The calcitonin-like sequence of the β CGRP gene

Maria Alevizaki, Akio Shiraishi*, Feyruz V. Rassool+, Gwendolynn J.M. Ferrier, Iain MacIntyre and Steve Legon

Dept of Chemical Pathology and +*MRC Leukaemia Unit, Royal Postgraduate Medical School, Hammersmith Hospital, Ducane Road, London W12 OHS, England*

Received 1 August 1986

We have identified a region within the β CGRP gene which has the potential to encode a novel calcitonin-like peptide. The gene is located on the short arm of chromosome 11 (11p 12–14.2) and we suggest that it resulted from a local duplication of the α gene. We have been unable to detect the corresponding mRNA in a variety of tissues which express α-calcitonin. It is not clear whether this sequence can be expressed in man.

Calcitonin gene related peptide *βCGRP gene* *Calcitonin* *Gene duplication*

1. INTRODUCTION

Study of calcitonin gene expression in the rat revealed that two different mRNA species can be produced from the same primary transcript, one encoding calcitonin (CT) and the other calcitonin gene related peptide (CGRP) [1]. The 2 mature mRNAs share a region at their 5′-end (exons 1–3); exon 4 is calcitonin-specific and exons 5 and 6 are CGRP-specific. Similar findings were soon reported for the human calcitonin gene [2,3]. Recently a second gene generating an mRNA related to CGRP has been identified in both rat and man [4,5]. This is known as the β CGRP gene – the organisation of its mRNA appears to be similar to that of α CGRP [4]. While no mRNA related to calcitonin has yet been identified from this gene, evidence both for and against the existence of calcitonin related areas within the β gene has been published [4,6]. The possibility of a second calcitonin arising from the β gene seemed particularly interesting in view of reports that a salmon calcitonin-like peptide can be detected in normal individuals as well as in patients with small cell carcinoma of the lung [7,8].

We report here the isolation of a genomic clone containing most of the β CT/CGRP gene. We have performed in situ hybridisation studies with regions lacking strong homology with the α gene and have shown that it is located close to the α gene on the short arm of chromosome 11 (11p 12–14.2). We have detected an area of the β gene which is related to the CT-specific exon of the α gene. Despite a 67% homology with exon 4 of the α gene it does not appear to code for a protein similar to the calcitonin precursor and we have been unable to detect any mRNA containing this sequence. The CT-like region of the β gene does however contain an open reading frame of 65 amino acids. Although this is probably not translated it does have considerable homology with the mammalian calcitonins. Of particular note is the finding that this potential peptide has features characteristic of both the bovine and human type of calcitonin.

* Present address: Sankyo Co. Ltd, Bioscience Research Laboratories, No 2-58 Hiromachi 1-chome, Shinagawa-ku, Tokyo 140, Japan

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

2. MATERIALS AND METHODS

A human genomic library was constructed in λ EMBL-4 using 15–20 kb *Sau*3A fragments. The library was screened with a probe derived from phTB58 [9] containing exons 5 and 6 of the α gene. The chain-termination method of Sanger et al. [10] was used for sequencing using ^{35}S-labelled dATP. Both strands were sequenced in multiple determinations. Plaque screening and DNA transfers were done on nitrocellulose filters hybridising in 50% formamide, 5 × SSPE (1 × SSPE: 180 mM NaCl, 10 mM Na phosphate buffer, pH 7.7, 1 mM EDTA), 0.1% dried milk, 0.2% SDS and 10% dextran sulphate at 42°C. DNA probes were nick-translated to a specific activity of 10^8–10^9 dpm/μg DNA using [^{32}P]dCTP. Filters were washed in 0.1 × SSPE/0.2% SDS at 65°C. For non-stringent hybridisation the hybridisation temperature was reduced to 35°C and the washing was done in 2 × SSPE at 50°C. mRNA was prepared using guanidinium/CsCl [11] and poly(U)-Sepharose columns. RNAs were electrophoresed on formaldehyde/agarose gels, transferred to Pall filters, and hybridised as for DNA. To reprobe, the filter was washed at 80°C for 15 min in 2 mM Tris-HCl, pH 8.0, 2 mM EDTA with 0.5% SDS.

In situ hybridisation: female metaphase chromosomes were prepared from peripheral lymphocytes [12] and hybridised with ^{125}I-labelled probes [13]. Metaphases were banded according to Zabel et al. [14], photographed and analysed.

3. RESULTS

A library of human genomic clones was constructed and screened with a probe representing human α CGRP. Three clones were identified which did not correspond with the known restriction map of the α CT/CGRP gene and one of these, MGL23, was selected for detailed analysis. Areas hybridising strongly with exons 3 and 5 of the α gene [6,9] were detected in this clone. We have now determined the nucleotide sequence of its CGRP coding region (not shown) and have confirmed that it is indeed β CGRP. Restriction fragments lacking strong homology with the α gene were labelled with ^{125}I by nick-translation and hybridised to human metaphase chromosomes. Specific hybridisation was found to the short arm of chromosome 11 in the region 11p 12–14.2 (fig.1). This agrees with the findings of Hoppener et al. [15] (11 q1–p ter) and corresponds with the position of the α CT/CGRP gene [16] (11 p13–15) to within the limits of resolution attainable with a ^{125}I-labelled probe.

Under stringent hybridisation conditions restriction fragments were identified within the genomic clone that gave a weak signal with probes specific for the calcitonin coding exon of the α gene. Non-stringent hybridisation conditions were therefore used subsequently and strong signals were obtained. The nucleotide sequence of this region was determined and compared with the corresponding sequence of the α gene (fig.2). The overall

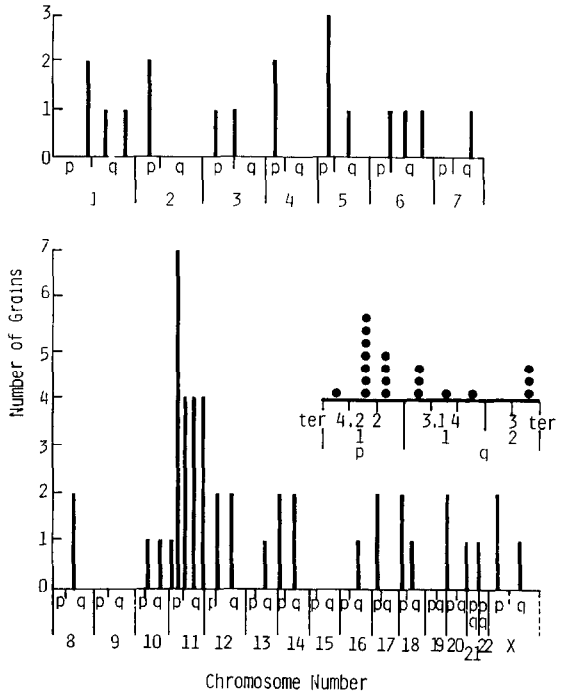

Fig.1. Localization of the β CGRP gene. 50 metaphases were analysed. The average grain count was 1.4 per metaphase and 28% of grains were localised on chromosome 11. The main histogram shows the grain count in the major regions of each chromosome. Inset: grain distribution along chromosome 11. Chromosome 11 has been divided into 8 regions. On the p arm, cent–p12, p12–p14.2 and p14.2–p ter; on the q arm, cent–q13.1, q13.1–q14, q14–q20, q20–q23 and q23–q ter.

```
   CCATGGGGACAGTCCCTAGTGCATGGTACTGTCTGGCATGTCC TTCCCTTGCAG |
                                                ** ****   ******  *** |
                                               GGTATGTGTTTTCCCTC CAG |

CTTGAGCAGTCCTAGATTTAAGTAGCATATAGTAATCTGAGTACCTGCTTGCAGGGCACA
* **   *** ** **** ****  *       ************ *** *** *******
CCTGGACAGCCCCAGATCTAAGCGGTGCG  GTAATCTGAGTACTTGCATGCTGGGCACA
                    ^^^^^^

TACTTGCAGTACCTGAAAAACTTTCATATGTTCCCTGGCATCAACTTCGGGCCTGAAATT
***  **** ** * ** ** ***** * ******     *          **** ** *
TACACGCAGGACTTCAACAAGTTTCACACGTTCCCCCAAACTGCAATTGGGGTTGGAGCA

CCTGGCAAGAATAGGGACATAGTCAACAGC    TTGCAGAGGGACCACTACCCG ACTCC
***** ***** ***** ** ***  ****    *** **** *****   ***  *
CCTGGAAAGAAAAGGGATAT GTC  CAGCGACTTGGAGAGAGACCATCGCCCTCATGTT
      ^^^^^^

AGGGTCCCCCAGA TGGCAGCTG AACTTCTCTCAACTCT CCTGATTCCCCTTCTTGC
**  * ******* ** ** **  **** *** *    ** *** *** **********
AGCATGCCCCAGAATGCCAACT AAACTCCTCCCTTTCCTTCCTAATTTCCCTTCTTGCA
                       + ++

TCCACTTTAT GAACCTGATGCATGTGGATT    CCTCTCTGATTTGTCTTCATGCTGG
***  **  *  *** ************ **    ******** *   ****   *****
TCC  TTCCTATAACTTGATGCATGTGGTTTGGTTCCTCTCTGGTGGCTCTTTGGGCTGG

TATTGGTATTTTTGCTTATGACAGAGAATGTTTTGAAGACCTCAGGATGGAAGGGAAGAC
*******   *** *** ** ***** **** *  ** ** *** ***** *** ****
TATTGGTGGCTTTCCTTGTGGCAGAGGATGTCTC AA ACTTCAAGATGGGAGGAAAGAG
    T
AGCAGGACTTACTGAACACGTT   AGAGA                TAAAAGAAAATAAGGG
*********       *** ***   *****                **  ******* ****
AGCAGGACTC      ACAGGTTGGAAGAGAATCACCTGGGAAAATACCAGAAAATGAGGG

AAGCTTCTTGAGACTGT AGAGGGTGTTATGACAGAGGCATCCAATTT CTGCTTCTAAAT
  ***  ***** *    ****  *** **  *****   ***  * * ******** ***
CCGCT  TTGAGTCCCCCAGAGA TGTCAT  CAGAGC  TCCTCTGTCCTGCTTCTGAAT

GTACTACGATAAAATAAGCACGTCCTTAATGCCTTCGGATTAGATGAATCAT CTATTTTT
** **  ***                                *** ** **** *   *******
GTGCT  GAT                              CATTTGAGGAATAAAATTATTTTT
                                                 ------

CTAA | AAGGA CTGAGCTGCGGTGCTCATTGCTCTGGTAC = BETA
*    | ** ** ******** **** *********** *
CCCC | AAAGATCTGAGCTGTGGTGGTCATTGCTCTGATCT = ALPHA
```

Fig.2. The β gene calcitonin-like region. The nucleotide sequence of exon 4 of the α gene [6,9] is shown aligned with the corresponding region of the β gene. Important features are indicated as follows: (|) limits of exon 4; (^^^) codons for paired dibasic amino acids bounding calcitonin sequence; (+++) termination triplet of α gene; (---) poly(A) addition sequence; (*) common residues. Residue 407 in the β gene sequence is uncertain, giving clear but conflicting data when the 2 strands were sequenced.

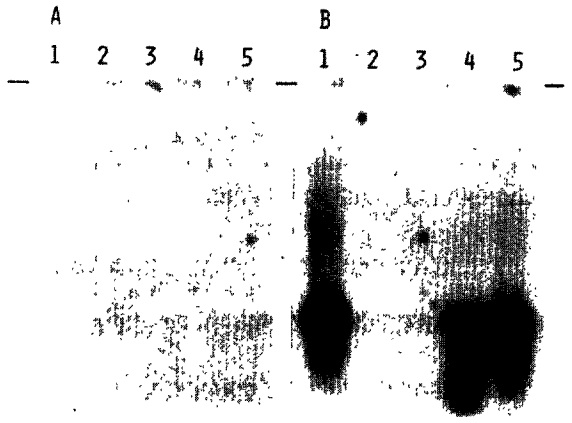

Fig.3. Northern blot analysis of human mRNA samples. (A) Probing with β gene sequence; (B) probing the same filter with α-calcitonin probe. Tracks: 1, 3 μg BEN cell mRNA (bronchial carcinoma [21]); 2, 3 μg HL60 mRNA (leukaemia cell line); 3, 3 μg phaeochromocytoma mRNA; 4, 0.5 μg sporadic type medullary thyroid carcinoma; 5, 0.5 μg familial type medullary thyroid carcinoma.

homology is 67.6%. The calcitonin coding sequence has not been conserved to a greater degree than other regions and in fact, over its first 200 bases, the 3'-non-coding region shows slightly more homology with its β counterpart.

As the β gene sequence contains an open reading frame of 65 amino acids (see below) the possibility that it might be expressed could not be ignored. A Northern blot was performed with mRNA from various tissues where expression might be expected. No signal was detected on the filter after 3 weeks exposure (fig.3A). The same filter was then reprobed with an α-calcitonin-specific probe (fig.3B). Calcitonin message was detected after 2 days exposure.

4. DISCUSSION

We have presented the structure of the calcitonin-like region of the gene encoding β CGRP. The principal interest in this region lies in the possibility that it encodes a previously unidentified regulatory peptide. Our attempts to detect an mRNA representing this region were unsuccessful (fig.3) but this must be treated with caution since we cannot know where or at what level such an mRNA might be expressed.

If this sequence is to be expressed in the manner of the calcitonin coding exon of the α gene then it would be excised from a much longer transcript. It is known that the β gene has a high degree of homology with exons 3 and 5 of the α gene [4] and that these are spliced together to produce β CGRP. The splice donor signals at the 3'-end of exon 3 are therefore functional and if suitable signals exist around the exon 4-like sequence then an alternate mRNA could be produced. The consensus sequence preceding a splice acceptor site is $(Y)_{11}$NYAG (Y = pyrimidine, N = any base) [17] and sequences conforming precisely to this paradigm are found before both α and β gene sequences in fig.2. The other likely requirement for the production of a message would be the conservation of sequences involved in the formation of the polyadenylated 3'-terminus. Such sequences are believed to comprise both the region preceding the polyadenylation site (including the hexanucleotide AATAAA) and sequences beyond this point [18]. The former is a region of strikingly low homology between α and β genes whilst the latter shows 87% homology. It seems unlikely that normal processing can occur at this site but the possibility cannot be ruled out.

If a calcitonin-like mRNA were somehow produced then the reading frame of the protein would be defined by the 3/4 splice and, since β CGRP would share the same splice donor site in exon 3, we can deduce the reading frame from that of β CGRP [4]. This reading frame terminates after only 8 amino acids. There is an open reading frame of 65 amino acids which starts at position 76 of fig.2. We have considered the possibility of its expression from an alternative splice acceptor site but there is no other suitable sequence. It therefore seems unlikely that this sequence is expressed in man. However, it is possible that the predicted peptide or a close homologue would have useful biological properties. A stretch of 33 amino acids shows homology with the calcitonins though this region is not flanked by paired dibasic amino acids as would be expected if it were to be cleaved from a larger precursor protein.

Three types of calcitonin (salmon, bovine and human) have been described [19,20] and a comparison with the β gene peptide is presented in

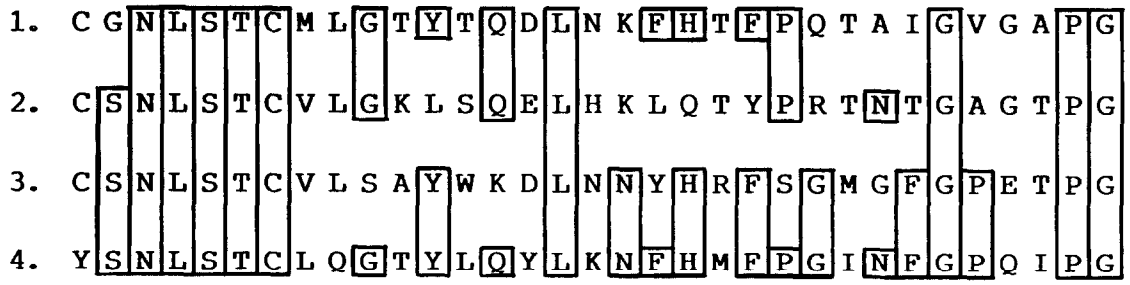

Fig.4. Comparison of β gene peptide with calcitonins. The predicted β gene peptide (4) is compared with the 3 known types of calcitonin sequence. Consensus human type (1), salmon type (2) and bovine type (3) calcitonin sequences were compiled from published sources [19,20], selecting the most abundant residue within each group. Where 2 amino acids were equally frequent, the one found in other types was chosen and in the one case where this did not permit assignment, the human amino acid was chosen rather than the rat.

fig.4. The β gene peptide has specific features in common with both mammalian-type sequences and might be considered as a separate type. The homology of the exon 4-like DNA with the α gene is roughly constant at 67% throughout its length (fig.1). This is comparable with the CGRP 3′-non-coding sequences (65% homologous) and in marked contrast to the CGRP coding sequences which are 92% homologous [4]. We suggest therefore that the β gene may have arisen from a gene duplication on the short arm of chromosome 11. The exon 4-like region of the new gene then lost its potential to code for a calcitonin. Whether the β gene is now committed only to CGRP production or has some other alternative product is a question whose answer must await complete sequencing of the gene.

ACKNOWLEDGEMENTS

We thank Dr R. Craig for providing the cDNA clones pHT B3 and pHT B58 from which our probes were derived. M.A. is in receipt of The Edmund Davis Scholarship in Medicine of The University of London; A.S. was a Sankyo Training Fellow. This work was supported in part by The Armand Bartos Foundation.

REFERENCES

[1] Amara, S.G., Jonas, V., Rosenfeld, M.G., Ong, E.S. and Evans, R.M. (1982) Nature 298, 240–244.
[2] Steenbergh, P.H., Hoppener, J.W.M., Zandberg, J., Van de Ven, Jansz, H.S. and Lips, C.J.M. (1984) J. Clin. Endocrinol. Metab. 59, 358–360.
[3] Morris, H.R., Panico, M., Etienne, T., Tippins, J., Girgis, S.I. and MacIntyre, I. (1984) Nature 308, 746–748.
[4] Steenbergh, P.H., Hoppener, J.W.M., Zandberg, J., Lips, C.J.M. and Jansz, H.S. (1985) FEBS Lett. 183, 403–407.
[5] Amara, S.G., Arriza, J.L., Leff, S.E., Swanson, L.W., Evans, R.M. and Rosenfeld, M.G. (1985) Science 229, 1094–1096.
[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] Tobler, P.H., Tschopp, F.A., Dambacher, M.A. and Fischer, J.A. (1984) Clin. Endocrinol. 20, 253–259.
[8] Gropp, C., Luster, W. and Havemann, K. (1985) Br. J. Cancer 51, 897–901.
[9] 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.
[10] Sanger, F., Coulson, A.R., Barell, B.G., Smith, A.J. and Roe, B.A. (1980) J. Mol. Biol. 143, 161–178.
[11] Glisin, V., Crkvenjakov, R. and Byus, C. (1974) Biochemistry 13, 2633–2637.
[12] Ferguson-Smith, M.A. (1974) La Ric. Clin. Lab. 4, 297–335.
[13] Boyd, B.Y., Buckle, V.J., Munro, E.A., Choo, K.H., Migeon, B.R. and Craig, I. (1984) Ann. Hum. Genet. 48, 145–152.
[14] Zabel, B.U., Naylor, S.L., Sakaguchi, A.Y., Bell, G.I. and Snows, T.B. (1983) Proc. Natl. Acad. Sci. USA 80, 6932–6936.

[15] Hoppener, 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.
[16] Przepiorka, D., Baylin, S.B., McBride, O.W., Testa, J.R., de Bustros, A. and Nelkin, B.D. (1984) Biochem. Biophys. Res. Commun. 120, 493–499.
[17] Mount, S.M. (1982) Nucleic Acids Res. 10, 459–472.
[18] Birnstiel, M.L., Busslinger, M. and Strub, K. (1985) Cell 41, 349–359.
[19] MacIntyre, I. (1978) J. Endocrinol. Invest. 1, 277–284.
[20] Lasmoles, F., Desplan, C., Segond, N., Jullienne, A., Treilhou-Lahille, F., Milhaud, G. and Moukhtar, M.S. (1985) FEBS Lett. 180, 113–116.
[21] Ellison, M., Woodhouse, D., Hillyard, C., Dowsett, M., Coombes, R.C., Gilby, E.D., Greenberg, P.B. and Neville, A.M. (1975) Br. J. Cancer 32, 373–379.