer elements located in the exon and the suboptimal splice signals located in the intron. Disrupting the helix inhibits in vitro splicing in a fashion similar to deleting the exon splice enhancers (ESE).

## ACKNOWLEDGMENT

We thank Dr. Liberti for reading the manuscript and for his helpful advice.

## REFERENCES

1. Horowitz, D. S., and Krainer, A. R. (1993) *Trends Genet. 10*, 100−106.
2. Smith, C. W. J., Patton, J. G., and Nadal-Grinard, B. (1989) *Ann. Rev. Genet. 23*, 527−577.
3. Black, D. L. (1995) *RNA 1*, 763−771.
4. Estes, P. A., Cooke, N. E., and Liebhaber, S. A. (1992) *J. Biol. Chem. 21*, 14902−14908.
5. Chan, R. C., and Black, D. L. (1995) *Mol. Cell. Biol. 15*, 6377−6385.
6. Tanaka, K., Watakabe, A., and Shimura, Y. (1994). *Mol. Cell. Biol. 14*, 1347−1354.
7. Lynch, K. W., and Maniatis, T. (1996) *Genes Dev. 10*, 2089−2101.
8. Eperon, L. P., Graham, I. R., Griffiths, A. D., and Eperon, I. C. (1988) *Cell 54*, 393−401.
9. Goguel, V., Wang, Y., and Rosbash, M. (1993). *Mol. Cell. Biol. 13*, 6841−6848.
10. Libri, D., Piseri, A., and Fiszman, M. Y. (1991) *Science 252*, 1842−1845.
11. Blanchette, M., and Chabot, B. (1997) *RNA 3*, 405−419.
12. Watakabe, A., Inoue, K., Sakamoto, H., and Shimura, Y. (1989) *Nucleic Acids Res. 17*, 8159−8169.
13. Balvay, L., Libri, D., and Fiszman, M. Y. (1993) *BioEssays 15*, 165−169.
14. Libri, D., Stutz, F., McCarthy, T., and Rosbash, M. (1995) *RNA 1*, 425−436.
15. Domenjoud, L., Gallinaro, H., Kister, L., Meyer, S., and Jacob, M. (1991) *Mol. Cell. Biol. 11*, 4581−4590.
16. d'Orval, B. C., d'Aubenton-Carafa, Y., Marie, J., and Brody, E. (1991) *J. Mol. Biol. 221*, 837−856.
17. Rosenfeld, M. G., Mermod, J., Amara, S. G., Swanson, L. W., Sawchenko, P. E., River, J., Vale, W. W., and Evans, R. M. (1983) *Nature 304*, 129−135.
18. Emeson, R. B., Hedjran, F., Yeakley, J. M., Guise, J. W., and Rosenfeld, M. G. (1989) *Nature 341*, 76−80.
19. Roesser, J. R., Liittschwager, K., and Leff, S. E. (1993) *J. Biol. Chem. 238*, 8366−8375.
20. Lou, H., Gagel, R. F., and Berget, S. M. (1995) *Genes Dev. 20*, 208−219.
21. Crenshaw, E. B., Russo, A. F., Swanson, L. W., and Rosenfeld, M. G. (1987) *Cell 49*, 389−398.
22. Cote, G. J., Nguyen, I. N., Berget, S. M., and Gagel, R. F. (1990) *Mol. Endocrinol.* 1744−1749.
23. Cote, G. J., Stolow, D. T., Peleg, S., Berget, S. M., and Gagel, R. F. (1992) *Nucleic Acids Res. 20*, 2361−2366.
24. Sanger, F., Nicklen, A., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463−5467.
25. Feldon, B., Florentz, C., Geige, R., and Westhof, E. (1994) *J. Mol. Biol. 235*, 508−531.
26. Dignam, D., Lebovitz, R. M., and Roeder, R. G. (1983) *Nucleic Acids Res. 11* (5), 1475−1481.
27. Krainer, A. R., Conway, G. C., and Kozak, D. (1990) *Genes Dev. 4*, 1158−1171.
28. Johnson, K. H., and Gray, D. M. (1992) *J. Biomol. Struct. Dyn. 9*, 733−745.
29. Sarkar, M., Sigurdsson, S., Sen, S., Rozners, E., Sjoberg, B., Stromberg, R., and Graslund, A. (1996) *Biochemistry 35*, 4678−4688.
30. Johnson, K. H., and Gray, D. M. (1991) *Biopolymers 31*, 385−395.
31. Shi, P., Brinton, M. A., Veal, J. M., Zhong, Y., and Wilson, W. D. (1996) *Biochemistry 35*, 4222−4230.
32. Marky, L. A., and Breslauer, K. J. (1987) *Biopolymers 26*, 1601−1620.
33. Zuo, E. T., Tanious, F. T., Wilson, W. D., Zon, G., Tan, G., and Wartell, R. M. (1990) *Biochemistry 29*, 4446−4456.
34. Sharp, P. (1994) *Cell 77*, 805−815.
35. Samuels, M. E., Bopp, D., Colvin, R. A., Roscigno, R. F., Garcia-Blanco, M. A., and Schedl, P. (1994) *Mol. Cell. Biol. 14*, 4975−4990.
36. Roscigno, R. F., Weiner, M., and Garcia-Blanco, M. A. (1993) *J. Biol. Chem. 268*, 11222−11229.
37. Tranguch, A. J., Kindelberger, D. W., Rohlman, C. E., Lee, J., and Engelke, D. R. (1994) *Biochemistry 33*, 1778−1787.
38. Wu, M., McDowell, J. A., and Turner, D. H. (1995) *Biochemistry 34*, 3204−3211.
39. Leff, S. E., Evans, R. M., and Rosenfeld, M. G. (1987) *Cell 48*, 517−524.
40. Baron, C., Westhof, E., Bock, A., and Geige, R. (1993) *J. Mol. Biol. 231*, 274−292.
41. Zuker, M. (1989) *Science 244*, 48−52.
42. Baker, B., Muckenthaler, M., Vives, E., Blanchard, A., Braddock, M., Nacken, W., Kingsman, A. J., and Kingsman, S. M. (1994) *Nucleic Acids Res. 22* (16), 3365−3372.
43. Dabeva, M. D., and Warner, J. R. (1993) *J. Biol. Chem. 268* (26), 19669−19674.
44. van Oers, C. C. M., Adema, G. J., Zandberg, H., Moen, T. C., and Baas, P. D. (1994) *Mol. Cell. Biol. 14*, 951−960.

BI9808058