Alternative RNA processing – its role in regulating expression of calcitonin/calcitonin gene-related peptide

H Lou and R F Gagel

Section of Endocrine Neoplasia and Hormonal Disorders, University of Texas M D Anderson Cancer Center, Box 15, 1515 Holcombe Blvd, Houston, Texas 77030, USA

Requests for offprints should be addressed to R F Gagel, Section of Endocrine Neoplasia and Hormonal Disorders, Box 15, 1515 Holcombe Blvd, Houston, Texas 77030, USA

Abstract

The calcitonin/calcitonin gene-related peptide (CT/CGRP) gene is one of the earliest studied examples of alternative RNA processing. The regulatory mechanisms controlling this event are poorly understood. We have identified and characterized an intron element residing in intron 4 of the human CT/CGRP gene. This intron element functions to enhance polyadenylation of an embedded alternative 3′-terminal exon within the CT/CGRP gene and is potentially involved in tissue-specific regulation of CT/CGRP RNA processing.

Journal of Endocrinology (1998) 156, 401–405

In eukaryotic nuclei, every pre-messenger RNA (pre-mRNA) undergoes post-transcriptional processing to generate mature messenger RNA (mRNA), which is exported to the cytoplasm for translation. Post-transcriptional modification is a complex process, including capping at the 5′ end, polyadenylation at the 3′ end, and splicing or removal of intervening sequences (introns); splicing results in the precise ligation of coding-sequence-containing exons. Transcription and RNA processing are tightly coupled (Steinmetz 1997): there is compelling evidence that interaction of RNA processing factors with the newly transcribed RNA occurs during transcription (Daneholt 1997). Capping and polyadenylation occur concomitantly with transcription, but splicing occurs during or after transcription (Beyer & Osheim 1988).

Errors of splicing and polyadenylation can cause the production of non-functional proteins, and form the molecular basis of a large number of diseases, including endocrine diseases (Harteveld et al. 1994, Nakai & Sakamoto 1994). For example, in one case of hereditary vitamin D rickets, a G to C point mutation at the +5 position of the 5′ donor splice site of intron 4 caused exon 4 skipping, which leads to the production of a truncated protein lacking the hormone-binding domain, through a frame-shift in translation (Hawa et al. 1996). In another example, in patients from several families in which familial persistent hyperinsulinemic hypoglycemia of infancy occurred, a mutation at the 3′ acceptor splice site was identified to cause altered splicing with cryptic 3′ splice sites leading to a defective sulfonylurea receptor (Thomas et al. 1995). In a third example, several common mutations have been mapped to CYP21B in steroid 21-hydroxylase deficiency, the leading cause of impaired cortisol synthesis in congenital adrenal hyperplasia. One of the mutations, located in intron 2, causes abnormal splicing (Owerbach et al. 1990).

Splicing is mediated by several families of trans-acting factors that interact with specific RNA sequences (Moore et al. 1993). These include 5′ and 3′ (polypyrimidine-tract and branchpoint sequence) splice site sequences that define exon–intron boundaries (Fig. 1; Moore et al. 1993). Several new classes of exon and intron elements have now been identified that interact with the rapidly growing family of RNA binding proteins (Lou et al. 1995). Two prominent members of this family include small nuclear RNA (snRNA)-containing small nuclear ribonucleoprotein complexes (snRNP) and arginine–serine-rich RNA binding proteins (SR proteins). The former recognize 5′ and 3′ splice site signals; the latter interact with both intronic and exonic elements (Moore et al. 1993). A conceptual problem has been the identification of relatively short exons (<300 nucleotides) in the context of large introns. There is now compelling evidence that exons are defined by interaction of splicing factors with their cognate 3′ and 5′ splice sites. Communication between the factors across the exon results in recognition and permits precise joining of widely separated exons (Robberson et al. 1990).
Alternative RNA processing adds an additional level of complexity to this process. Alternative RNA processing is a ubiquitous process of fundamental importance for creating biological diversity. The inclusion or exclusion of an exon from a final mRNA can completely alter the biological behavior of a molecule, thus allowing modulation of protein functions (Leff et al. 1986). The differential expression of exons is often subject to developmental or tissue-specific regulations. An increasing number of genes have been shown to undergo alternative RNA processing. These include some of the important nuclear hormone receptors and ligands, for example growth hormone receptor (Dastot et al. 1996), luteinizing hormone receptor (Bachich et al. 1994), parathyroid hormone-related protein (Southby et al. 1996) and calcitonin/calcitonin gene-related peptide (CT/CGRP) (Amara et al. 1982, Rosenfeld et al. 1984).

Alternative processing of rat and human CT/CGRP transcript is one of the earliest studied examples of alternative RNA processing (Amara et al. 1982, Rosenfeld et al. 1984). Two tissue types in which this gene is transcribed, the thyroidal C cell and specific neural cell types, provide clearly defined examples of the two alternative processing pathways. The pre-mRNA of the CT/CGRP gene contains six exons and can be processed in two ways (Fig. 2A). The key regulatory event in the processing choice is inclusion or exclusion of the alternative 3′-terminal exon, exon 4. In thyroidal C cells, 95% of the CT/CGRP pre-mRNA is processed to include exon 4, with usage of the exon 4 polyadenylation site; this leads to production of calcitonin peptide. In contrast, in neuronal cells, 99% of the CT/CGRP pre-mRNA is processed to exclude exon 4, resulting in inclusion of exons 5 and 6, with usage of the exon 6 polyadenylation site; this leads to the production of CGRP peptide. Therefore, through alternative RNA processing, two peptide hormones having completely different structures and functions are generated. Interestingly, in malignant C cells, both calcitonin and CGRP are produced in equal amounts, resulting from changes in splicing pathways (Benett & Amara 1993).

To understand the underlying mechanism that controls the regulation of CT/CGRP alternative RNA processing, several research groups have developed cell model systems to mimic the tissue-specific RNA processing pathways (Benett & Amara 1993). Our group has chosen HeLa cells for calcitonin processing (exon 4 inclusion) and T98G, a human glioblastoma cell line, for CGRP processing (exon 4 exclusion) (Fig. 2A; Lou et al. 1994, 1995). Because the fate of exon 4 is a key issue for this processing choice, that exon became the initial focus of these studies.

Figure 1 Diagram showing the constitutive exon recognition of a generic internal exon and a generic terminal exon. The sequence elements involved in splicing and polyadenylation and the factors binding to these sequences are illustrated (Moore et al. 1993, Proudfoot 1996). Y, pyrimidine; R, purine, n>13 for most of the mammalian introns.

Figure 2 A. Schematic representation of alternative RNA processing of CT/CGRP. Cell models used for this study are indicated. The black oval represents the intron element. B. Sequences of the two splicing signals in the intron element.
Exon 4 has weak RNA processing signals, in common with other examples of alternatively utilized exons. The first weak signal is the 3' splice site, which includes a non-canonical branchpoint nucleotide uridine in human or cytidine in rat, instead of the canonical branchpoint adenosine (Adema et al. 1988, Benett & Amara 1993) and a short, purine-interrupted polypyrimidine-tract (see Fig. 1 for consensus sequences). The second weak signal is the polyadenylation signal, including a suboptimal G/U sequence downstream of the polyadenylation cleavage site, that is required for binding of the 64 kDa CstF polyadenylation factor (Fig. 1). Terminal exons, for example exons 4 and 6 of the CT/CGRP gene, form a unique adenylation factor (Fig. 1). Terminal exons, for example exons 4 and 6 of the CT/CGRP gene, form a unique adenylation factor (Fig. 1). Terminal exons, for example exons 4 and 6 of the CT/CGRP gene, form a unique adenylation factor (Fig. 1). Terminal exons, for example exons 4 and 6 of the CT/CGRP gene, form a unique adenylation factor (Fig. 1).

Figure 3 Interactions of splicing factors with the CT/CGRP intron element. Potential interactions between factors are indicated by thick arrows.

Figure 3 Interactions of splicing factors with the CT/CGRP intron element. Potential interactions between factors are indicated by thick arrows.

The concept is evolving in the field of alternative RNA processing that normal RNA processing factors do not recognize these weak sequences well, and additional factors are required to recognize weak exons. We hypothesized that these factors bind to exon 4 or flanking intronic sequences to enhance recognition of this exon. That these factors are ubiquitous is suggested by experiments with transgenic mice expressing rat CT/CGRP gene, which showed exon 4 inclusion in most tissues, and exclusion only in neuronal and cardiac tissues (Crenshaw et al. 1987).

Support for our hypothesis is provided by the identification of several sequence elements, including three exon 4 sequence motifs, a sequence upstream of the exon 4, and a sequence downstream of exon 4 (Emeson et al. 1989, Cote et al. 1992, Roesser et al. 1993, van Oers et al. 1994, Yeakley et al. 1993). We have focused on a sequence element located downstream of exon 4, because it has several interesting features. First, the initial 276-nucleotide intron 4 sequence, extending from the AATAAA hexanucleotide of exon 4, is 63% identical in human, mouse and rat species (Lou et al. 1995). This is unusual for intron sequences, and suggests that it has important functions, as intron sequences are not normally conserved. Secondly, buried in the middle of intron 4, there are two splicing signals: a perfect 5' splice site sequence and a pyrimidine-tract sequence that immediately precedes the 5' splice site sequence (Fig. 2B; Lou et al. 1995). These two sequences are out of context; they are normally found at the 5' and 3' ends of an intron (Fig. 1). Particularly intriguing is the presence of a 5' splice site sequence downstream of a polyadenylation signal, because 5' splice site sequences in last exons have been shown to affect polyadenylation both positively and negatively (Furth et al. 1994, Niwa et al. 1992, Wassarman & Steitz 1993). We hypothesized that a good 5' splice site sequence located downstream of exon 4 could facilitate exon 4 polyadenylation.

We established the importance of the two intronic splicing signals by extensive mutational analyses (Lou et al. 1995). Deletion or mutation of the 5' splice site sequence or the pyrimidine-tract sequence dramatically decreased recognition of exon 4 when the mutant construct was transfected into a cell type (HeLa) that normally includes exon 4 (Lou et al. 1995). This effect is caused, at least in part, by regulation of polyadenylation; the presence of intact 5' splice site and pyrimidine-tract sequences was necessary for polyadenylation. Mutation of these sequences resulted in decreased polyadenylation in an in vitro polyadenylation cleavage assay, utilizing a truncated exon 4 polyadenylation site and HeLa cell nuclear extract (Lou et al. 1996). This is the first example of regulation of polyadenylation by an intron element. As the 5' splice site and pyrimidine-tract sequences are general splicing signals, it is conceivable that splicing factors bind to these sequences and interact with polyadenylation factors. The 5' splice site is usually recognized by U1 snRNP through base pairing and can also bind other splicing factors, including members of the SR family of proteins (Kohtz et al. 1994). We demonstrated that U1 snRNA binds to the 5' splice site sequence in the calcitonin intron element, by use of RNase H protection (Fig. 3, Lou et al. 1996) and the psoralen cross-linking assay (a method for detecting
RNA–RNA interactions; H Lou, unpublished observation. The functional relevance of the base pairing between U1 snRNA and the intronic 5′ splice site was shown by demonstrating the abolition of exon 4 in vitro polyadenylation cleavage activity in the absence of U1 snRNA (Lou et al. 1996).

The 5′ splice site sequence in the calcitonin intron element also binds other splicing factors, analogous to a true 5′ splice site. One of the factors involved is alternative splicing factor (ASF/SF2) (Fig. 3). The interaction of splicing factors with this element raised a question: why is this 5′ splice site sequence not utilized as a 5′ splice site? A clue is provided by the binding of polypyrimidine-tract binding protein (PTB) to the pyrimidine-tract sequence preceding the 5′ splice site sequence (Lou et al. 1996; Fig. 3). PTB is generally considered to be a negative splicing factor. It has been shown to bind to the pyrimidine-tract at the 3′ splice site and compete with the essential splicing factor U2 snRNP auxiliary factor (U2AF) – an interaction that is required for splicing (Zamore & Green 1991). PTB has also been implicated in the tissue-specific repression of splicing for several genes (Lin & Patton 1995, Chan & Black 1997). It is therefore possible that PTB prevents U1 snRNA from interacting with other splicing factors, thereby preventing the potential 5′ splice site in the enhancer from being used as a true 5′ splice site. The U1 snRNA bound at the 5′ splice site sequence would then be available to interact with non-splicing components, for example polyadenylation factors, resulting in enhanced exon 4 polyadenylation. We are currently investigating the mechanism by which splicing components assembled on the calcitonin intron element interact with polyadenylation factors.

If this intron element has a role in the cell-specific regulation of CT/CGRP RNA processing, how is the regulation achieved? One simple, but not applicable, explanation might be the presence of different amounts of the factors that interact with the intron element in cells that include or exclude exon 4; however, western blot analysis showed similar levels of PTB or ASF/SF2 in HeLa and T98G cells (unpublished observation). There are several other possible explanations. First, PTB or ASF/SF2 could be modified differently in the two cell types, for example by differential phosphorylation. Secondly, additional unidentified proteins may be involved in the regulation. A third possibility, which we favor, is the presence of a factor that interferes with the interaction between U1 snRNA, ASF/SF2, PTB or other positive-acting factors and the intron element in cells that exclude exon 4. We and others are searching for such a factor. A few neuron-specific RNA binding proteins have been isolated (Koushika et al. 1996). It would not be surprising if one of these neuron-specific proteins were shown to be involved in the regulation of CT/CGRP RNA processing.

Acknowledgements

This work was carried out in collaboration with Dr Susan Berget in Baylor College of Medicine. The authors thank Yun Yang for technical assistance. This work was supported by NIH grants RO1-DK38146 (R F G) and 2P30-CA16672 (M D Anderson Cancer Center) and the American Cancer Society (NP-695 to Susan Berget).

References

Adema GJ, Bovenberg RAL, Janse HS & Baas PD 1988 Unusual branch point selection involved in splicing of the alternatively processed calcitonin/CGRP-I pre-mRNA. Nucleic Acids Research 16 9513–9526.


Furth PA, Choe W-T, Rex JH, Bryne JC & Baker CC 1994 Sequences homologous to 5′ splice sites are required for the inhibitory activity of papillomavirus late 3′ untranslated regions. Molecular and Cellular Biology 14 5278–5289.


Proudfoot N 1996 Ending the message is not so simple. Cell 87 779–781.

Robberson BL, Cote GJ & Berget SM 1990 Exon definition may facilitate splice site selection in RNAs with multiple exons. Molecular and Cellular Biology 10 84–94.


Steinmetz EJ 1997 Pre-mRNA processing and the CTD of RNA polymerase II: the tail that wags the dog? Cell 89 491–494.


Received 8 August 1997
Accepted 14 October 1997