Volume 233, number 1, 57–63 FEB 05899 June 1988

A third human CALC (pseudo)gene on chromosome 11

J.W.M. Höppener*, P.H. Steenbergh+, J. Zandberg+, G.J. Adema+, A.H.M. Geurts van Kessel, C.J.M. Lips*° and H.S. Jansz*+

*Institute of Molecular Biology, State University of Utrecht, Padualaan 8, 3584 CH Utrecht, +Laboratory for Physiological Chemistry, State University of Utrecht, Vondellaan 24, 3521 GG Utrecht, °Department of Internal Medicine, University Hospital, State University of Utrecht, Catharijnesingel 101, 3511 GV Utrecht and Department of Cell Biology and Genetics, Erasmus University, PB 1738, 3000 DR Rotterdam, The Netherlands

Received 30 March 1988

A genomic locus in man (CALC-III) containing nucleotide sequences highly homologous to both exon 2 and exon 3 of the CALC-I and -II genes, is described in this paper. The CALC-I gene produces calcitonin (CT) (encoded by exon 4) or calcitonin gene-related peptide (CGRP) (encoded by exon 5) in a tissue-specific fashion. The CALC-II gene produces a second human CGRP, but probably not a second CT. The CALC-III gene does not seem to encode a CT- or CGRP-related polypeptide hormone and is probably a pseudogene. Like the other two CALC genes, the CALC-III gene is located on human chromosome 11.

Calcitonin; Gene duplication; CALC-III (pseudo)gene; Chromosomal localization; (Man)

1. INTRODUCTION

Analysis of rat calcitonin (CT) gene expression has provided the first example of tissue-specific regulation of RNA processing pathways in the endocrine and nervous systems, determining which of two polypeptide hormone precursors is produced [1,2]. In the thyroid gland CT mRNA is almost exclusively produced, whereas in most neuronal tissues mRNA encoding the precursor of calcitonin gene-related peptide (CGRP) is almost exclusively produced from the rat CT/CGRP gene. Exons 1–3 of this gene are represented in both CT and CGRP mRNA. In addition, CT mRNA contains exon 4 (CT-encoding) and CGRP mRNA contains exon 5 (CGRP-encoding) and exon 6 sequences.

Structural analysis of the human CT gene has revealed a similar organization and mode of expression to that of the corresponding rat gene [3,4]. In medullary thyroid carcinoma (MTC; a tumour of the CT-producing thyroidal C-cells), both CT mRNA and CGRP mRNA are produced from the human CT/CGRP gene (or 'CALC-I gene', [5]). In addition, a structurally related gene has been identified in the human genome, encoding a second CGRP (CGRP-II) differing from the previously known human CGRP (CGRP-I) in 3 of the 37 amino acids. This second human CT/CGRP gene (or 'CALC-II gene', [5]) is probably a pseudogene for a second human CT [6,7]. CGRP-II mRNA has been detected in MTC [8] and in Ewing sarcoma cell lines [9]. A gene encoding a second CGRP (β-CGRP) has also been identified in rat, but as in man a second CT-encoding exon has not been reported for this gene [10].

CGRP-I has been isolated and sequenced from MTC [11] and both CGRP-I and CGRP-II have been isolated and sequenced from spinal cord [12]. CGRPs have been shown to stimulate the rate and force of contraction of the heart and to be potent vasodilators (for reviews see [13,14]).

Here we report the characterization and

Correspondence address: J.W.M. Höppener, Institute of Molecular Biology, State University of Utrecht, Padualaan 8, 3584 CH Utrecht, The Netherlands

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/88/$3.50

chromosomal localization of a locus in the human genome containing nucleotide sequences highly homologous to both exon 2 and 3 of the known CALC genes. However, this third CALC locus does not seem to encode a CT- or CGRP-related polypeptide hormone and is probably a pseudogene.

2. MATERIALS AND METHODS

2.1. *Phage λ library of human genomic DNA*

An amplified human genomic DNA library in phage λEMBL3 [15] was obtained from Dr G. Grosveld (Erasmus University, Rotterdam, The Netherlands). It was prepared from high molecular mass chromosomal DNA from chronic myeloid leukemia cells.

2.2. *Southern blotting and hybridizations*

Restriction enzyme-digested DNA was size-fractionated by electrophoresis in 0.8% agarose gels. Blotting was performed overnight as described [16] onto Hybond-N membranes, using 20 × standard saline citrate (20 × SSC).

Double-stranded (ds) DNA probes were labeled using [α-^{32}P]dCTP by random priming [17] with the Klenow fragment of DNA polymerase I. Hybridizations were performed in the presence of 50% formamide and 6 × SSC at 42°C overnight. Filters were washed twice at 65°C in 2 × SSC/0.5% SDS for 30 min and exposed to Fuji-RX films.

2.3. *Nucleotide sequence analysis*

DNA restriction enzyme-generated 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, fragments labeled at only one of their 3′- or 5′-termini were isolated from polyacrylamide gels or agarose gels and sequenced using the chemical modification technique [18]. Fragments subcloned in M13 vectors were sequenced using the dideoxynucleotide chain termination method [19].

2.4. *Chromosomal assignment*

Chromosomal assignment of the CALC-III gene was performed as described previously [20]. Chromosomal DNA was isolated from a panel of 23 human-Chinese hamster and human-mouse somatic cell hybrids and their parents and digested with *Taq*I. A Southern blot prepared from these DNA digests was hybridized to a probe recognizing the CALC-III gene and hybrids were analyzed for the presence of the CALC-III gene-specific restriction fragment of 3.2 kb.

3. RESULTS AND DISCUSSION

Analysis of human genomic DNA with probes derived from the CALC-I or CALC-II genes, in combination with the known structural organization of these genes, provided evidence for the existence of a yet unknown CALC-related gene in the human genome. The possibility that this gene encodes a CT- or CGRP-like polypeptide was not ruled out. As shown in figs 1 and 2, a genomic 3.2 kb *Taq*I fragment (previously termed 3.0 kb in [8,20]) was identified which hybridized to probes containing exon 2 and/or exon 3 sequences of the CALC-I gene or the CALC-II gene, but which does not belong to either one of these two genes. The 8.0 kb and 6.5 kb fragments in fig.1 are the two alleles of a CALC-I gene-specific restriction fragment length polymorphism for *Taq*I [21] and the 2.3 kb *Taq*I fragment is derived from the CALC-II gene [6].

An amplified human genomic DNA library in phage λEMBL3 was screened with a probe containing exon 2 plus 3 sequences from the CALC-I gene (*Bgl*II-*Pst*I fragment from phC36 [3]), in order to isolate a stretch of chromosomal DNA comprising the 3.2 kb *Taq*I fragment. Of the 300000 plaques tested, 12 positive clones were selected for a second screening. Because a specific probe for the genomic locus containing the 3.2 kb *Taq*I fragment was not available, clones containing this fragment were selected by virtue of the fact that they do not hybridize to gene-specific probes (both upstream and downstream of the exon 2 plus 3 region) from the CALC-I or the CALC-II gene. One clone of the 12 selected (later termed λTG3) did not hybridize to probes containing either exon 1 or exon 4 of the known CALC genes (probes d–g

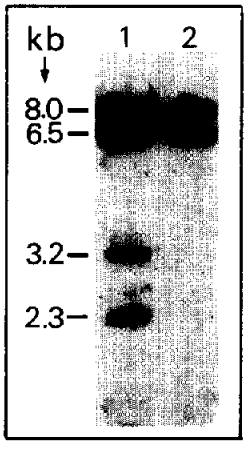

Fig.1. Southern blot hybridization analysis of *Taq*I-digested human placenta DNA to: (lane 1) plasmid phC36, containing exons 2, 3 and 4 of the CALC-I gene [3]; (lane 2) plasmid pTT1062, containing the 3′-terminal 475 bp of exon 4 (CT-encoding) of the CALC-I gene [29].

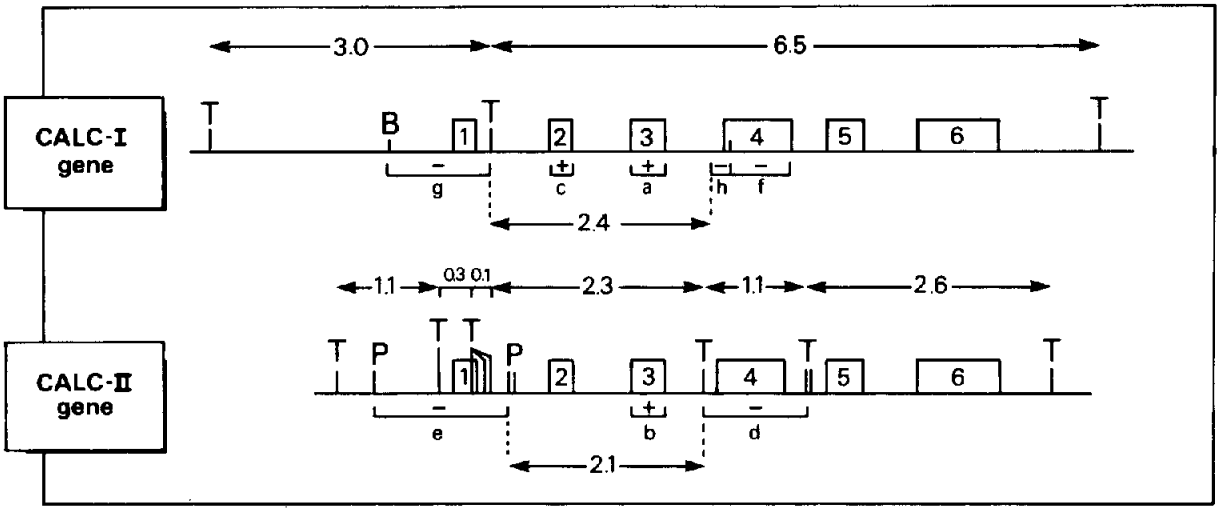

Fig.2. Schematic representation (not drawn to scale) of the CALC-I and the CALC-II gene, with the locations of *Taq*I restriction enzyme recognition sites (T) and the distances between them (in kb) indicated above. In addition, the locations of a *Bam*HI site (B) in the CALC-I gene and three *Pst*I sites (P) in the CALC-II gene are represented. Exons are numbered from 1 to 6 and the locations of probes (a–h) which do (+) or do not (−) detect the 3.2 kb genomic *Taq*I fragment are indicated. The location and the maximum size of a fragment derived from the CALC-I or the CALC-II gene which could hypothetically be contained within the genomic 3.2 kb *Taq*I fragment (which hybridizes to exon 2 and exon 3 probes), is indicated below the gene structures.

Fig.3. (A) Comparison of the exon 2-like nucleotide sequence of λTG3 with that of exon 2 of the CALC-I gene. The nature of the nucleotide at position 48 in the exon 2-like sequence could not be determined unambiguously. The ATG initiation-codon in exon 2 of the CALC-I gene is underlined. Gaps (−) were introduced to maximize homologies, asterisks (*) indicate differences in nucleotide sequence. Exon-intron boundaries are indicated by arrows. (B) Comparison of the exon 3-like nucleotide sequence of λTG3 with that of exon 3 of the CALC-I gene. An insertion of 3 nucleotides (corresponding to an alanine-encoding triplet) in exon 3 of the CALC-I gene is boxed.

in fig.2) and was shown to contain a *Taq*I fragment of 3.2 kb containing both exon 2- and exon 3-related nucleotide sequences. Restriction mapping of λTG3 DNA placed the 3.2 kb *Taq*I fragment at a distance of approx. 12 kb from the downstream insert/λEMBL3 junction in λTG3 (not shown).

The exon 2-like sequence of λTG3 consists of 95 nucleotides with approx. 80% homology to exon 2 of both CALC genes (figs 3A and 4). The 3'-splice site is considerably mutated. The ATG initiation-codon is mutated to AAA and no other ATG sequence is present within or immediately preceding the exon 2-like sequence. The 5'-splice site is very well conserved.

The exon 3-like sequence of λTG3 consists of 136 nucleotides with approx. 90% homology to exon 3 of both CALC genes (figs 3B and 4). The 3'- and 5'-splice sites are very well conserved. Three nucleotides encoding an alanine residue in exon 3 of the CALC-I gene are absent in both the CALC-II gene and in λTG3. An additional dinucleotide deletion located at a position 8 and 9 nucleotides upstream of the 5'-splice site in λTG3 is not present in either the CALC-I gene or the CALC-II gene (and would have caused a shift in the translational reading frame used for the precursors produced from these two genes).

In both CALC genes, exon 4 (-like) sequences (encoding CT in the CALC-I gene) start at a position located about 1 kb downstream of the 3'-end of exon 3. CGRP-encoding exon 5 sequences start another 1.2 kb further downstream in both of these genes [3,6]. Nucleotide sequence analysis of a region of approx. 4 kb downstream of the 3'-end of the exon 3-like sequence in λTG3 did not reveal significant homologies with either exon 4 or exon 5 of the known CALC genes. The existence of a salmon CT (SCT)-like polypeptide has been demonstrated immunochemically in man [22,23] as well as in several other species [24,25]. In addition, in vitro translation of RNA isolated from human

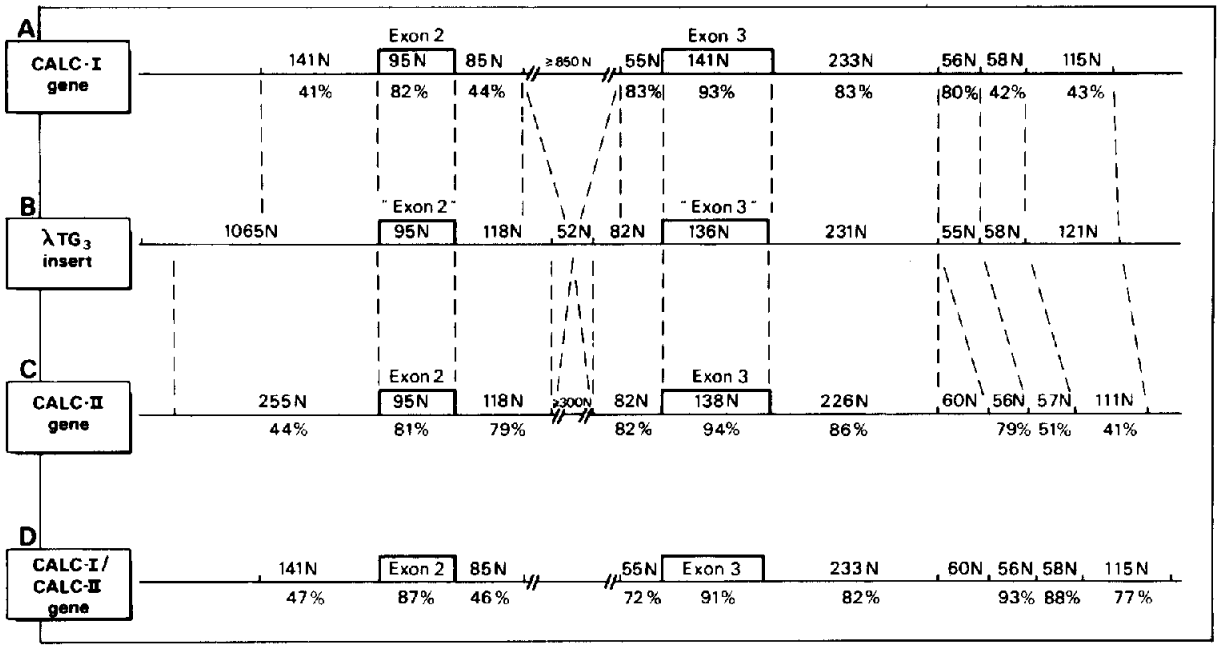

Fig.4. Schematic representation illustrating comparisons between the exon 2- and exon 3-like regions in λTG3 with the corresponding regions in the CALC-I and the CALC-II gene, respectively. Exons 2 and 3 are represented as open boxes. The size [in nucleotides (N)] of exon and intron sequences (as far as sequence data were available) is indicated in the boxes and above the genomic structures, respectively. The percentage of homology in nucleotide sequence between corresponding regions is indicated underneath the genomic structures in (A) for the comparison of the λTG3 insert with the CALC-I gene, in (C) for the comparison of the λTG3 insert with the CALC-II gene, and in (D) for the comparison of the CALC-I and the CALC-II genes to one another. The position and (estimated) size of deletions within the insert of λTG3 are also indicated (see text for further explanation).

and murine MTCs directed the synthesis of polypeptides which could be specifically precipitated with antibodies raised against SCT [26]. However, no significant homology with the nucleotide sequence of a cDNA encoding SCT [27] was detected in the 4 kb downstream of the exon 3-like sequence in λTG3. Also, DNA probes specific for human CT (exon 4) or CGRP (exon 5) or probes comprising exon 1 or 6 of the CALC-I or -II gene did not hybridize to λTG3 DNA under non-stringent conditions.

A comparison between nucleotide sequences of the λTG3 insert with corresponding regions of both CALC genes is presented in fig.4. In the region upstream of the exon 2-like sequence in λTG3, no significant homology with either one of the CALC genes is found. Nucleotide sequences immediately downstream of the exon 2-like sequence in λTG3 are clearly more homologous to the CALC-II gene than to the CALC-I gene. Intron 2 in both CALC genes is at least 300 nucleotides longer than the intron 2-like sequence in λTG3. Nucleotide sequences immediately downstream of the exon 3-like sequence in λTG3 are about 80% homologous to both CALC genes, up to a position about 290 bp downstream of exon 3 in the CALC-I gene. From that point onwards, homology with either one of the two CALC genes decreases very rapidly, although intron 3 in the CALC-I and -II genes stays about 80% homologous over the next 168–173 nucleotides sequenced from both genes (unpublished). A stretch of 60 bp within intron 3 of the CALC-II gene (starting at a position 226 bp downstream of exon 3) is missing from both the CALC-I gene and from λTG3. These results suggest that, during the course of evolution, a part of the sequences represented in λTG3 (comprising a region of approx. 770 bp) was generated by duplication of only part of the CALC-II gene and subsequent integration elsewhere in the genome (the 3.2 kb *Taq*I fragment is not present in two cosmid clones containing the CALC-I gene, nor in two other cosmid clones containing the CALC-II gene). Thereafter, an insertion of 60 bp in intron 3 of the CALC-II gene has occurred. The duplicated DNA segment might start at a position close to the intron 1/exon 2 boundary and terminate in intron 3 at a position located about 280 bp downstream of exon 3 (excluding the 60 bp insertion).

Chromosomal localization of the genomic locus containing the λTG3 insert [henceforth termed the CALC-III (pseudo)gene or CALC-III locus] was performed using the somatic cell hybrid technique. As are the CALC-I and -II genes [20,21], the CALC-III (pseudo)gene is located on human chromosome 11 (table 1). For the other chromosomes, the discordance is between 30% and 78%. Limited regional localization, using somatic cell hybrids containing translocation-derived chromosomes, placed the CALC-III locus in the q12-p12 or p15.1-pter region of chromosome 11 (fig.5). The CALC-II gene has been assigned to the same region [28]. These data support the

Table 1

Segregation of the CALC-III locus with human chromosomes in 23 human-rodent somatic cell hybrid clones

Chromosome	Chromosome/CALC-III locus number of clones[a]				
	+/+	+/−	−/+	−/−	% discordance
1	10	2	7	4	39
2	2	3	15	3	78
3	10	4	7	2	48
4	11	3	6	3	39
5	11	4	6	2	43
6	8	3	9	3	52
7	7	2	10	4	52
8	11	5	6	1	48
9	8	5	9	1	61
10	7	2	10	4	52
11[b]	10	0	0	4	0
12	6	5	11	1	70
13	7	2	10	4	52
14	7	5	10	1	65
15	9	4	8	2	52
16	13	4	4	2	35
17	11	4	6	2	43
18	9	1	8	5	39
19	9	2	8	4	43
20	11	4	6	2	43
21	15	5	2	1	30
22	14	5	3	1	35
X	12	4	5	2	39
Y	3	0	14	6	61

[a] Numbers in columns indicate the number of hybrid clones tested which contain both the numbered chromosome and the CALC-III locus (+/+), the numbered chromosome but not the CALC-III locus (+/−), the CALC-III locus but not the numbered chromosome (−/+), or neither the CALC-III locus nor the numbered chromosome (−/−)

[b] Hybrids containing chromosome 11-derived fragments (see fig.5) were excluded from these data

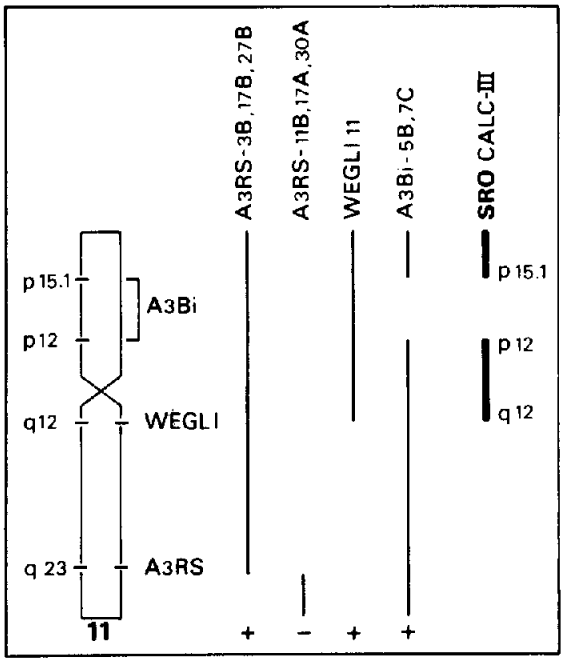

Fig.5. Regional localization of the CALC-III locus using somatic cell hybrids containing translocation-derived fragments of human chromosome 11 (represented as vertical bars). + and − indicate the presence and absence, respectively, of the CALC-III gene-specific 3.2 kb *Taq*I fragment as revealed by Southern blot hybridization analysis of chromosomal DNA from these hybrid clones. The A3RS hybrids (breakpoint 11q23) [30], the WEGLI hybrids (breakpoint 11q12) [31] and the A3Bi hybrids (deletion 11p12-11p15.1) [28] have been described previously. The SRO (shortest region of overlap) for the CALC-III locus is 11q12-11p12 plus 11p15.1-11pter.

hypothesis that all three CALC genes arose by local duplications of a common ancestral gene.

As yet we cannot exclude the possibility that the exon 2- and/or exon 3-like sequences in λTG3 are expressed at the RNA level, ligated to exons not related to either exon 4 or exon 5 (and/or exon 1) of the CALC-I or -II genes. In that case, an ATG initiation-codon might be present in an exon upstream of the exon 2-like sequence, and the dinucleotide deletion just upstream of the 5′-splice site in the exon 3-like sequence might not have to cause a frame-shift in translation of that RNA hypothetically transcribed from the CALC-III gene. The fact that the exon 3-like sequence is better conserved (as compared to the CALC-I and -II genes) than the exon 2- and intron 2-like sequences might indicate that this sequence has (or had) a biological function. However, as yet we have no data to support this hypothesis and at present it seems most likely that CALC-III is a pseudogene.

Acknowledgements: This research was supported by Ciba-Geigy AG, Basel, Switzerland. The authors would like to thank E. Kluis, I. Hassink and M.-J. Mens-Meert for assistance in preparation of the manuscript.

REFERENCES

[1] Rosenfeld, M.G., Amara, S.G. and Evans, R.M. (1984) Science 225, 1315–1320.
[2] Amara, S.G., Evans, R.M. and Rosenfeld, M.G. (1984) Mol. Cell. Biol. 4, 2151–2160.
[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.) pp.23–31, Elsevier, Amsterdam, New York.
[4] Steenbergh, P.H., Höppener, J.W.M., Zandberg, J., Van de Ven, W.J.M., Jansz, H.S. and Lips, C.J.M. (1984) J. Clin. Endocrinol. Metab. 59, 358–360.
[5] Grzeschik, K.-H. and Kazazian, H.H., jr (1985) Cytogenet. Cell Genet. 40, 179–205.
[6] Steenbergh, P.H., Höppener, J.W.M., Zandberg, J., Visser, A., Lips, C.J.M. and Jansz, H.S. (1986) FEBS Lett. 209, 97–103.
[7] Alevizaki, M., Shipaishi, A., Rassool, F.V., Ferrier, G.J.M., MacIntyre, I. and Legon, S. (1986) FEBS Lett. 206, 47–52.
[8] Steenbergh, P.H., Höppener, J.W.M., Zandberg, J., Lips, C.J.M. and Jansz, H.S. (1985) FEBS Lett. 183, 403–407.
[9] Höppener, J.W.M., Steenbergh, P.H., Slebos, R.J.C., Visser, A., Lips, C.J.M., Jansz, H.S., Bechet, J.M., Lenoir, G.M., Born, W., Haller-Brem, S., Petermann, J.B. and Fischer, J.A. (1987) J. Clin. Endocrinol. Metab. 64, 809–817.
[10] Amara, S.G., Arriza, J.L., Leff, S.E., Swanson, L.W., Evans, R.M. and Rosenfeld, M.G. (1985) Science 229, 1094–1097.
[11] Morris, H.R., Panico, M., Etienne, T., Tippins, J., Girgis, S.I. and McIntyre, I. (1984) Nature 308, 746–748.
[12] Petermann, J.B., Born, W., Chang, J.-Y. and Fischer, J.A. (1987) J. Biol. Chem. 262, 542–545.
[13] Fischer, J.A. and Born, W. (1985) Peptides 6 (suppl.3), 265–271.
[14] Goodman, E.C. and Iversen, L.L. (1986) Life Sci. 38, 2169–2178.
[15] Frischauf, A.-M., Lehrach, H., Poustka, A. and Murray, N. (1983) J. Mol. Biol. 170, 827–842.
[16] Southern, E.M. (1975) J. Mol. Biol. 98, 503–517.
[17] Feinberg, A. and Vogelstein, B. (1983) Anal. Biochem. 132, 6–13.
[18] Maxam, A.M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560–564.
[19] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.

[20] 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, 259–264.
[21] 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.
[22] Fischer, J.A., Tobler, P.H., Henke, H. and Tschopp, F.A. (1983) J. Clin. Endocrinol. Metab. 57, 1314–1316.
[23] Tobler, P.H., Tschopp, F.A., Dambacher, M.A. and Fischer, J.A. (1984) Clin. Endocrinol. 20, 253–259.
[24] Perez Cano, R., Girgis, S.I., Galan Galan, F. and McIntyre, I. (1982) J. Endocrinol. 92, 351–355.
[25] Perez Cano, R., Girgis, S.I. and McIntyre, I. (1982) Acta Endocrinol. (Copenh.) 100, 256–261.
[26] Lasmoles, F., Jullienne, A., Day, F., Minvielle, S., Milhaud, G. and Moukhtar, M.S. (1985) EMBO J. 4, 2603–2607.
[27] Pöschl, E., Lindley, I., Hofer, E., Seifert, J.-M., Brunowsky, W. and Besemer, J. (1987) FEBS Lett. 226, 96–100.
[28] Mannens, M., Slater, R.M., Heyting, C., Geurts van Kessel, A.H.M., Goedde-Salz, E., Frants, R.R., Van Ommen, G.J.B. and Pearson, P.L. (1987) Hum. Genet. 75, 180–187.
[29] Nelkin, B.D., Rosenfeld, K.I., DeBustros, A., Leong, S.S., Roos, B.A. and Baylin, S.B. (1984) Biochem. Biophys. Res. Commun. 123, 648–655.
[30] Sacchi, N., Watson, D.K., Geurts van Kessel, A.H.M., Hagemeijer, A., Kersey, J., Drabkin, H.D., Patterson, D. and Papas, T.S. (1986) Science 231, 379–382.
[31] Geurts van Kessel, A.H.M., Tetteroo, P.A.T., Van Agthoven, T., Paulussen, R., Van Dongen, J., Hagemeijer, A. and Von dem Borne, A.E.G. Kr. (1984) J. Immunol. 133, 1265–1269.