Structure and expression of the human calcitonin/CGRP genes

P.H. Steenbergh[+°], J.W.M. Höppener[+], J. Zandberg[+°], A. Visser[+°], C.J.M. Lips and H.S. Jansz[+°]

[+]*Institute of Molecular Biology, Padualaan 8, 3508 TB Utrecht,* [°]*Laboratory for Physiological Chemistry* and *Department of Internal Medicine of the University Hospital, State University of Utrecht, Utrecht, The Netherlands*

Received 8 September 1986; revised version received 14 October 1986

Recently, we have reported the isolation of cDNA encoding a second human calcitonin gene-related peptide (hCGRP-II) [(1985) FEBS Lett. 183, 403–407]. In this report we describe the isolation and characterization of the gene encoding hCGRP-II. This gene, designated CALC-II, is structurally closely related to the known CALC-I gene encoding human calcitonin (hCT) and hCGRP-I. In constrast to CALC-I, CALC-II does not seem to be alternatively expressed. The formation of a second, hCT-like mRNA by differential splicing of CALC-II transcripts is unlikely in view of the structure of CALC-II, and could not be demonstrated in tissues known to express CALC-I and CALC-II.

Calcitonin *CGRP* *Gene family* *Alternative splicing* *(Human)*

1. INTRODUCTION

The rat calcitonin (CT) gene is one of the first mammalian genes which have been shown to generate two mRNAs encoding different biologically active peptides as a consequence of alternative RNA processing events. The products of these mRNAs are the precursor to the calcium-regulating hormone calcitonin in the thyroidal C-cell, and the precursor to the calcitonin gene-related peptide (CGRP) in nervous tissues [1].

Analysis of the structure of the human calcitonin (hCT) gene by our own and other laboratories has demonstrated that the same mechanism of alternative RNA processing leads to the synthesis of hCTmRNA (exons 1–4) and hCGRPmRNA (exons 1–3,5 and 6) [2–6]. During these studies we isolated, from a cDNA library of human medullary thyroid carcinoma (MTC) mRNA, a clone which was clearly derived from transcription of a second hCT/CGRP gene [7]. Nucleotide sequence analysis of this cDNA clone revealed that it contained regions with greater than 90% homology to part of exon 3 and the entire exon 5, and about 65% homology to exon 6 of the first hCT/CGRP gene. The cDNA sequence predicted the existence of a second hCGRP (hCGRP-II), differing from the known hCGRP (hCGRP-I) in 3 of its 37 amino acids.

Southern-blot analysis of total human DNA hybridized to exon 3 and exon 5 sequences of the first hCT/CGRP gene as probes confirmed the existence of the second gene [7]. The second hCT/hCGRP gene has been shown to be a single copy gene located on chromosome 11, the same chromosome that carries CALC-I [8,9].

In accordance with the report of the committee on the genetic constitution of human chromosomes 10, 11, and 12 [10], the first and second hCT/ CGRP genes are in the following referred to as CALC-I and CALC-II.

This report presents the structure of CALC-II and compares it to the structure of CALC-I. The complete structure of the hCGRP-II precursor polypeptide molecule is predicted from the nucleotide sequencing results. The possibility of a second mature mRNA generated from CALC-II is assessed, both theoretically on the basis of the gene

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/86/$3.50

structure, and empirically by screening of a variety of tissues for the presence of this RNA by Northern blot analysis.

2. MATERIALS AND METHODS

2.1. *Cosmid libraries*

Using the pJB8 cosmid vector cloning system, representative human gene libraries were constructed with high molecular mass DNA from human medullary thyroid carcinoma (MTC) and human acute lymphatic leukaemia cells (ALL) after partial digestion with *Sau3A* [11]. Following in vitro packaging, transduction to *E. coli* 1046 and selection on ampicillin containing medium, approx. 100 000 colonies were assayed for CALC-II sequences.

2.2. *Subcloning*

Subclones of cosmid fragments in pBR322 and M13 vectors, after digestion with the appropriate restriction enzymes, were obtained by ligation using T_4 ligase. *E. coli* K12 1592 cells were transformed with the resulting chimaeric plasmids. Transformants were grown on nitrocellulose filters. Replica filters were screened by hybridization to sequence specific DNA probes.

2.3. *DNA probes*

Total plasmid or ds M13 DNA containing specific inserts, or isolated restriction enzyme fragments, were labeled by nick-translation to a specific activity of $1\text{–}5 \times 10^8$ dpm/μg and used as hybridization probes after heat denaturation. Synthetic oligonucleotides, with free 3′- and 5′-terminal hydroxyl groups [12] were radiolabeled to the same specific activity with [γ-^{32}P]ATP and T_4 polynucleotide kinase.

2.4. *Nucleotide sequence analysis*

ds DNA restriction enzyme fragments were labeled 3′- or 5′-terminally using [α-^{32}P]ddATP and terminal deoxynucleotidyl transferase, or [γ-^{32}P]ATP and T_4 polynucleotide kinase, respectively. After a second restriction enzyme digestion and electrophoretic separation, sequences of fragments labeled at only one of their 3′- or 5′-termini were analyzed using the chemical modification technique [13]. Fragments subcloned into M13 vectors (mp8, mp9, mp10, and mp11) were analyzed using the dideoxy method [14].

2.5. *Northern blots*

Total cellular RNA was isolated from cultured cells and tissues by the guanidine thiocyanate procedure [15]. Poly(A)-rich RNA was prepared by oligo(dT)-cellulose chromatography. RNA samples were treated with glyoxal and DMSO, size fractionated on agarose slabgel and transferred to Gene-Screen membranes (New England Nuclear,

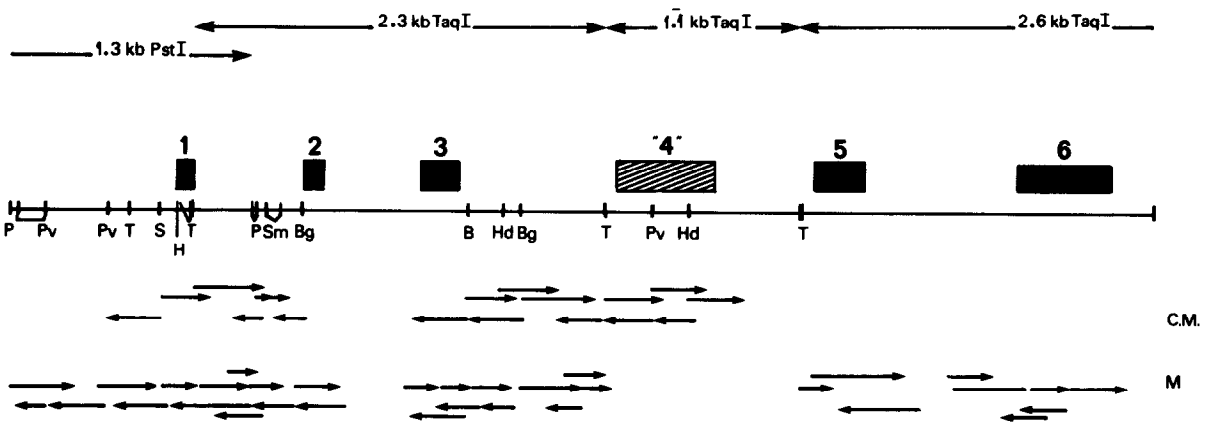

Fig.1. Schematic representation of CALC-II. The exons are indicated by the black boxes, the region corresponding to exon 4 of CALC-I by the hatched box. Restriction enzyme fragments contained in subclones used for further structural analysis are indicated at the top. Other restriction enzyme recognition sites relevant to sequence analysis experiments are indicated on the line below the exons (B, *Bam*HI; Bg, *Bgl*II; H, *Hin*cII; Hd, *Hin*dIII; P, *Pst*I; Pv, *Pvu*II; S, *Sac*I; Sm, *Sma*I; T, *Taq*I). Sequences derived by the chemical modification (C.M.) or M13 (M) dideoxy sequencing techniques are indicated by arrows.

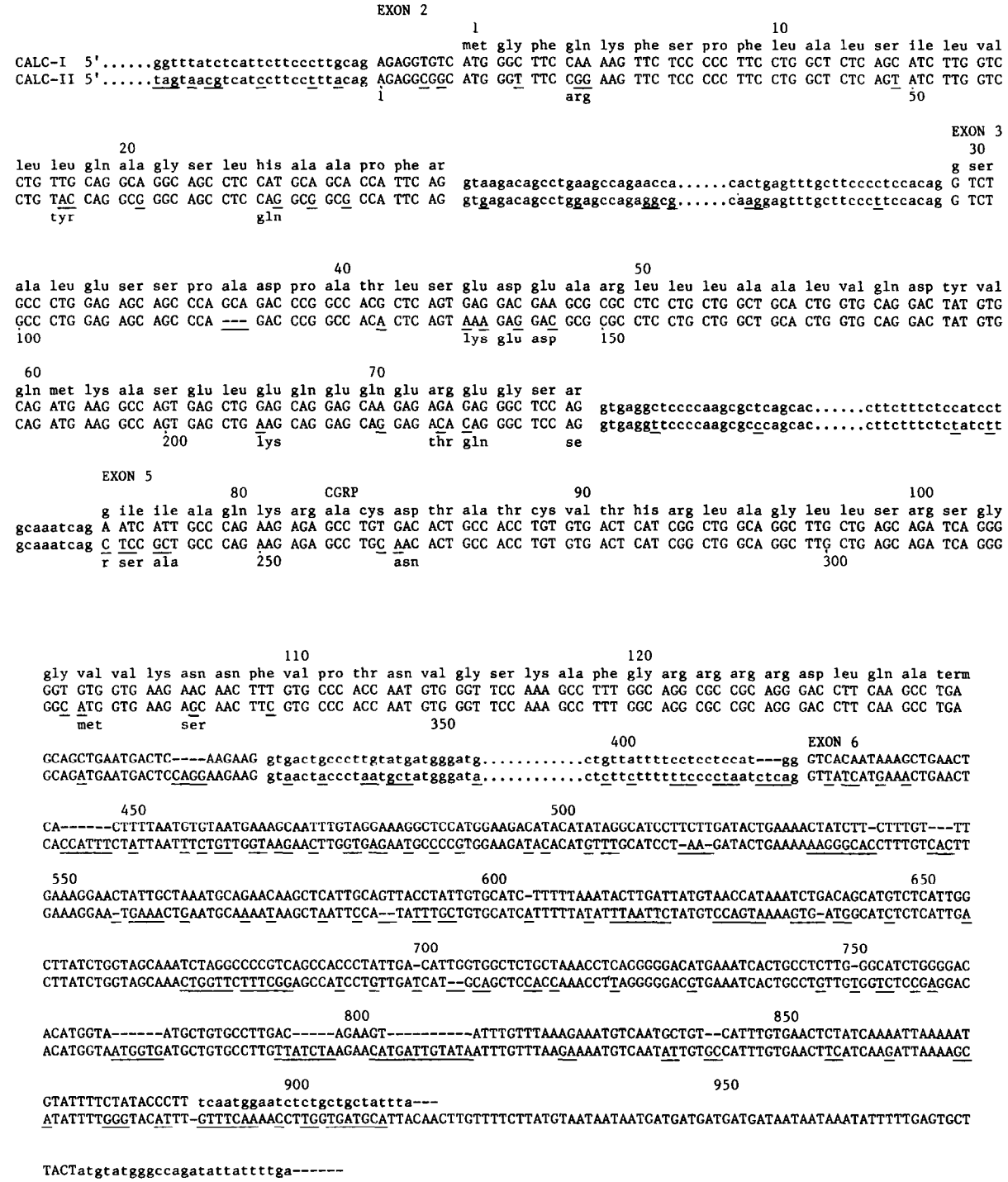

Fig.2. Comparison of the nucleotide sequences of exons 2, 3, 5 and 6 of CALC-I with the equivalent regions of CALC-II. Gaps were introduced to maximize homologies. Base substitutions in CALC-II are underlined. The amino acid sequence of the hCGRP-I precursor is indicated; substitutions in the hCGRP-II precursor are indicated underneath.

Boston, USA). The membranes were pre-hybridized, hybridized to radiolabeled probe and washed as prescribed by the manufacturer. When oligonucleotides were used as hybridization probes, the hybridization temperature was lowered from 42 to 32°C, and the washing procedure was carried out at 30 rather than 65°C.

3. RESULTS

The insert of cDNA clone pCGRP-4 [7] was used as probe in the screening of cosmid libraries of human DNA for CALC-II specific clones. The 3′-terminal *Sac*I-*Pst*I part of this insert, containing sequences partially homologous to exons 5 and 6 of CALC-I, should detect a 2.6 kb *Taq*I fragment within CALC-II; the 5′-terminal *Pst*I-*Sac*I part, containing sequences partially homologous to exon 3 of CALC-I, should detect a 2.3 kb *Taq*I fragment derived from CALC-II [7]. Cos1CALC-II, isolated from the cosmid library of human DNA from MTC tissue, was shown to contain the 2.6 kb *Taq*I fragment, but not the 2.3 kb *Taq*I fragment, indicating that only part of CALC-II is present within this clone. Cos2CALC-II, containing the entire CALC-II gene, was isolated from the library constructed with DNA from acute lymphatic leukaemia cells. The 2.6 kb *Taq*I fragment was subcloned from Cos1CALC-II. 2.3 kb and 1.1 kb *Taq*I fragments and a 1.3 kb *Pst*I fragment were subcloned from cos2CALC-II (fig.1).

Nucleotide sequences within these subcloned fragments were analyzed using the chemical modification technique, and by dideoxy sequencing after further subcloning into M13 vectors (fig.1). These analyses revealed the presence within CALC-II of regions with about 90% homology to exons 2, 3 and 5 of CALC-I. The sequence of the non-coding exon 6 of CALC-I is 65% homologous to the corresponding region of CALC-II. From these sequences, the structure of the complete precursor polypeptide for hCGRP-II can be predicted (fig.2).

Between exons 3 and 5 of CALC-II, a region with about 50% homology to exon 4 of CALC-I (the hCT encoding exon) is present (fig.3). The 1.1 kb *Taq*I fragment encompassing this region was used as probe in the screening of human tissues for

```
                                 EXON 4

                                                                             calcitonin
                                leu  asp  ser  pro  arg  ser  lys  arg       cys  gly  asn  leu  ser  thr  cys
CALC-I     5'-----ttccc-tgcag C CTG  GA-C AGC  CCC  AGA  TCT  AAG  CGG  --   TGC  GGT  AAT  CTG  AGT  ACT  TGC
CALC-II    5'-----ttcccttgcag C -TA  GAGC AGT  CCT  AGA  TTT  AAG  TAG  CA   TAT  AGT  AAT  CTG  AGT  ACC  TGC
                                                                             tyr  ser

met  leu  gly  thr  tyr  thr  gln  asp  phe  asn  lys  phe  his  thr  phe  pro  gln  thr  ala  ile  gly  val
ATG  CTG  GGC  ACA  TAC  ACG  CAG  GAC  TTC  AAC  AAG  TTT  CAC  ACG  TTC  CCC  CAA  ACT  GCA  ATT  GGG  GTT
TTG  CAG  GGC  ACA  TAC  TTG  CAG  TAC  CTG  AAA  AAC  TTT  CAT  ATG  TTC  CCT  GGC  ATC  AAC  TTC  GGG  CCT
leu  gln                 leu       tyr  leu  lys  asn            met            gly  ile  asn  phe       pro

gly  ala  pro  gly  lys  lys  arg  asp  met  ser  ser  asp  leu  glu  arg  asp  his  arg  pro  his  val  ser
GGA  GCA  CCT  GGA  AAG  AAA  AGG  GAT  ATG  TCC  AGC  GAC  TTG  CAG  AGA  GAC  CAT  CGC  CCT  CAT  GTT  AGC
GAA  ATT  CCT  GGC  AAG  AAT  AGG  GAC  ATA  GTC  AAC  AGC  TTG  CAG  AGG  GAC  CAC  TAC  CCG  ACT  CCA  TGG
glu  ile                 asn            ile  val  asn  ser       gln                 tyr       thr  pro  trp

met  pro  gln  asn  ala  asn  stop                               stop
ATG  CCC  CAG  AAT  GCC  AAC  TAA  ACT  CCT  CCC  TTT  CCT  TCC  TAA  TTTCCCTTCTTGCATCC--TTCCTATAACTTGATGCAT
TCC  CCC  GGG  TGG  CAG  CTG  AAC  TTC  TCT  CAA  CTC  TCC  ---  TGA  TTCCCCTTCTTGC-TCCACTT-TATGAACCTGATGCAT
ser       gly  trp  gln  leu  asn  phe  ser  gln  leu  ser       stop

GTGGTTTGGTTCCTCTCTGGTGGCTCTTTGGGCTGGTATTGGTGGCTTTCCTTGTGGCAGAGGATGTCTCAAACTTCAGATGGGAGGAAAGAGAGCAGGACTCACAGG
GTGGATT----CCTCTCTGATTTGTCTTCATGCTGGTATTGGTATTTTTGCTTATGACAGAGAATGTTTTGAAGACCTCAGGATGGAAGGGAAGACAG------CAGG

--TTGGAAGAGAATCACCTGGGAAAATACCAGAAAATGAGGGCCGCTTTGAGTCCCCCAGAGA---TGTCATCAGAGCTCCTCTGTCCTGCTTCTGAATGTG-CTGAT
ACTT---ACTGAA-CACGTTAGAGA-TAAAAGAAAATAAGGGAAGCTTCTTGAGACTGT-AGAGGGTGTTATGACAGAG--GC-ATCCAATTTCTGCTTCTAAATG-T

CATTTGA-GGAATAAAATTATTTTTCCCCAAAGA tctgagctgtggtgg-tcattgctct...........
ACTACGATAAAATAAGCACGTCCTT----AATGC cttggcgttagatgaatcaat-ctat...........
```

Fig.3. Nucleotide sequence of the exon 4-like region in CALC-II compared to exon 4 of CALC-I. Gaps were introduced to maximize homology. The amino acids encoded by the CALC-I sequence are indicated; hCT is underlined. The amino acids which theoretically might be encoded by this part of CALC-II and which differ from those encoded by CALC-I are indicated underneath.

alternative expression of CALC-II, as has been established for CALC-I. Total cellular RNA (50 μg) and poly(A)-rich RNA (20–50 μg) from human MTCs (sporadic and familial), pituitary, hypothalamus, thyroid, phaeochromocytoma and lung tumours of various types was electrophoresed, Northern-blotted and hybridized to this probe. In none of these tissues could mRNA containing sequences derived from the exon 4-like region of CALC-II be demonstrated.

A region of 111 nucleotides with about 50% homology to the sequence directly preceding exon 1 and part of exon 1 of CALC-I is located within the 1.3 kb *Pst*I fragment of CALC-II (fig.1) at approximately the same position as is exon 1 in CALC-I. This region is presented in fig.4. Subsequently, it has been shown that transcripts of this region of CALC-II are present in hCGRP-II mRNA, which was isolated from Ewing sarcoma cells. These cells have been shown to express CALC- II, but not CALC-I, using probes specific for either one of these genes (manuscript in preparation). The 20-mer synthetic oligonucleotide 5′-GATGCTGAGAGCCAGGAAGG-3′, complementary to part of exon 2 of CALC-I, and, with one mismatch, to nucleotides 33 to 52 of exon 2 of CALC-II (fig.2) was labeled 5′-terminally, hybridized to poly(A)-rich RNA from Ewing sarcoma cells and extended using reverse transcriptase. The nucleotide sequence of the products revealed that exon 2 of CALC-II is linked to the sequence in fig.4 as indicated, and allowed unambiguous confirmation of the sequence up to position 70 in fig.4.

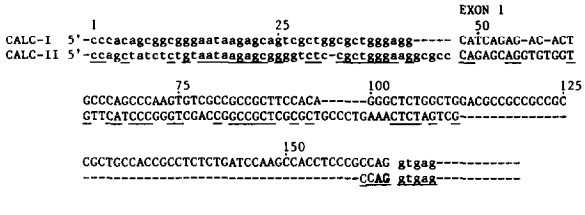

Fig.4. Comparison of nucleotide sequences of exon 1 of CALC-I and a region at the equivalent position in CALC-II. Gaps have been introduced to maximize homology. Identical nucleotides in both sequences are underlined.

4. DISCUSSION

We have previously reported the nucleotide sequence of cDNA corresponding to hCGRP-II mRNA isolated from human MTC [7], hCGRP-II has since been proven to have cardiovascular activity comparable in its effects and potency to hCGRP-I (I. Marshall et al., in preparation; A. Franco-Cereceda et al., in preparation).

In the present report we show that CALC-II has a striking resemblance in overall structure to the known CALC-I gene. From the sequence of exons 2, 3 and 5 of CALC-II the structure of the precursor polypeptide for hCGRP-II can be predicted (fig.2). The sequence contains an open reading frame of 381 nucleotides, coding for a precursor protein of 127 amino acids. The first 237 nucleotides specify the 79 amino acid NH_2-terminal peptide. Comparison to the equivalent region in hCGRP-I mRNA reveals 27 base substitutions, insertions, or deletions, altering 13 of the NH_2-terminal amino acids. The next 120 nucleotides encode hCGRP-II, and a COOH-terminal tetrapeptide. Within this region only 6 base substitutions occur, causing 3 of the 37 amino acids of hCGRP-II to differ from hCGRP-I.

The nucleotide sequence further predicts that hCGRP-II, like hCGRP-I, is excised from its protein precursor by proteolytic cleavage at pairs of basic amino acids flanking hCGRP-II and that hCGRP-II is COOH-terminally amidated. The COOH-terminal tetrapeptide resulting from cleavage of the hCGRP-II precursor is identical to that of the hCGRP-I precursor.

The nucleotide sequences of the cDNA clone [7] and the genomic clone differ at four points. One of these differences (G is A in cDNA) is located within the coding region of the NH_2-terminal peptide, at nucleotide 229 in fig.2. As a consequence, the *Sac*I recognition site (5′-GAGCTC-3′) present in the cDNA clone is absent in the genomic clone, and the serine residue predicted by the cDNA is changed to a glycine residue by the genomic sequence. The remaining three differences are located within the non-coding exon 6 (see below).

It has been suggested that CT and CGRP exons derive from a common primordial gene, and that the CT/CGRP gene arose by duplication and sequence divergence events [6]. The present data indicate that biological diversity was increased

further by gene duplication, resulting in CALC-I and CALC-II. Since also in rat a second CGRP has been found [16], it is probable that gene duplication occurred prior to the rodent-primate split 70 million years ago. Fig.5 compares the amino acid sequences of the precursor polypeptides of CGRPs in rat and human. Identical amino acids are found at 78 positions in all four precursors; of these 78 positions, 43 are located within the 49 COOH-terminal residues, which constitute the CGRPs and the COOH-terminal peptides. The number of differences between the human NH_2-terminal peptides (13 out of 80 amino acids) is strikingly lower than that between the rat NH_2-terminal peptides (41 out of 88 amino acids). This might reflect a difference in biological function of the NH_2-terminal peptides in rat and human. The two additional potential cleavage sites (Arg-Lys) in the rat β-CGRP precursor [16] are absent from the hCGRP-II precursor.

In hCGRP-I mRNA, transcripts of two non-coding exons, exons 1 and 6, are contained. Within CALC-II, a region 65% homologous to exon 6 of CALC-I is present. As shown by the cDNA data [7], exon 6 of CALC-II is at least 551 nucleotides long, as opposed to 431 nucleotides in CALC-I. The AATAAA sequence (nucleotides 957–962 in fig.2) is most likely to function as poly(A)-addition enzyme recognition signal, and polyadenylation probably takes place at the T-residue at position 980.

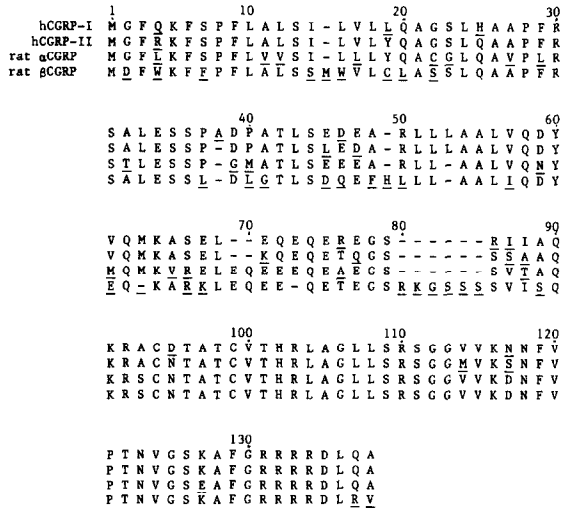

Fig.5. Amino acid sequences of the precursor polyepeptides of the human and rat CGRPs. Amino acids differing from the consensus residue at that point are underlined. Gaps are included in order to obtain maximal homology.

Three differences in the nucleotide sequence of exon 6 were observed between the CALC-II gene and the cDNA sequence reported earlier [7]. These differences are located at nucleotides 450 (T is C in cDNA), 489 (G is C in cDNA, resulting in the presence of a *Bst*NI site in the cDNA at this point), and 698 (G is A in cDNA) (fig.2). These differences may be the result of reverse-transcription errors and/or naturally occurring nucleotide sequence polymorphisms.

The sequence presented in fig.4 shows the highest degree of homology to exon 1 of CALC-I within the region examined. Its position and the fact that the putative RNA polymerase recognition signal in CALC-I (nucleotides 12–22 in fig.4) is well conserved in CALC-II make it likely that this region functions as exon 1 in CALC-II.

In view of the fact that CALC-II has so many features in common with CALC-I, the question arises whether transcripts of CALC-II may be processed in such a way that an mRNA equivalent to hCT mRNA is formed. Examination of the exon 4-like region of CALC-II (fig.3) makes this unlikely. Splicing at the site equivalent to the exon 3-exon 4 junction in hCT mRNA would result in stop-codons within the reading frame of the precursor polypeptide. Furthermore, the cysteine residue at position 1 of all CTs studied so far is not encoded by this sequence, and the amino-terminal Lys-Arg excision signal is lacking.

In conclusion, CALC-II seems to be a pseudogene for hCT, but a structural gene for hCGRP. Since antibodies raised to hCGRP-I are bound to cross-react with hCGRP-II, it is unclear whether hCGRP-I or -II has been detected by immunochemical methods in human tissues. The presence of immunoreactive hCGRP may be the result of expression of CALC-II and not of alternate expression of CALC-I. A genomic clone, λhCa12, containing a second human calcitonin gene has also been isolated by Jonas et al. [6]. Restriction analysis and partial sequence analysis was stated to indicate that 'λhCa12 contains regions with homology to the common region, calcitonin and CGRP exons, but that this gene encodes neither the calcitonin nor CGRP mRNA' (i.e. hCGRP-I

mRNA). Since according to our data, the human genome does not contain regions hybridizing to CGRP specific probes other than those in CALC-I and CALC-II, it seems probable that λhCa12 contains CALC-II. If so, hybridization under stringent conditions of a calcitonin-specific rat cDNA probe to λhCALC2 [6] remains to be explained, unless this probe also contained 'common region' derived sequences.

ACKNOWLEDGEMENTS

The authors wish to thank Dr W.J.M. van de Ven (University of Nijmegen) for making the cosmid library of human ALL-DNA available to them, Dr J.H. van Boom (University of Leiden) for the synthesis of oligonucleotides, and Mrs E. den Aantrekker for assistance in the preparation of the manuscript. This work was supported by the Ciba-Geigy Company and by the Foundation for Chemical Research (SON) with financial aid from the Netherlands Organisation for the Advancement of Pure Research (ZWO).

REFERENCES

[1] Amara, S.G., Evans, R.M. and Rosenfeld, M.G. (1984) Mol. Cell Biol. 4, 2151–2160.
[2] Steenbergh, P.H., Höppener, J.W.M., Zandberg, J., Van der Ven, W.J.M., Jansz, H.S. and Lips, C.J.M. (1984) J. Clin. Endocrinol. Metab. 59, 358–360.
[3] Steenbergh, P.H., Höppener, J.W.M., Zandberg, J., Cremers, A.F.M., Jansz, H.S. and Lips, C.J.M. (1985) in: Calcitonin (Pecile, A. ed.) Elsevier Science Publishers, Amsterdam, New York.
[4] Le Moullec, J.M., Jullienne, A., Chenais, J., Lasmoles, F., Guliana, J.M., Milhaud, G. and Moukthar, M.S. (1984) FEBS Lett. 167, 93–97.
[5] Edbrooke, M.R., Parker, D., McVey, J.H., Riley, J.H., Sorenson, G.D., Pettengill, O.S. and Craig, R.K. (1985) EMBO J. 4, 715–724.
[6] Jonas, V., Lin, C.R., Kawashima, E., Semon, D., Swanson, L.W., Mermod, J.-J., Evans, R.M. and Rosenfeld, M.G. (1985) Proc. Natl. Acad. Sci. USA 82, 1994–1998.
[7] Steenbergh, P.H., Höppener, J.W.M., Zandberg, J., Lips, C.J.M. and Jansz, H.S. (1985) FEBS Lett. 183, 403–407.
[8] Höppener, J.W.M., Steenbergh, P.H., Zandberg, J., Bakker, E., Pearson, P.L., Geurts van Kessel, A.H.M., Jansz, H.S. and Lips, C.J.M. (1984) Hum. Genet. 66, 309–312.
[9] Höppener, J.W.M., Steenbergh, P.H., Zandberg, J., Geurts van Kessel, A.H.M., Baylin, S.B., Nelkin, B.D., Jansz, H.S. and Lips, C.J.M. (1985) Hum. Genet. 70, 259–263.
[10] Grzeschik, K.-H. and Kazazian, N.H. (1985) Cytogenet. Cell Genet. 40, 179–205.
[11] Ish-Horowic, D. and Burke, J.F. (1981) Nucleic Acids Res. 9, 2989–2998.
[12] Marugg, J.E., McLaughlin, L.W., Piel, N., Tromp, M., Van der Marel, G.A. and Van Boven, J.M. (1983) Tetrahedron Lett. 24, 3989–3992.
[13] Maxam, A.M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560–564.
[14] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[15] Chirgwin, J.M., Przybala, A.E., MacDonald, R.J. and Rutter, W.J. (1979) Biochemistry 18, 5294–5299.
[16] Amara, S.G., Arriza, J.L., Leff, S.E., Swanson, L.W., Evans, R.M. and Rosenfeld, M.G. (1985) Science 229, 1094–1097.