Alternative RNA processing in calcitonin gene expression generates mRNAs encoding different polypeptide products

Susan G. Amara, Vivian Jonas & Michael G. Rosenfeld
Division of Endocrinology, School of Medicine University of California, San Diego La Jolla, California 92093, USA

Estelita S. Ong, & Ronald M. Evans
Tumor Virology Laboratory, The Salk Institute, San Diego, California 92138, USA

Alternative processing of RNA transcripts from the calcitonin gene results in the production of distinct mRNAs encoding the hormone calcitonin or a predicted product referred to as calcitonin gene-related peptide (CGRP). The calcitonin mRNA predominates in the thyroid while the CGRP-specific mRNA appears to predominate in the hypothalamus. These observations lead us to propose a model in which developmental regulation of RNA processing is used to increase the diversity of neuroendocrine gene expression.

A LARGE array of hormonal signals provides the specificity of intracellular communication within and between organs. The existence of gene families[1] and the presence of multiple hormones within a single primary translation product[2–4] are two means by which a diversity of peptide hormones can be generated. The discontinuity of genetic regions encoding mature RNAs and the complexity of RNA processing pathways suggest that alternative splicing events could be an additional mechanism for increasing the flexibility of gene expression in the neuroendocrine system. The potential versatility provided by alternative RNA processing events has been used effectively by several eukaryotic viruses to generate multiple protein products from a single transcription unit[5,6]. We show here that multiple mRNAs are generated from the calcitonin gene as a consequence of alternative RNA processing events. These events ultimately produce different protein products and seem to occur in a tissue-specific fashion. Such a mechanism could serve to increase the diversity of proteins encoded by endocrine genes.

The production of multiple calcitonin-related mRNAs was first noted during the spontaneous and permanent switching of serially transplanted rat medullary thyroid carcinoma (MTC) lines from states of 'high' to 'low' calcitonin production[7]. This conversion is associated with the appearance of new calcitonin cDNA-reactive mRNAs, referred to as calcitonin -gene-related product (CGRP) mRNAs, 50–250 nucleotides larger than calcitonin mRNA. These RNAs can function as mRNAs as they are present on polyribosomes and can be translated *in vitro* after being isolated by hybridization to a specific immobilized cDNA clone. A new 16,000 molecular weight protein is the major cell-free translation product directed by the CGRP mRNAs[8]; this protein does not contain immunoreactive calcitonin.

To study the switching events occurring in the tumours, cDNA clones to calcitonin (pCal) and calcitonin gene-related ($pCGRP_1$, $pCGRP_2$) mRNAs have been generated and a calcitonin genomic fragment referred to as λ-Cal_1 has been isolated. In this article we examine the structural basis for differences between mature calcitonin mRNA (Cal mRNA) and one of the variant mRNAs (CGRP mRNA) by examination of both cDNA and genomic clones.

Cal and CGRP mRNA sequences are present in the same gene

To determine the relationship of Cal and CGRP sequences within the calcitonin gene, fragments or clones bearing calcitonin-specific or CGRP-specific sequences were used as hybridization probes to locate and orientate exons within the genomic clone, λ-Cal_1. These regions of homology are closely linked as both probes react with a common 8.2-kilobase (kb) *Eco*RI fragment and a common 3.0-kb *Pst*I fragment (Fig. 1*a*, *b*). Fragments specific to 5′ and 3′ regions of calcitonin mRNA were generated as hybridization probes to determine the transcriptional orientation of the gene. These were prepared by cleaving pCal at the single *Bgl*II site, located 11 nucleotides 5′ to the start of the calcitonin coding sequence[9]. A cDNA clone $pCGRP_1$ was used as a specific probe for CGRP sequences because it contains a region unique to CGRP mRNAs and does not hybridize Cal mRNA. The pCal- and $pCGRP_1$-reactive *Eco*RI and *Pst*I fragments of λ-Cal_1 correspond precisely to those identified in digests of genomic DNA isolated from rat liver or MTC tumours. Figure 1*c*, *d* shows the results of this analysis. Comparison of this hybridization with that in *a* and *b* indicates that CGRP sequences map to the 3′ portion of this gene. Further restriction mapping establishes that the 8.2-kb *Bgl*II fragment which hybridizes only with $pCGRP_1$ (Fig. 1*a*, lane 2) represents the 3′ end of the λ-Cal_1 clonal insert. These results support the notion that the 3′ ends of Cal and CGRP mRNAs are encoded by the same gene.

Comparison of $pCGRP_2$ and pCal cDNA coding domains

Previously described RNA hybridization analysis[8] and the genomic mapping observations presented above suggest that Cal and CGRP mRNAs contain common or homologous 5′ sequences and nonhomologous 3′ sequences. This prediction was tested by restriction analysis and sequencing of CGRP cDNA clones. One clone, designated as $pCGRP_2$, containing an insert of >600 nucleotides, was sequenced to identify the

0028–0836/82/280240—05$01.00

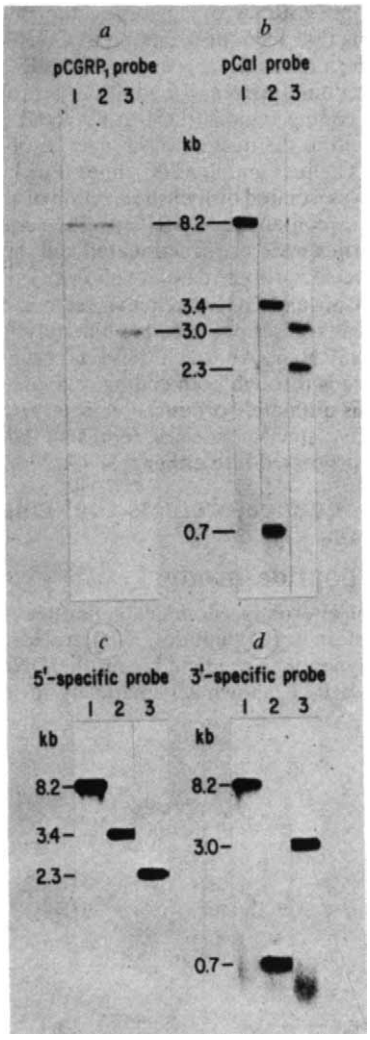

Fig. 1 Localization of pCGRP sequences in the calcitonin genomic clone λ-Cal$_1$. The rat calcitonin genomic clone (λ-Cal$_1$) was digested to completion with the restriction enzymes *Eco*RI, *Bgl*II or *Pst*I and electrophoresed on an 0.8% agaraose gel. The restriction fragments were transferred to a nitrocellulose filter[27]. Filters were hybridized to $\sim 5 \times 10^6$ c.p.m. ml^{-1} of ^{32}P-labelled nick-translated probes, washed and autoradiographed as previously described[8]. The clone pCGRP$_1$ used in this analysis contains a region unique to CGRP mRNA and does not hybridize to Cal mRNA. *a*, Hybridization of pCGRP$_1$ to λ-Cal$_1$ digested with: *Eco*RI (lane 1), *Bgl*II (lane 2), *Pst*I (lane 3). *b*, Hybridization of pCal to a duplicate blot of *a*. *c*, Hybridization of the 5′ fragment of the *Bgl*II-digested pCal insert to λ-Cal$_1$ DNA digested with: *Eco*RI (lane 1), *Bgl*II (lane 2), *Pst*I (lane 3). *d*, Hybridization of the 3′ fragment of *Bgl*II-digested pCal insert to the identical blot shown in *c*. Sizes are based on the migration of *Hin*dIII-digested λ DNA fragments.

point of divergence between Cal and CGRP mRNAs. The sequence of this clone reveals the entire coding region of CGRP mRNA with an open reading frame of 384 nucleotides capable of coding for a primary translation product of 128 amino acids.

The sequences of pCGRP$_2$ and pCal are compared in Fig. 2*a*. The 5′ sequences of pCal and pCGRP$_2$ are identical from the first nucleotide in the pCGRP$_2$ insert to nucleotide 227 in the coding region of both cDNA clones. As previously reported, the structure of pCal predicts only one open reading frame which encodes a precursor protein whose characteristics correspond to those determined for precalcitonin and its component polypeptides *in vivo* and *in vitro*[9–11]. Thus, the identity of the 5′-proximal mRNA sequences including the AUG for both mRNAs suggests that the first 76 amino acids of their translation products must also be identical. The nucleic acid sequence diverges 228 nucleotides into the coding sequence, following an AG, and precisely identifies the junction point that distinguishes these two mRNAs. Beyond nucleotide 227, the two mRNAs share no homology and encode completely different amino acid sequences. The junction of shared and divergent sequence in these two mRNAs is detailed in Fig. 2*c*. These data are most compatible with the interpretation that a single 5′ sequence represents a discrete domain which can be selectively spliced onto Cal- or CGRP-specific exons. In such an instance, the mRNA junction point must correspond to an intervening sequence in which a unique donor site can splice onto one of two alternative acceptor sites. We examined this prediction by a detailed structural analysis of the calcitonin gene.

The point of divergence between Cal and CGRP mRNAs corresponds to a genomic splice junction

Detailed restriction endonuclease mapping using the cDNA clones as probes and partial DNA sequence analysis of the calcitonin genomic clone (λ-Cal$_1$) confirms the prediction that exons with sequences precisely corresponding to the divergent 3′-coding regions of pCal and pCGRP$_2$ are present in one genomic transcription unit. As noted in Fig. 2*b*, all exons identified in the genomic clone have sequences identical to those determined for the two cDNA clones. The 5′ boundaries of the Cal- and CGRP-specific genomic coding sequences establish that splice junctions occur precisely at the point of the divergence between Cal and CGRP$_2$ mRNAs. Figure 2*c* (A) and (B) details the sequences at the junctions for the pCal and pCGRP$_2$ sequences. The observation that an AG precedes both exons in the genomic clone is in accordance with the GT . . . AG rule for splice junctions[12] and is consistent with the prediction that each exon can be alternatively spliced into the mature mRNAs. The AG which appears adjacent to the splice site in the two cDNA clones is presumably contributed by the 3′ end of the preceding exon.

Nuclear RNA transcripts contain both Cal and CGRP sequences

The structure of Cal and CGRP mRNAs and of the calcitonin genomic clone demonstrate that the same region of DNA can be used to generate multiple mature mRNAs. However, there are several possible mechanisms by which each mature mRNA could be produced. Because the sequences unique to CGRP mRNA are located 3′ to Cal mRNA sequences in the gene, it is possible that the calcitonin primary transcript could terminate at a polyadenylation site 5′ with respect to CGRP sequences, while transcription to the next polyadenylation site could give rise to the CGRP primary transcript. This model predicts two different primary transcripts and would parallel the molecular events demonstrated in the case of late adenovirus mRNA production; these events have also been proposed for the alternative generation of membrane and secretory forms of immunoglobulin μ heavy chains[13,14] and the simultaneous synthesis of μ and δ heavy chains[15].

This possibility was investigated by analysing putative nuclear RNA precursors for the presence of both Cal and CGRP mRNA-specific sequences. We have previously identified reactive pCal RNA precursors 6.6, 5.2, 4.2 and 3.8 kb in tumour lines producing CGRP calcitonin mRNA (MTC$_L$ lines)[7]. The comparably sized precursors observed in lines producing predominantly calcitonin mRNA (MTC$_H$ lines) could reflect the production of small amounts of CGRP mRNAs[8], the existence of common precursors for Cal and CGRP mRNAs, or the presence of CGRP RNA precursors not processed to mature CGRP mRNAs in this tissue. RNA from normal thyroid glands was also analysed. Virtually all the pCal-reactive RNA detected in the thyroid gland is the 1,050-nucleotide calcitonin mRNA. The presence of CGRP-specific sequences within the large

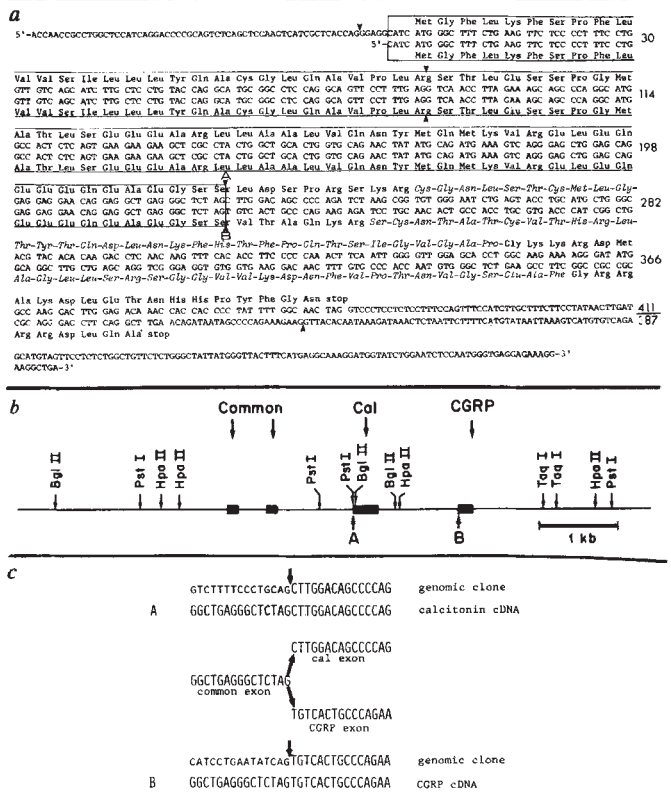

Fig. 2 The nucleotide sequence of pCal and $pCGRP_2$ and the location of corresponding regions within the calcitonin gene. The plasmid pCal, containing DNA complementary to calcitonin mRNA, was cloned and sequenced as described previously[9]. To construct plasmids complementary to CGRP mRNAs, size-fractionated mRNA 800–1,500 nucleotides long was prepared from an MTC_L line highly enriched in CGRP mRNAs and used to generate double-stranded DNA using reverse transcriptase. The double-stranded DNAs were inserted into the *Pst*I site of pBR322 by dG-dC tailing and the recombinant plasmid was used to transform an *Escherichia coli* K-12 host (SF8)[9]. Transformants were identified by their differential antibiotic sensitivity. The plasmid $pCGRP_2$ was selected for further study based on its restriction map and its hybridization to the CGRP-specific 8.2-kb *Bgl*II genomic clone fragment. *a*, The DNA sequence of the 627-base insert in pCal is shown above the sequence of 482 bases of $pCGRP_2$. Amino acid sequences of products encoded by the two clones are indicated above and below. Calcitonin and the putative 37 amino acid cleavage product of $pCGRP_2$ (CGPR) are indicated by italics. The boxed area denotes the region of common nucleotide sequence between the two clones. DNA sequencing has demonstrated that the entire coding sequence of both cDNA clones is present with precise correspondence in the genomic clone λ-Cal_1. The genomic DNA sequence corresponding to the first 58 nucleotides of the 5′ noncoding portion of pCal and the last 67 nucleotides of the 3′ noncoding portion of $pCGRP_2$ has not been completed. The sequence for most of pCal and the strategy used has been described previously[10]. To sequence $pCGRP_2$, plasmid was digested with *Dde*I, *Hae*III, *Alu*I or *Rsa*I, end-labelled with [γ-^{32}P]ATP and T4 polynucleotide kinase (Bethesda Research Laboratories) and either strand separated or cleaved with a second restriction enzyme before electrophoresis using procedures described by Maxam and Gilbert[28]. The cDNA insert was sequenced in multiple determinations, in both strands and across all sites used for labelling. Arrowheads denote sites of intervening sequences in λ-Cal_1 determined by sequencing: A and B are the junctions of the Cal-encoding and CGRP-encoding exons, respectively. Nucleotides are numbered on the right beginning with the first nucleotide by the coding region of both clones. *b*, Restriction map of λ-Cal_1. Endonuclease cleavage sites of λ-Cal_1 were determined by analysis of non-limit restriction digests as described by Smith and Birnstiel[29] and/or by comparison of double and single enzyme digestion patterns determined for the clone. Small fragments shown by hybridization to contain regions of pCal and $pCGRP_2$ sequences were isolated and 5′-end labelled for Maxam–Gilbert sequencing as above. The location of four sequenced exons is represented by black boxes. Fragments generated by *Rsa*I and *Dde*I cleavage were strand separated and sequenced completely in both strands. Sequence from *Bgl*II and *Hae*III sites was obtained by multiple determinations of sequence from the same strand. Other sequences were obtained by dideoxy sequencing of sheared genomic fragments cloned into M13[30–32]. Details of the various sequencing strategies are available from the authors on request. Arrows A and B indicate the location of the intron–exon junctions for Cal and CGRP exons, respectively. *c*, The cDNA sequences at the junction of the region of identity and non-identity in pCal and $pCGRP_2$. The genomic sequences at intron–exon junctions corresponding to those labelled A and B in panels *a* and *b* are also labelled A and B in *c*. Arrows indicate putative splice sites. Intervening sequences are shown using small letters.

nuclear calcitonin mRNA precursors in the thyroid (Fig. 3*a*, lane T) suggests that sequences unique to CGRP mRNAs may be present in the nuclear RNA transcripts of cells making almost exclusively calcitonin mRNA. In addition, a probe specific for the calcitonin coding region (the 0.7-kb *Bgl*II genomic fragment) hybridized to the nuclear RNA species of identical sizes in thyroid, MTC_H lines and in MTC_L lines. Furthermore, intact nuclear RNA was isolated by hybrid selection of a tumour RNA, using a plasmid containing CGRP-specific sequences. These purified transcripts were size-fractionated and shown to hybridize with Cal-specific probes (data not shown).

Thus, both Cal- and CGRP-reactive regions are present on common nuclear transcripts in cells ultimately producing only Cal or CGRP mRNAs. Although these data suggest that transcription proceeds through both coding regions, irrespective of which mRNA is ultimately produced, it is nevertheless possible that the reactive nuclear species represent large transcripts which are not processed into either mRNA.

$pCGRP_2$ sequence predicts that calcitonin gene encodes an additional peptide product

The splicing of alternative domains generates mRNAs which can encode different polypeptides. The structural relationship of the two polypeptides predicted from the DNA sequence of pCal and $pCGRP_2$ is shown schematically in Fig. 4. For the

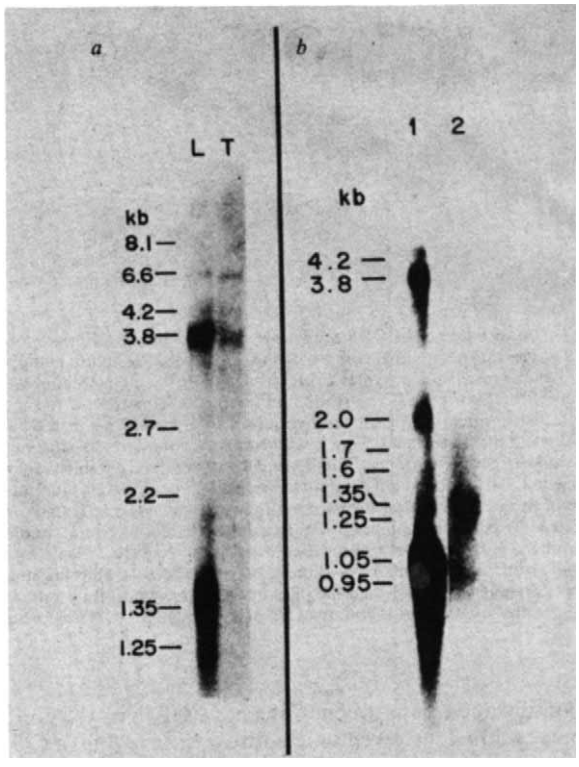

Fig. 3 *a*, Analysis of calcitonin mRNA precursors for the presence of CGRP regions. Total poly(A)-rich RNAs from an MTC_L line and from normal thyroids were denatured and electrophoresed on 1.5% agarose formaldehyde gel as described previously[33]. RNA was transferred to nitrocellulose[27], washed in prehybridization buffer and hybridized to $pCGRP_1$ nick-translated to a specific activity of 8×10^8 c.p.m. per μg using [α-^{32}P]dCTP as the labelled nucleotide. The $pCGRP_1$ probe is specific for CGRP sequences. Lane L, 5 μg MTC_L tumour mRNA; lane T, 5 μg thyroid mRNA. *b*, Analysis of hypothalamic and thyroid mRNA species hybridizing the calcitonin cDNA clone pCal. Poly(A)-rich mRNA from normal thyroid and gradient-enriched 8–15S RNA from rat hypothalamus were subjected to formaldehyde gel and Northern blotting procedures as in *a* except that transfer was to diazotysed paper[34]. Immobilized RNA was hybridized with pCal nick-translated to a specific activity of 5×10^8 c.p.m. per μg (ref. 33). Lane 1, 5 μg poly(A)-rich thyroid mRNA; lane 2, 8 μg gradient-enriched poly(A)-rich hypothalamic mRNA. RNA sizes are based on migration of 18S and 28S RNAs.

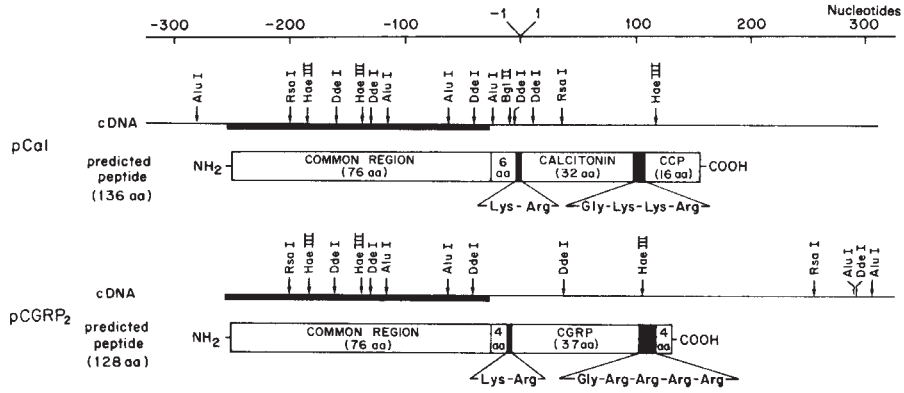

Fig. 4 Schematic representation of pCal and $pCGRP_2$ DNA inserts and the structure of the precursors they encode. Restriction sites used for DNA sequence analysis are those indicated and regions of identical nucleotide sequence are underlined in black. Nucleotide residues numbered on the scale above begin with the first residue in the coding region for calcitonin. The structure of the predicted protein precursors are represented by the bar below. Potential proteolytic processing sites and the resulting cleavage products are also noted.

calcitonin precursor, the points of excision can be accurately determined from knowledge of the mature calcitonin peptide, as we have previously discussed in detail[9,10]. The excision occurs at paired or multiple basic residues and this processing information is encoded in the calcitonin exon. The DNA sequence of $pCGRP_2$ predicts a 128 amino acid, 16,000 molecular weight protein which has the same 76 amino acid N-terminal sequence as the calcitonin precursor. The 52 C-terminal residues of the predicted protein contain similar processing signals which would allow excision of a 37 amino acid peptide. The Lys-Arg dipeptide, encoded by nucleotides 241–246 (see Fig. 2*a*), is proposed to represent the N-terminal cleavage site for excision of this novel product. The presence of glycine followed by three basic amino acid residues is thought to serve as a signal for C-terminal amidation and cleavage in three amidated peptides for which the structure of the precursor is known[4,9,16]. The sequence Gly-Arg-Arg-Arg-Arg in the product encoded by CGRP mRNA may signal the generation of a peptide terminating in a phenylalanine-amide. We refer to this 37 amino acid excised peptide as calcitonin gene-related peptide (CGRP). We note that in parallel with calcitonin, the CGRP exon encodes all the processing signals necessary to generate CGRP.

Expression of Cal and CGRP mRNAs is tissue specific

The demonstration that the 'switching' of MTC tumours from production of Cal to CGRP mRNA is the consequence of alternative RNA splicing events prompted the speculation that such events might occur physiologically during differentiation. This possibility was examined by size-fractionation of RNA from several tissues and challenge with Cal- and CGRP-specific probes. The results of one such study are shown in Fig. 3*b*. The hypothalamus contains poly(A)-rich RNAs which hybridize specifically with the Cal-specific probe and which are the same size as the CGRP RNA species present in the MTC_L tumours (Fig. 3*b*, panel 2). These RNA species hybridize to the CGRP-specific probe. A small amount of RNA corresponding to calcitonin mRNA is also present in these cells while virtually all the reactive RNA in the thyroid is calcitonin mRNA (compare Fig. 3*a*, panel T, with Fig. 3*b*, lane 1). These data suggest that the alternative RNA splicing events might occur physiologically in the expression of the calcitonin gene. It is tempting to speculate that CGRP, the predicted polypeptide product of $CGRP_2$ mRNA, is a hypothalamic neuropeptide that may exert hormonal effects of its own.

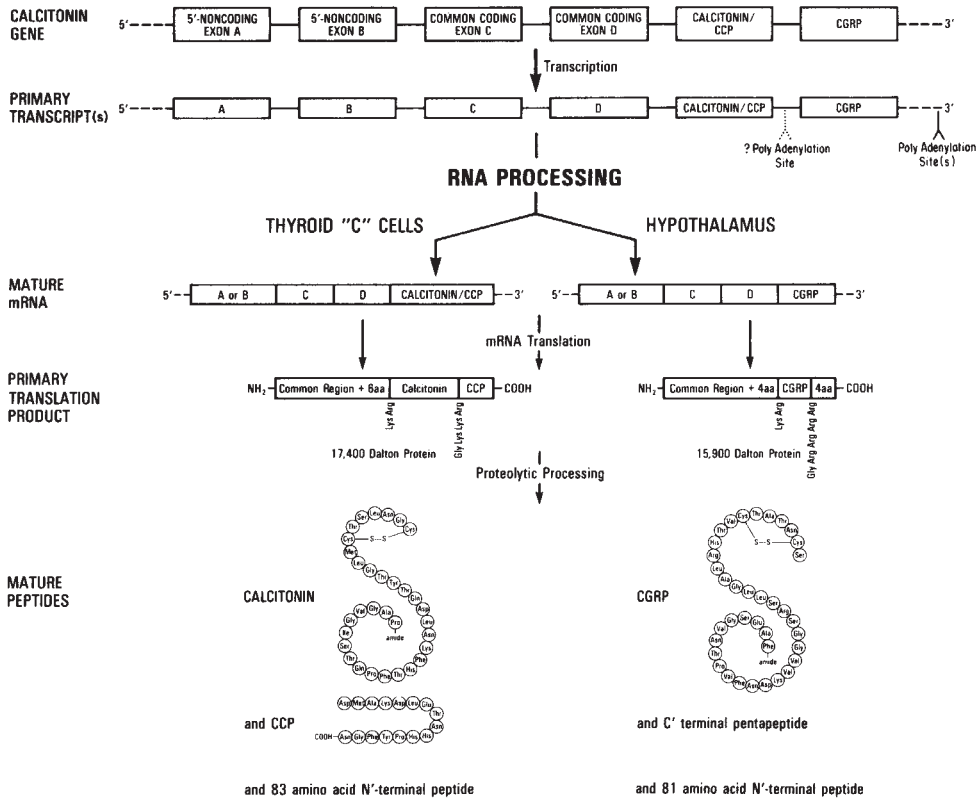

Fig. 5 A model to describe the tissue-specific expression of the calcitonin gene during 'peptide switching' events. The 5′ and 3′ termini are indicated as dotted lines because their structural organization remains unknown. The structural organization of coding blocks A and B, and the presence of A within CGRP mRNAs are not unequivocally established. Possible sites of polyadenylation of the primary transcript(s) are indicated. The forked arrow indicates tissues in which pCal reactive RNAs predominate (thyroid) or pCGRP reactive RNAs predominate (hypothalamus). The predicted proteolytic products of the primary translation product directed by Cal mRNA are established, those of CGRP mRNA, predicted.

Concluding remarks

The discovery of 'split genes' and definition of hnRNA structure[5,6,17–21] raised the possibility of important regulatory events at the level of processing of initial RNA transcripts. As presented here, such regulation of RNA processing events may be necessary for the expression of certain neuroendocrine genes and may increase the diversity of the resultant peptide products. Genomic mapping data are consistent with the existence of a single calcitonin gene[8]; if a second gene exists it would have to be extremely similar in size and organization. Sequence analysis establishes that Cal and CGRP domains are linked in a single genomic clone whose exons correspond precisely in sequence to the cloned mRNAs. Based on the structure of this gene and its RNA products, we propose a model in which genomic regions can represent discrete hormone encoding domains whose ultimate expression is dependent on differential RNA processing events (Fig. 5). Thus, these domains may be included or excluded during the maturation process to generate multiple mRNAs encoding alternative hormones or neuropeptides. We refer to the consequences of these RNA processing events as 'peptide switching' because the splicing of alternative exons results in the expression of different peptides.

Several explanations could account for the alternative expression of Cal and CGRP mRNAs. First, there could be two overlapping transcription units with different cap and poly(A) sites. Second, there could be a single cap site with two or more poly(A) sites. Multiple RNAs would arise from a choice of poly(A) site either during transcription or due to post-transcriptional cleavage and re-polyadenylation. Third, there could be a small rearrangement or modification of gene structure affecting a splice site or polyadenylation signal. Finally, there could be true regulation of splicing choice for a single primary transcript.

The last possibility would parallel events observed in SV40 and adenovirus gene expression, where several mRNAs can be generated with common 5′ and 3′ termini and yet encode different proteins[5,6]. This process requires the differential use of multiple and optional 5′ and 3′ splicing sites. It has been suggested that the L-1 transcription unit of adenovirus displays splice-choice control[22]. Should such an analogy apply to expression of calcitonin it remains to be defined how such a process is regulated. Splicing regulation could arise from changes in enzyme specificity, small nuclear RNAs and/or other factors directly affecting the splicing process and could account for the accumulation of the larger nuclear RNA transcripts of the calcitonin gene during switching to production of CGRP mRNAs[7].

Certain observations regarding the expression of the calcitonin gene must be accounted for in any model of alternative processing. Both the calcitonin genomic clone and the larger nuclear RNA transcripts contain sequences specific to both the mature calcitonin and CGRP mRNAs. If the observed nuclear transcripts are in fact precursors of both mRNAs, then the expression of either Cal or CGRP mRNAs requires splicing out of an internal exon. This would be in contrast to the RNA processing events proposed in the case of immunoglobulin heavy chain gene expression, where a polyadenylation site intercedes between the exons involved in switching[13–15]. On the other hand, if a polyadenylation site is present between Cal and CGRP exons, the precursors of the two mRNAs will be distinct. In this case, the nuclear CGRP-reactive RNAs in the thyroid would represent species not destined to become either RNA or would require a cleavage and re-polyadenylation to be processed into calcitonin mRNA.

Although it has been proposed that, in the case of liver and salivary gland α-amylase and dihydrofolate reductase gene expression[23], differential 5′ transcription initiation and/or alternative polyadenylation sites[24,25] result in mRNA structural polymorphism, this does not affect the structure of the protein-encoding regions of the mature mRNAs. In fact, 5′ heterogeneity also seems to occur in calcitonin mRNAs, where one splice junction precedes by nine nucleotides the initiator methionine codon (Fig. 2*a*). The location of this intervening sequence corresponds precisely with the 3′ boundary of a region of 5′ noncoding sequence in pCal that diverges from a sequence reported for another calcitonin clone[26]. The use of 'optional' 5′ exons, in addition to the coding exon switch discussed here, further increases the mRNA polymorphism associated with calcitonin gene expression, but does not alter the coding region of the resultant mRNAs.

The apparent tissue specificity of the observed splicing events implies developmental regulation of these processes. Calcitonin gene expression therefore represents a model system for exploring mechanisms responsible for both transcriptional and post-transcriptional regulation of specific gene expression during development. Whatever the responsible molecular mechanisms, the unexpected consequence of these RNA processing events is that the same gene which generates the hormone calcitonin in thyroid C-cells is apparently responsible for the expression of a new polypeptide product in the hypothalamus.

We thank Drs James E. Darnell, Francis Crick, Geoffrey Wahl, Jean-Jacques Mermod and Geoffrey Murdoch for advice and helpful discussions regarding the manuscript, Arnold Loera and Rodrigo Franco for providing experimental data, our collaborators Drs Bernard A. Roos and Roger Birnbaum for their continued support in these studies, Drs Tom Sargent and James Bonner for providing the rat liver DNA library, Maureen Brennan and Margaret Richards for preparation of the manuscript. This work was supported by grants from the American Cancer Society and the NIH.

Received 9 February; accepted 4 May 1982.

1. Blundell, T. L. & Humbel, R. E. *Nature* **287,** 781–787 (1980).
2. Roberts, J. L., Phillips, M., Rosa, P. A. Herbert, E. *Biochemistry* **17,** 3609–3618 (1978).
3. Eipper, B. A. & Mains, R. E. *J. biol. Chem.* **253,** 5732–5744 (1978).
4. Nakanishi, S. *et al. Nature* **278,** 423–427 (1979).
5. Ziff, E. B. *Nature* **287,** 491–499 (1980).
6. Darnell, J. E. *Prog. Nucleic Acid Res. molec Biol.* **22,** 327–353 (1978).
7. Rosenfeld, M. G., Amara, S. G., Roos, B. A., Ong, E. S. & Evans, R. M. *Nature* **290,** 63–65 (1980).
8. Rosenfeld, M. G. *et al. Proc. natn. Acad. Sci. U.S.A.* **79,** 1717–1721 (1982).
9. Amara, S. G., David, D. N., Rosenfeld, M. G., Roos, B. A. & Evans, R. M. *Proc. natn. Acad. Sci. U.S.A.* **77,** 4444–4448 (1980).
10. Amara, S. G. *et al. J. biol. Chem.* **257,** 2129–2132 (1982).
11. Birnbaum, R. S., O'Neal, J. A., Muszynski, M., Aron, D. C. & Roos, B. A. *J. biol. Chem.* **257,** 241–244 (1982).
12. Breathnach, F., Benoist, C., O'Hare, K., Gannon, F. & Chambon, P. *Proc. natn. Acad. Sci. U.S.A.* **75,** 4853–4857 (1978).
13. Alt, F. W. *et al. Cell* **20,** 293–301 (1980).
14. Early, P. *et al. Cell* **20,** 313–319 (1980).
15. Maki, R. *et al. Cell* **24,** 353–365 (1981).
16. Kreil, G., Suchanek, G. & Kindas-Mugge, I. *Fedn Proc.* **36,** 2081–2086 (1977).
17. Chow, L. T., Gelinas, R. E., Broker, T. R. & Roberts, R. J. *Cell* **12,** 1–8 (1977).
18. Klessig, D. F. *Cell* **12,** 9–21 (1977).
19. Sambrook, J. *Nature* **268,** 101–104 (1977).
20. Jeffreys, A. J. & Flavell, R. A. *Cell* **12,** 1097–1108 (1977).
21. Tilghman, S. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **75,** 725–729 (1978).
22. Nevins, J. R. *Cell* **28,** 1–2 (1982).
23. Tosi, M., Young, R. A., Hagenbuchle, O. & Schibler, U. *Nucleic Acids Res.* **9,** 2313–2323 (1981).
24. Setzer, D. R., McGrogan, M., Nunberg, J. H. & Schimke, R. T. *Cell* **22,** 361–370 (1980).
25. Young, R. A., Hagenbuchle, O. & Schibler, U. *Cell* **23,** 451–458 (1981).
26. Jacobs, J. W. *et al. Science* **213,** 457–459 (1981).
27. Southern, E. M. *J. molec. Biol.* **98,** 503–517 (1975).
28. Maxam, A. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74,** 560–564 (1977).
29. Smith, H. O. & Birnstiel, M. L. *Nucleic Acids Res.* **3,** 2387–2398 (1976).
30. Messing, J., Crea, R. & Seeburg, P. M. *Nucleic Acids Res.* **9,** 309–321 (1981).
31. Sanger, F., Nicklen, S. & Coulsen, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74,** 5463–5467 (1977).
32. Fuhrman, S. A., Deininger, P. L., Laporte, P., Friedmann, T. & Geiduschek, E. P. *Nucleic Acids Res.* **9,** 6439–6456 (1981).
33. Potter, E., Nicolaisen, A. K., Ong, E. S., Evans, R. M. & Rosenfeld, M. G. *Proc. natn. Acad. Sci. U.S.A.* **78,** 6662–6666 (1981).
34. Wahl, G. M., Stern, M. & Stark, G. R. *Proc. natn. Acad. Sci. U.S.A.* **76,** 3683–3687 (1979).