Structure and methylation of the human calcitonin/α-CGRP gene

P.M.Broad, A.J.Symes, R.V.Thakker[1] + and R.K.Craig*§

Cancer Research Campaign Endocrine Tumour Molecular Biology Group, Medical Molecular Biology Unit, Department of Biochemistry, University College and Middlesex School of Medicine, The Windeyer Building, Cleveland Street, London W1P 6DB and [1]Department of Medicine, University College and Middlesex School of Medicine, The Middlesex Hospital, Mortimer Street, London W1P 7PN, UK

Received May 3, 1989; Accepted August 3, 1989 EMBL accession no. X15943

ABSTRACT

We report a detailed analysis of the human calcitonin/α-CGRP gene locus. About 39kb of DNA containing the gene has been mapped and a common *Pvu* II RFLP identified downstream of the gene. DNA sequence analysis revealed an extensive CpG island containing several rare restriction enzyme sites at the 5′ end of the gene. The structure of this island is unusual in that it contains two distinct CpG-rich regions, one located around exon 1 and the other about 1.5kb further upstream. *Msp* I sites within both CpG-rich regions were found to be unmethylated, regardless of whether the calcitonin/α-CGRP gene was being expressed. However, a correlation was found between demethylation of *Msp* I sites in intron 2, downstream of the CpG island, and calcitonin/α-CGRP gene expression. DNA sequence analysis also revealed the presence of several binding sites for constitutive and regulatory transcription factors in the promoter of the gene. These results suggest that both unmethylated CpG islands and specific demethylation of internal sequences may play a role in the activation of calcitonin/α-CGRP gene transcription.

INTRODUCTION

The calcitonin/α-CGRP gene encodes the calcium-lowering hormone calcitonin and the neuropeptide α-CGRP. These peptides are processed from precursor proteins encoded by calcitonin mRNA and α-CGRP mRNA respectively. Both mRNA species are generated from a common primary transcript by tissue-specific alternative RNA processing events (1,2,3). Previous studies on the human calcitonin/α-CGRP gene have demonstrated that it is organised in a similar fashion to the rat gene (4,5,6,7,8). The structure of the calcitonin/α-CGRP gene appears to have been conserved in evolution since the chicken calcitonin gene can generate an mRNA encoding a putative CGRP molecule (9). The human calcitonin/α-CGRP gene is part of a gene family containing at least two other members. The β-CGRP gene (10) encodes a CGRP molecule (β-CGRP) closely related to α-CGRP (11) but is not capable of generating a calcitonin-like mRNA (12). Another cross-hybridising gene appears to be a pseudogene (13). All three of these genes are localised to chromosome 11p (12,13,14).

As an initial step towards the study of the tissue-specific regulation of human calcitonin/α-CGRP gene expression in thyroid and neural tissues, and the ectopic expression of this gene in lung carcinoma, we have determined the complete sequence of this gene and over 1.8kb of 5′ flanking sequence. An extensive CpG island was found at the 5′ end of the gene. CpG islands are unmethylated regions of DNA which are rich in CpG and have a high G+C content. Such islands are associated with all housekeeping genes and some tissue-specific genes (15,16). The CpG island at the calcitonin/α-CGRP gene is unmethylated whereas an internal region of the gene exhibits a methylation pattern which correlates with

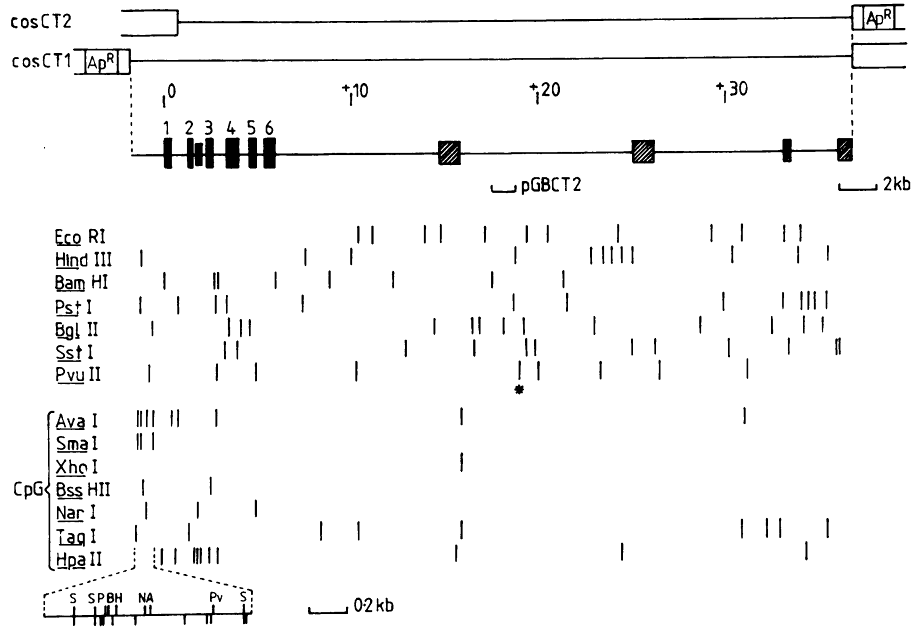

Figure 1: *Restriction map of DNA at the human calcitonin/α-CGRP gene locus*
The structure and restriction map of the inserts of cosCT1 and cosCT2 are shown. The orientation of the vector cos202 with respect to each insert is given by the position of the ampicillin restistance gene (Ap^R) gene. Below cosCT1 and cosCT2 is a map of the 38.5kb of genomic DNA which the inserts span. The positions of the six exons of the calcitonin/α-CGRP gene (numbered shaded boxes) and the five restriction fragments containing Alu repeats (hatched boxes) are shown. The scale (in kb) is relative to the transcription start site of the gene. The position of the 1.3kb *Bam* HI-*Hind* III fragment subcloned into pGEM4Blue to yield pGBCT2 is indicated. The *Pvu* II site present in cosCT1 but absent from cosCT2 is marked by an asterisk. The restriction enzyme sites mapped are grouped into those which contain the CpG dinucleotide (lower group) and those which do not (upper group). *Hpa* II sites within the region −1.9kb to −0.8kb are shown on an expanded scale, where they are marked below the line. Other restriction sites in this region are marked above the line (A, *Ava* I; B, *Bss* HII; H, *Hind* III; N, *Nar* I; P, *Pst* I; Pv, *Pvu* II;S, *Sma* I ; note that the *Sma* I sites are also *Ava* I sites).

calcitonin/α-CGRP gene expression. Restriction analysis of 31 kb of DNA downstream of the gene identified a common *Pvu* II restriction fragment length polymorphism. This should prove to be a useful marker in the analysis of allele loss from chromosome 11p in a variety of tumours types.

MATERIALS AND METHODS

Isolation and analysis of cosmid clones

Cosmid clones were isolated from the HPB-ALL cos202 human genomic library constructed by Kioussis *et al.* (17). The library was screened according to the procedure of Grosveld *et al.* (18). Cosmid DNA was restriction mapped by a combination of Southern blotting and subcloning of fragments into pGEM1 and pGEM4Blue (P+S Biochemicals). Alu repeat sequences were positioned by using the insert of pMA60, an Alu-repeat-containing clone (M.S. Davies and R.K.C., unpublished) as a hybridisation probe on Southern blots of digested cosmid DNA. Sequencing was performed by subcloning suitable restriction fragments of cosCT1 subclones into M13mp18 and M13mp19 and sequencing them using

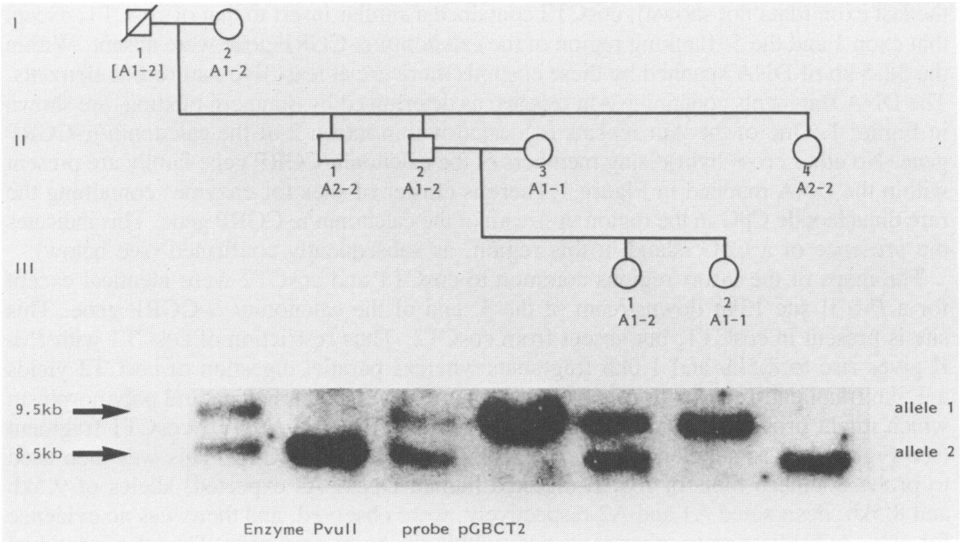

Figure 2: *Codominant inheritance of a Pvu II restriction fragment length polymorphism*
Autoradiograph of a Southern blot of *Pvu* II-digested human DNA from a three generation family hybridised to pGBCT2 is shown. The family tree is drawn so that each member appears above his or her allele pattern. The 9.5kb *Pvu* II allele is designated A1 and the 8.5kb *Pvu* is designated A2.

the dideoxy chain termination method (19,20); with either the M13 primer or synthetic oligonucleotides. Both strands were sequenced through all sites used for subcloning.

Genomic DNA isolation and Southern blot analysis

Human genomic DNA was isolated as descibed by Broad *et al.* (21). 5μg samples were restricted and then electrophoresed through 0.8%(w/v) agarose gels (for *Pvu* II RFLP analysis) or 1.1%(w/v) gels (for methylation analysis).Southern blotting (22) was performed using Hybond-N membranes (Amersham). Southern blots were hybridised to plasmids or restriction fragments labelled by nick-translation (23) or oligo-labelling (24). After hybridisation, the blots were washed in finally in 0.1×SSC/0.1% SDS at 65°C before exposure to preflashed Fuji RX film between intensifying screens for 72 hours (RFLP analysis) or Kodak XAR-5 film for 2−5 days (methylation analysis) at −70°C.

RESULTS AND DISCUSSION

Isolation and characterisation of cosmid clones of the human calcitonin/ α-CGRP gene.

The HPB-ALL cos202 human cosmid library constructed by Kioussis *et al.*, (17) was screened with a 370bp *Pst* I-*Bgl* II fragment from the 5′ end of the insert of the cDNA clone phT-B58 (7). This fragment derives from exon 4 of the calcitonin/α-CGRP gene. Two cosmids, designated cosCT1 and cosCT2, were isolated from the library. Å restriction map of the inserts of these cosmids is presented in Figure 1. cosCT1 contains all six exons of the calcitonin/ α-CGRP gene, spanning 5.7kb, together with 1.8kb of 5′ flanking sequence and 31kb of 3′ flanking sequence . The calcitonin/α-CGRP gene and upstream sequences were subcloned from this cosmid. Using Southern blotting it was determined that the copy of the calcitonin/α-CGRP gene present in cosCT1 was unrearranged between a *Hind* III site 1.5kb upstream of the first exon and a *Bgl* II site 7.5kb downstream of the last exon (data not shown). cosCT2 contained a similar insert to that of cosCT1, except that exon 1 and the 5′ flanking region of the calcitonin/α-CGRP gene were absent. Within the 38.5 kb of DNA spanned by these cosmids there are at least five Alu repeat elements. The DNA fragments containing Alu repeats, as determined by Southern blotting, are shown in Figure 1. One of the Alu repeats is located within intron 2 of the calcitonin/α-CGRP gene. No other cross-hybridising members of the calcitonin/CGRP gene family are present within the DNA mapped in Figure 1.There is cluster of sites for enzymes containing the rare dinucleotide CpG in the region upstream of the calcitonin/α-CGRP gene. This indicates the presence of a CpG island in this region, as subsequently confirmed (see below).

The maps of the insert regions common to cosCT1 and cosCT2 were identical except for a *Pvu* II site 13kb downstream of the 3′ end of the calcitonin/ α-CGRP gene. This site is present in cosCT1, but absent from cosCT2. Thus restriction of cosCT1 with *Pvu* II gives rise to 8.5kb and 1.0kb fragments whereas parallel digestion of cosCT2 yields a 9.5kb fragment. In order to determine whether this was a common natural polymorphism which might prove useful in linkage studies, a 1.3kb *Bam* HI-*Hind* III cosCT1 fragment (see Figure 1) was subcloned into pGEM4Blue, yielding pGBCT2. This was then used to probe Southern blots of *Pvu* II-digested human DNA. As expected, alleles of 9.5kb and 8.5kb, designated A1 and A2 respectively, were observed, and there was no evidence for cross-hybridisation to other sequences within the human genome. Thus the individual from whom the cosmid library was generated is heterozygous for this *Pvu* II polymorphism. Figure 2 shows a Southern blot of a family in which all three genotypes occur. The frequencies of alleles A1 and A2 were determined by analysis of 47 unrelated Caucasians; five were homozygous for A1, eleven were homozygous for A2, and thirty one were heterozygous, giving an allele frequency for A1 and A2, of 0.436, and 0.564 respectively. The polymorphic information content of this locus, calculated using the formula of Botstein *et al.*, (25) is 0.371, a value close to the theoretical maximum of a two allele system. Analysis of the inheritance of alleles A1 and A2 in eleven three-generation families comprising 175 individuals confirmed that they segregated in a Mendelian fashion and were inherited codominantly. The only previously reported polymorphism at the calcitonin/α-CGRP gene is a polymorphic *Taq* I site. The *Taq* I RFLP gives rise to alleles of 7.0 and 9.0kb, detectable with a calcitonin cDNA probe (26,11). Within cosCT1 a *Taq* I site 2.4kb 3′ to exon 6 of the calcitonin/α-CGRP gene is positioned 7.0kb downstream of a *Taq* I site in intron 1 and 2.0kb upstream of the next site (Figure 1), indicating that this *Taq* I site is the polymorphic site and that the smaller (7.0kb) *Taq* I allele is present in both cosCT1 and cosCT2.

The human calcitinon/α-CGRP gene has been localised to chromosome 11p15 (26,14). The cluster of rare-cutting enzymes and the *Pvu* II RFLP reported here should facilitate the construction of physical and genetic maps of this region. The *Pvu* II RFLP should also assist the analysis of allele loss from 11p observed in a variety of tumours (27,28,29).

Characterisation of a CpG island at the 5′ end of the calcitonin/α-CGRP gene

As first demonstrated in the rat, the calcitonin/α-CGRP gene comprises six exons (2). Exons 1, 2, 3 and 4 are spliced together to yield calcitonin mRNA. Exons 1, 2, 3, 5 and 6 are spliced together to yield α-CGRP mRNA. Thus alternative 3′ splice sites precede exons 4 and 5 and alternative polyadenylation sites are located at the ends of exons 4 and 6. Two groups have reported exon sequences of the human calcitonin/α-CGRP gene (6,8).

```
                     GATCAATTAAGGGCATCTTAGAAGTTAGGCGTT  -1801
  CCCGCTGCCTCCTTTGAGCACGGAGGCCACCAACCCCTAGGGGGAAGAGA  -1751
  TGTAGCGCCGAGGCAGGGGTGTCGTGCTAAGAAATTTCGACGCTTCTGGGG  -1701
  ACTGAGGACAAAGGTGCGGACACGACCCCGGGGTACCTGGAGTTCCGTGA  -1651
  CTCGCGCCACGGACGGCACACCTAGGGGCTAATTTCTGCTCTGCCTCAAA  -1601
  GAACCTCAAGCTAGAGTCCTTGCCTCCGCCCACAGCCCCGGGATGCCGCT  -1551
  GCTGCGCTCACCGCACAGGCAGCGCCCGGACCGGCTGCAGCAGATCGCGC  -1501
  GCTGCGCGTTCCACCGGGAGATGGTGGAGACGCTGAAAAGCTTCTTTCTT  -1451
  GCCACTCTGGACGCTGTGGGCGGCAAGCGCCTTAGTCCCTACCTCTGCTG  -1401
  AGCTGAACGCTCAGGCACAGTGGAACTGAAACCCGGTTCTGCGGGATGTG  -1351
  AGAGCTGTTGAGGTCACGCGTAATTGGGTGTGATGGAGGGCGCCTGTTCG  -1301
  TGATGTGTGCAGGTTTGATGCAAGCAGGTCATCGTCGTGCCGAGTGTGTGG  -1251
  ATGCGACCGCCCGAGAGACTCGGAGGCAGGCTTGGGACACGTTTGAGTGA  -1201
  ACACCTCAGGATACTCTTCTGGCCAGTATCTGTTTTTTAGTGTCTGTGAT  -1151
  TCAGAGTGGGCACATGTTGGGAGACAGTAATGGGTTTGGGTGTGTGTAAA  -1101
  TGAGTGTGACCGGAAGCGAGTGTGAGCTTGATCTAGGCAGGGACCACACA  -1051
  GCACTGTCACACCTGCCTGCTCTTTAGTAGAGGACTGAAGTGCGGGGGTG  -1001
  GGGGTACGGGGCCGGAATAGAATGTCTCTGGGACATCTTGGCAAACAGCA   -951
  GCCGGAAGCAAAGGGGCAGCTGTGCAAACGGCTCAGGCAGGTGATGGATG   -901
  GCAGGGTAGGAAGGGGGAGGTCCAGAGGTCTGGATGGAGGCTTCCGCATC   -851
  TGTACCTTGCAACTCACCCCTCAGGCCCAGCAGGTCATCGGCCCCCTCCT   -801
  CACACATGTAATGGATCTGAAGAGTACCCCGGGACAGTCCGGGGAGATGG   -751
  AGATTCGGAAAGTATCCATGGAGATCTTACAGAATCCCCTATGCGGACCA   -701
  GGAAACTCTTGTAGATCCCTGCCTATCTGAGGCCCAGGCGCTGGGCTGTT   -651
  TCTCACAATATTCCTTCAAGATGAGATTGTGGTCCCCATTTCAAAGATGA   -601
  GTACACTGAGCCTCTGTGAAGTTACTTGCCCATGATCACACAACCAGGAA   -551
  TTGGGCCAACTGTAATTGAACTCCTGTCTAACAAAGTTCTTGCTCCCAGC   -501
  TCCGTCTCTTGTTTCCCACGAGCCCTGGCCCTCTGTGGGTAATACCAGCT   -451
  ACTGGAGTCAGATTTCTTGGGCCCAGAACCCACCCTTAGGGGCATTAACC   -401
  TTTAAAATCTCACTTGGCAGGGTCTGGATCAGAGTTGGAAGAGTCCCTAC   -351
  AATCCTGGACCCTTTCCGCCAAATCGTGAAACCAGGGGTGGAGTGGGGCG   -301
  AGGGTTCAAAACCAGGCCGGACTGAGAGGTGAAATTCACCATGACGTCAA   -251
  ACTGCCCTCAAATTCCCGCTCACTTTAAGGGCGTTACTTGTTGGTGCCCC   -201
  CACCATCCCCCACCATTTCCATCAATGACCTCAATGCAAATACAAGTGGG   -151
  ACGGTCCTGCTGGATCCTCCAGGTTCTGGAAGCATGAGGGTGACGCAACC   -101
  CAGGGGCAAAGGACCCCTCCGCCCATTGGTTGCTGTGCACTGGCGGAACT    -51
  TTCCCGACCCACAGCGGCGGCAATAAGAGCAGTCGCTGGCGCTGGGAGGC     -1
  ATCAGAGACACTGCCCAGCCCAAGTGTCGCCGCCGCTTCCACAGGGCTCT     50
1 GGCTGGACGCCGCCGCCGCCGCTGCCACCGCCTCTGATCCAAGCCACCTC    100
  CCGCCAGGTGAGCCCCGAGATTCTGGCTCAGGTATATGTCTCTCCCTCCC    150
  TCTCCCTCCATTCGTCATTTTCTCACTCCCTTTCCTCCTCTCCCTCTCTC    200
  TCCGTTAGTCTCTTCATCAGATAGTCTCTGTTAGTCCGCGATTTATACCA    250
  GCTCGTGCCCTAGGTTGGATCGGACAGTCTCAATCCCCCGGCTCGCTCTT    300
  CCTGCTCGGCTGCGGACTCCAGTCTTACTCTCTCGCACTGCACACAGGCT    350
  TAGGCCAGTCTCGGGACACTCAGGCTCCCCAGGGACCGCGCACAGAGCCT    400
```

```
  GAGGCAAGAGAAACTTTCCGCAGACGGTGCGATCAGGGACGGCGTCTGGA    450
  GCCCAGCAGTCCCAGGGAAATTGGTTCAGAACCTGGAACAGAGCGGATGG    500
  GTGGCAAATAGGCACGACGACTGAGGGACAAGCAGCCCTAAACTGCAAGC    550
  CCCAGTCACAGGCTCTGGGAGCAAAGAGAGAAAGTCTTGGGTCCCAATTT    600
  AAGAGTAAAACTTGGTTCCTGCAGGGGACTGAGGTGCCAGCAGGCGGAGC    650
  TTAGGCTGAGCACTAAAATTGATTCTTTTATTTCTCAGATTCCAATTGCA    700
  GAGATAATCTCTGTGGATAGATCCCACCCCTCCCACCAAGGTGTATGCAA    750
  AAGCCCCAGAAGAGGAGGACAGCTCTGGGTTTCCTTATGGATCAGCTTC    800
  TGTGACAGGGGCCTGGGAGCATGTCCTCCACAGAGACCTGGCTAACAGGT    850
  TTATATGTATCTGTGTGGCTTTGGGGAGAAGGGTAGGACTGAGACCTCCC    900
  CCACTCAGCTCTCCCTCTAGGATCGCTCCTCCAGCTCTATCAGACTGTGC    950
  ATGCTACAGGCCCACTGAACCTCCCAGAAGTCCACTGTGCTGATGAGGAA   1000
  ACAAAAAAGAAGAGAAATATAATGACCAGTCTGAAGCCACAAAGCCAGAG   1050
  CGTTAGAAGCAAGACAGAAAACCAGGATAGGAGCCAAGGTCTGTGGGCTC   1100
  CTAGAACTTCGAACTGGCGCAGGTGGTTTATCTCATTCTTCCCTTGCAGA   1150
           MetGlyPheGlnLysPheSerProPheLeuAlaLeuSerIle   (14)
2 GAGGTGTCATGGGCTTCCAAAAGTTCTCCCCCTTCCTGGCTCTCAGCATC   1200
  LeuValLeuLeuGluAlaGlySerLeuHisAlaAlaProPheAr         (29)
  TTGGTCCTGTTGCAGGCAGGCAGCCTCCATGCAGCACCATTCAGGTAAGA   1250
  CAGCCTGAAGCCAGAAGGACACTGGTATCAGACAATTCCATGCCTCTTAA   1300
  GTGTGGTGTACTGAATGCTTAAAAACCGACAGGGGGCCGGGGTAAGGTAT   1350
  GTACTCATGCACATATATAATTTTTTGATATAAAAATGTGTGGTACACAA   1400
  TTTATAAGTAGTGATAAATATTTAGACAATATTCTTCATTATAATTTTTT   1450
  TGAGACAGAGTCTCACTCTGTCATCCATGCTGGAGTGCAATGGCACCGTG   1500
  TCGGCTCACTGCAACTTCCATCTCCTGGGTTCAAGCAATTCTCCTGCCTC   1550
  AGCCTCCCAAGTAGCCGGGATTACAGGCGCCCACCACCATGCCCGGCTAA   1600
  TTTTTGTATTTTTAGTAGAGATAGGGTTTCACCATGTTGGCCAGGCTGGT   1650
  CTCAAACTCCTGACCTCAGGTGATCCACCTGCCTCAGCCTCCCAAAGTGC   1700
  TGAGATTACAGCCGTGAGCCACCGCGCCCGGCCTATTATAAATTTTATAT   1750
  AGCCCACTGAGTCTCACAAAATGCTTTTGTTGCTTTTAGCCAAACTTTTG   1800
  TATGTGTAGCCATCCTATTTATCCTATTCAGCCATGATTTGAAAAATGAA   1850
  TTGCATCCTATTCTGTTAAATAGTTTGCCACTAAATTTGTTATTGTTAAA   1900
  TCTGGTATTTTACCTGTTGATCTATTCTCACCCACATATAAATCTATTAA   1950
  ACTGAAACTATTTTCAGGTTCAGCACTAACTGTAAGTTTTTGCATTTAAT   2000
  TTTTAATAATGGCTGCACTCAGCTTGCAAAATTCTTGAAAATTTAACCAT   2050
  TAGCTTTCACAAGCCTATACAAACTGGCTCCAGCACACCACTGTTTAGAG   2100
  GCCACACCAGTGCCTGGGTCCTGAGGAGGACACTGGCCTTGTGCCCTGTC   2150
  CCCTAGGACTCCCGCTGGCCACATCCTCAGGGGAAGAAGCAAAGACCAGG   2200
  AAGCCTGGCTGCTTATCCTGGGGAGGGGCAGGCAGGGGCTCACAGCCTGC   2250
                          gSerAlaLeuGluSerSerProAlaA   (38)
  ACTGAGTTTGCTTCCCCTCCACAGGTCTGCCCTGGAGAGCAGCCCAGCAG   2300
  spProAlaThrLeuSerGluAspGluAlaArgLeuLeuLeuAlaAlaLeu   (54)
  ACCCGGCCACGCTCAGTGAGGACGAAGCGCGCCTCCTGCTGGCTGCACTG   2350
3
  ValGlnAspTyrValGlnMetLysAlaSerGluLeuGluGlnGluGlnGl   (71)
  GTGCAGGACTATGTGCAGATGAAGGCCAGTGAGCTGGAGCAGGAGCAAGA   2400
  uArgGluGlySerSe                                      (76)
  GAGAGAGGCTCCAGGTGAGGCTCCCCAAGCGCTCAGCACAGGGCCTCCT   2450
  CTCCCCGCAGCATACACAGGAAGGTGGATCCCGAGAGGTAGGAGAGAACA   2500
  CACTGGCCAGGAATCCAACAGGCTGTGTTGTTCACCGGGACCTGGGGCCC   2550
  AGCTGTTCTCAGCCTCCAAGGGAGACAGAGGTCCCACTGCAGCTGGAGGT   2600
  ACCGTGGTTAGACATAACAAAAGGCTCCGTTTCTGAAAGTTCTTAGGAAA   2650
  TGAAATGGGGAGGTGTGGAATCGCTCACTGTGGGAATTGTTCTTGCAGTA   2700
  CTGGGAGACCTCCCAGCACTGGATGAGTTAAGACTAGTAAGGGTGAAGTC   2750
  AGGGATACACATAGTACATCTCAGGAAGTTTTAGAAAGTTTGGATCCACC   2800
  TTGTTAATCTGCATACGACATTCTTTCACACCTGCAAGGAGACTTTCATT   2850
  TCACATTTGCAAGGTGAAGGTGAGGCCCTTGCAGGGGATGGGGATGGGTA   2900
  GAGCAGTGTCTGAGGTAGGTTTGAGCCTTTAAATGTGTGGCATCTGTGGA   2950
  GATGTGCATGTTGTCAGGAGGCCAGGGAGAGCACCCTTCCCCAGATCCAC   3000
  ATCCCTGTGTACCTTGAGCCTGAGCAGAGACCAGCCCCTGGCCTGGCCCC   3050
  AGCACTGCTCAGGCAGAGGCATGTGTTGCCCCTGCATCTGCCCTGAGAAC   3100
  CCCTCTGTCAAGCATGAAGGACTGAACAGCATGTGGAATGCCAGAAAAGA   3150
  TCATCCTTCCCCATCCAGCCCTTCCCTGCATTGCCCTAGCCCACTGCACC   3200
  TCTGAGCTCTCCTAATGAAGGATAGATAAGTGAGCTGCCCTCCTGCCTGC   3250
  CCCTCCTGCTCCCAGGGTCCCCTGCCTGGTCTAACCTTCTAAGTGACTGC   3300
  CCATGGGGACAGATTCTGGTGCATGGTACTGTCTGGTATGTGTTTTCCCT   3350

      rLeuAspSerProArgSerLysArgCysGlyAsnLeuSerThrCys   (91)
  GCAGCCTGGACAGCCCCAGATCTAAGCGGTGTGGTAATCTGAGTACTTGC   3400

  MetLeuGlyThrTyrThrGlnAspPheAsnLysPheHisThrProGl  (108)
  ATGCTGGGCACATACACGCAGGACTTCAACAAGTTTCACACGTTCCCCCA   3450

  nThrAlaIleGlyValGlyAlaProGlyLysLysArgAspMetSerSerA (125)
  AACTGCAATTGGGGTTGGAGCACCTGGAAAGAAAAGGGATATGTCCAGCG   3500

  spLeuGluArgAspHisArgProHisValSerMetProGlnAsnAlaAsn (141)
  ACTTGGAGAGAGACCATCGCCCTCATGTTAGCATGCCCCAGAATGCCAAC   3550

4 TAAACTCCTCCCTTTCCTTCCTAATTTCCCTTCTAGCATCCTTCCTATAA   3600
  CTTGATGCATGTGGTTTGGTTCCTCTCTGGTGGCTCTTTGGGCTGGTATT   3650
  GGTGGCTTTCCTTGTGGCAGAGGATGTCTCAAACTTCAGATGGGAGGAAA   3700
  GAGAGCAGGACTCACAGGTTGGAAGAGAATCACCTGGGAAAATACCAGAA   3750
  AATGAGGGCCGCTTTGAGTCCCCCAGAGATGTCATCAGAGCTCCTCTGTC   3800
  CTGCTTCTGAATGTGCTGATCATTTGAGGAATAAAATTATTTTTCCCCAA   3850
  AGATCTGAGCTGTGGTGGTCATTGCTCTGATCTATGTCCCAGGCTTCATA   3900
  GTGTCTAAGACCTATGCTTAGAAATAGCCTTAACCCTAGGCTAGCTGGAC   3950
  AGAGGATATGGTGGGTGGTCCCTTTGACCAAGCTCAAGCAGGAAGAACAG   4000
  GGGTCCTAAGGAGCAGGTAAGCACCTCTAGGACTTGATGCTGCAAACTCC   4050
  GCTCCTCTTCCAGGTAAGACTGAGGAATTTTTTATTTTCCTAAGAAAGGG   4100
  TATTTGGTGCCCGTGACTGGGGTGTAGATTTTATAGTCCTTTGTGAATGG   4150
  GGCTGGGTGTGGGACCATAATTCACTCCAGTGTCATAAACCTCCGCTTTG   4200
```

```
  ATTTTTAGTTAATTTATACAGGAAAGATTGGCTGTTACTGCTCCACATTC   4250
  CATAGCCAGTCATCCAGAGTCACCTTGGGTTTTCTGACACCCCTGGGAAT   4300
  ATCTATGGGGAGTGATCATGGCATTTTCCCTAATGGCCTTGTGATTTTCT   4350
  GCTCTGATAATTGTGTTTAGGAGAAACACTTAAAGTTAATTGGTGCCTTT   4400
  CAGCACAGCAACTTTACCATGAAGGTCCATGGGGCTGACCTCTCTCCCAG   4450

                                               gIle   (77)
  CCTCTCACTCACAGATCTTCTCTTCTTTCTCCATCCTGCAAATCAGAATC   4500

  IleAlaGlnLysArgAlaCysAspThrAlaThrCysValThrHisArgLe  (94)
  ATTGCCCAGAAGAGAGCCTGTGACACTGCCACCTGTGTGACTCATCGGCT   4550

  uAlaGlyLeuLeuSerArgSerGlyGlyValValLysAsnAsnPheValP (111)
5 GGCAGGCTTGCTGAGCAGATCAGGGGGTGTGGTGAAGAACAACTTTGTGC   4600

  roThrAsnValGlySerLysAlaPheGlyArgArgArgArgAspLeuGln (127)
  CCACCAATGTGGGTTCCAAAGCCTTTGGCAGGCGCCGCAGGGACCTTCAA   4650

  Ala                                                (128)
  GCCTGAGCAGCTGAATGACTCAAGAAGGTGACTGCCCTTGTATGATGGGA   4700

  TGGGAAGATGAATGACTGGTTTTTACTGGGGTGTAAAACCACTCTGAGCC   4750
  TCTCTGAGACCATGTGGTTTTAAAAAATCCATAAGGGAAGGTACCCACAC   4800
  CAGTATCTGAGTTCCAGTAGCTAAGACCCTAGAATTTGGATTCATCTCTG   4850
  TTTTTTCATGTCTCTCCTTGTAACCCTGAGATCATCAGACCAAGAAATAC   4900
  AGATCCTGTTTATTTAGAACACTGCTGTTTGACATTTATTAATTTTGATT   4950
  ATTCTAGCCTTTGAGTTTGAAAGATAATGACACTATCATTTTAGAGTATA   5000
  CATTACTATGTTGTGTACCTCTGAATTAAGTGTACAGACTTGAGTTCAAG   5050
  AACTGTATTTGCCCTTTACCAGCTATATGACCATGAACAGGTTACTTAAC   5100
  TCTCTCCAAGCCTCAGTTTCTCCATCTGTAAATTGAGGGCTACAATAGTA   5150
  CCTACCTCCAAGCATTACTATAAAGAGCAAGTGAGGTAATAGATGTTAAG   5200
  TCTGACAATTAACCAGTAACTAAATTCTAGCTGTTATTTTCTTCCTCCTA   5250
  GGTCACAATAAAGCTGAACTCCTTTTAATGTGTAATGAAAGCAATTTGTA   5300
  GGAAAGGCTCCATGGAAGACATACATATAGGCATCCTTCTTGATACTGAA   5350
  AACTATCTTCTTTGTTTGAAAGGAACTATTGCTAAATGCAGAACAAGCTC   5400
  ATTGCAGTTACCTATTGTGCATCTTTTTAAATACTTGATTATGTAACCAT   5450
6 AAATCTGACAGCATGTCTCATTGGCTTATCTGGTAGCAAATCTAGGCCCC   5500
  GTCAGCCACCCTATTGACATTGGTGGCTCTGCTAAACCTCAGGGGGACAT   5550
  GAAATCACTGCCTCTTGGGCATCTGGGGACACATGGTAATGCTGTGCCTT   5600
  GACAGAAGTATTTGTTTAAAGAAATGTCAATGCTGTCATTTGTGAACTCT   5650
  ATCAAAATTAAAAATGTATTTTCTATACCCTTCAATGGAATCTCTGCTGC   5700
  TATTTATGCTATTTTTCCCAGGAGAGTTCAGGGGCCTGTAAGCACTCTGT   5750
  TGACCAACCAATCTGACTCACTTAGCACAATGACAACATGCCTGGGTAGG   5800
  ATCC                                                 5804
```

Figure 3: *Sequence of the human calcitonin/α-CGRP gene and 5′ flanking region*

The sequence shown extends from the 5′ insert-vector boundary of cosCT1 to a *Bam* HI site 120bp downstream of exon 6. The sequence is numbered relative to the transcription start site, the numbers referring to the nucleotides at the end of each line. The six exons of the gene are boxed and numbered on the left. Exon 1 has two different 3′ ends depending on whether the upstream (major) 5′ splice site at +108 or downstream (minor) 5′ splice site at +132 is used in intron 1 splicing (32). The protein sequences of preprocalcitonin and preproCGRP are shown, and numbered in brackets on the right. The translation products of exons 2 and 3 are common to both precursors,exon 4 is specific to preprocalcitonin and exon 5 is specific to CGRP. Note that splicing of exon 3 to exon 5 generates an Arg at position 76 of preproCGRP. Consensus sequences in the promoter which might interact with transcription factors are boxed. These are a 'TATA' box (43) AATAA (−25 to −29), an Spl binding site (44) CCGCCC (−77 to −82), an octamer (45) ATGCAAAT (−160 to −167), a cAMP-responsive element GTGACGCA (−104 to −111) with one copy of the concensus TGACG (46), a cAMP-responsive element TGACGTCA (−253 to −260) with two inverted copies of the concensus TGACG, and a sequence TGACCTCA (−168 to −175) with

Intron sequences, apart from a few bases of consensus sequence at intron-exon boundaries, were not reported by these groups, although the sequence of intron 4, as derived from a partially-processed mRNA molecule, has been previously determined in this laboratory (7). The complete sequence of the exon-intron structure of the gene and 1.84kb of 5′ flanking sequence (up to the vector:insert boundary in cosCT1) was determined (Figure 3). There was almost complete agreement of exon sequences with previously published data (5,7,30,31,32; —see legend to Figure 3 for differences). Sequence analysis confirmed that within intron 2 there is an Alu repeat which is inverted with respect to the gene (see Figure 3 legend).

The dinucleotide CpG is under-represented in human genomic DNA except in CpG islands. These islands are regions of unmethylated G+C rich DNA where the level of CpG approaches that predicted from the G+C content (15,16). The presence of a cluster of *Hpa* II sites and rare 6-cutter sites containing the CpG dinucleotide at the 5′ end of the insert of cosCT1 suggested the presence of a CpG island (Figure 1). Figure 4 presents an analysis of 3.85kb, from the 5′ insert boundary of cosCT1 (−1.85kb relative to the transcription start site of the calcitonin/α-CGRP gene) to +2.0kb from the transcription start site. GpC dinucleotides are distributed throughout this sequence. The entire region upstream of the middle of intron 1 is CpG-rich. Within this region CpG dinucleotides are clustered around exon 1 and in an upstream region between −0.8 and −1.8kb. These two CpG-rich subregions correspond to peaks of G+C richness. A plot of observed/expected CpG frequency, calculated according to Gardiner-Garden and Frommer (33) shows these two CpG-rich subregions most clearly. Gardiner-Garden and Frommer (33) define CpG-rich regions as having a percentage G+C of greater than 50 and an observed/expected CpG of greater than 0.6. The upstream subregion (−1.25kb to −1.85kb, average G+C =59%, average observed/expected CpG =0.76) and the subregion around exon 1 (0 to $^{+}$0.55kb, average G+C=61%, average observed/expected CpG =0.66) fulfil these criteria. Between these subregions the observed/expected CpG drops down to around the genomic average (0.2−0.25) and between −0.4kb and −0.7kb the percentage G+C drops down to 48.

Sequence analysis thus reveals the presence of a CpG-rich region at the 5′ end of the calcitonin/α-CGRP gene. To test whether *Msp* I sites in the vicinity of the CpG-rich region were unmethylated, DNA fragments generated by *Msp* I and *Hpa* II digestion were analyzed by Southern blotting. DNA was analysed from sources which do not express the calcitonin/α-CGRP gene (liver and lymphocytes) and two sources which do express the calcitonin/α-CGRP gene; medullary carcinoma of the thyroid (MCT), a tumour of the C-cells of the thyroid, and the human lung carcinoma cell-line BEN, which ectopically expresses the gene (32,34). A series of probes spanning the 5′ flanking sequence, exon

homology to TPA-responsive elements (47). Three Spl binding sites within the upstream CpG-rich region (centred on −1241, −1430 and −1573) are also boxed. The 19 *Msp* I sites within the the sequenced region are underlined (see Figure 5C for a map of these sites). The following differences between the exon sequences of this clone and previously published cDNA sequences (7) were observed: in exon 4 the T at position 3382 is a C in cDNA (this is a silent substitution in codon 51 of preprocalcitonin) and the A at 3585 substitutes for a T in cDNA; in intron 4 the T at position 4290 is absent in cDNA and this clone lacks a C present in cDNA between positions 4426 and 4427; in exon 6 the T at 5434 is an A in cDNA, the T at 5363 is absent from cDNA, the C at 5673 is an A in cDNA and the A at 5675 is a C in cDNA. An Alu repeat, inverted relative to the orientation of the gene, is present in intron 2 (positions 1452−1734). This repeat has a homology of 89% to the concensus of Kariya *et al.*(48).

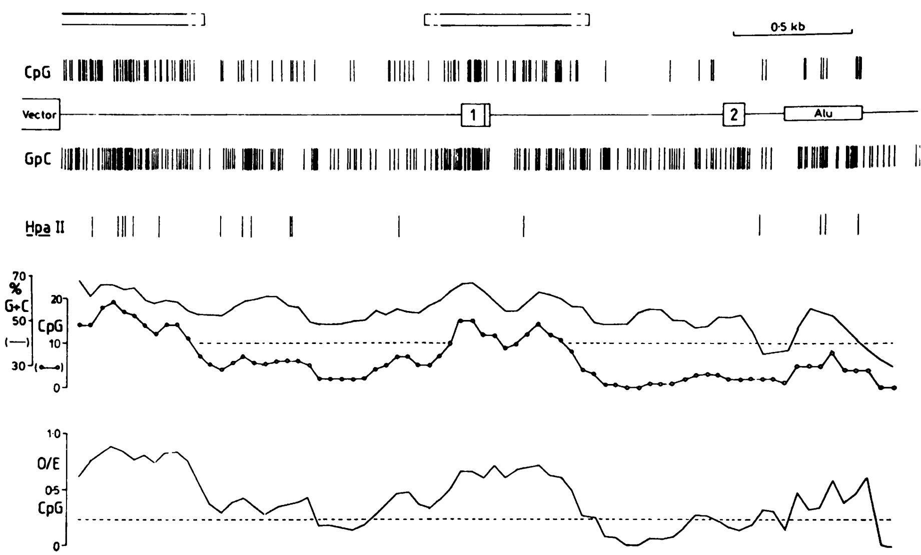

Figure 4: *CpG-rich regions at the human calcitonin/α-CGRP gene*
An analysis of 3.85kb of genomic DNA is shown. The positions of CpG and GpC dinucleotides, and of *Hpa* II sites , are marked. Below this is shown a graphical representation of G+C frequency and CpG frequency (per 200bp in 50bp steps); the dotted line at 40% G+C shows the average base composition of mammalian genomic DNA. Below this is plotted the observed over expected (O/E) frequency of CpG (per 200bp in 50bp steps); the dotted line at 0.25 shows the average O/E CpG for genomic DNA. The two CpG-rich regions revealed by graphical analysis are marked at the top of the figure, one encompassing exon 1 and the other extending from the vector-insert boundary to around −1.2kb.

1 and part of intron 1 were hybridised to Southern blots. None of these probes cross-hybridises to any other sequences in the human genome at high stringency.

Figure 5A shows the pattern resulting from hybridisation of a 0.73kb *Hin*d III-*Bgl* II fragment (probe I). Bands of approximately 0.28 and 0.18kb were observed in *Hpa* II and parallel *Msp* I digests. A faint band of 0.1kb can be seen in the MCT tracks and is visible in the other tracks on longer exposure. These are the bands expected if the *Msp* I sites 5−11 in Figure 5C are unmethylated, with the 0.1kb band being a doublet of 116 (sites 5−6) and 102 (sites 7−8). Although the 40bp fragment separating sites 7 and 8 was too faint to be seen on this blot, the identical size of the 0.1 and 0.18kb bands in both *Hpa* II and *Msp* I digests suggests that neither of these bands is methylated. The additional band of 0.4kb observed in liver was also seen in *Msp* I digests and so presumably does not reflect a partial *Hpa* II digest. The alternative explanation is that one of the *Msp* I sites, probably site 5 is polymorphic. Using probe I and adjacent fragments as hybridisation probes, bands of the expected size (predicted from the sequence of Figure 3) were observed in both *Hpa* II and *Msp* I digests (data not shown). These results indicate that *Msp* I sites 1−13 are unmethylated in all of the DNA samples examined, although some of the *Msp* I sites in this region are so close together that some methylation of these sites might not be detected by Southern blotting. *Msp* I sites 1−13 span the CpG-rich region of DNA extending from the upstream vector-insert boundary in cosCT1 down into intron 1 of the calcitonin/α-CGRP gene.

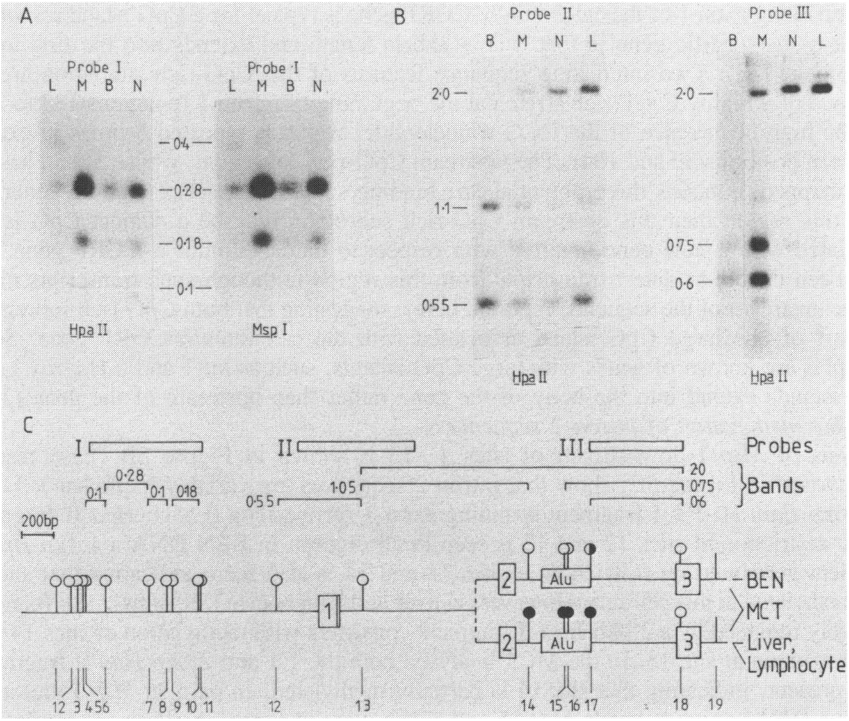

Figure 5: ***Methylation status of Msp I sites in the calcitonin/α-CGRP gene***

A Unmethylated *Msp* I sites in the CpG-rich region. An autoradiograph of a Southern blot hybridised to a 0.73kb *Hind* III-*Bam* HI fragment (probe I: positions −1463 to −729) is shown. Sizes of bands are shown in kb. DNA samples are BEN(B), MCT(M), liver(L) and lymphocyte(N). Parallel *Hpa* II and *Msp* I digests are shown.

B Variable methylation of *Msp* I sites in intron 2. Autoradiographs of Southern blots of *Hpa* II-digested DNA hybridised to a 0.76kb *Bam* HI-*Pst* I fragment (probe II: positions −139 to +619) and a 0.58kb *Msp* I fragment (probe III: positions 1728 to 2303) are shown. The larger bands correspond to those shown in Figure 5C and are due to partial or complete methylation of *Msp* I sites in intron 2.

C Summary of methylation data. A map of the region from the cosmid-vector boundary of cosCT1(−1.85kb) to the centre of intron 3 is shown. All the *Msp* I sites within the sequenced region are on this map and are numbered 1−19. Open circles show unmethylated sites, partially filled circles partially methylated sites and completely filled circles completely methylated sites. Site 19 was not detected in any of the genomic DNA samples and is presumably polymorphic. Above the map are shown the positions of probes I−III and the origins of the bands seen in Fig. 5A and 5B (sizes in kb).Sites 14−17 exhibit a different pattern of methylation in liver and lymphocyte as compared to BEN and MCT. This is indicated by the divergence in the downstream region of the map.The upper pattern is seen in BEN, the lower pattern is seen in liver and lymphocytes and a combination of the two is seen in MCT. The *Msp* I sites shown here agree with the map of Baylin *et al.*(36) if sites 1−9 (here) are equivalent to site M1 (36), sites 10 and 11 to M2, site 12 to M3, site 13 to M4, sites 14−17 to M5 and sites 18 and 19 to M6.

CpG islands are unmethylated irrespective of the expression of the associated gene (15). Thus, these results confirm that the CpG-rich region at the 5′ end of the calcitonin/α-CGRP gene is a CpG island. On close analysis, this CpG island appears to be composed of two distinct CpG-rich subregions. This is an unusual strucure for a CpG island. The CpG-rich subregions could be regarded as separate islands. Indeed the CpG-rich subregion encompassing exon 1 of the calcitonin/α-CGRP gene is typical for a CpG island associated with a tissue-specific gene in that it is ~1kb in length,and extends into the first intron of the gene (33). Two interesting sequence features of this CpG-rich subregion are the presence of a highly C+T-rich stretch at the beginning of intron 1 (positions 138 to 203) and the high occurrence of the CCG trinucleotide, which is repeated 8 times in exon 1 (between positions 30 and 104). The upstream CpG-rich subregion, whose 5′ end has not been mapped, contains three potential Sp1 binding sites. If transcripts can be generated from this region then this upstream CpG-rich subregion may be a distinct CpG island associated with a new gene inverted with respect to the calcitonin/ α-CGRP gene. We have been unable to detect transcripts from this region (although such transcripts might initiate upstream of the sequence reported here), suggesting that both CpG-rich subregions are part of one large CpG island associated with the calcitonin/α-CGRP gene. Some examples are known of genes with large CpG islands, such as *int*-1 and c-Hα-*ras* 1, but these islands extend into the body of the gene rather than upstream of the gene (33).

Variable methylation of intron 2 sequences

Analysis of *Msp* I downstream of sites 1−13 is shown in Figure 5B.These results, summarised in Figure 5C, show that intron 2 sequences are variably methylated. Using a 0.76kb *Bam* HI-*Pst* I fragment spanning exon 1 (probe II) a 0.55kb *Hpa* II fragment due to restriction at sites 12 and 13 is seen in all digests. In BEN DNA a 1.1kb *Hpa* II fragment generated by restriction at sites 13 and 14 is also seen, indicating that site 14 is unmethylated in this cell-line. However, in liver and lymphocyte DNA this 1.1kb fragment is largely replaced by a 2.0kb *Hpa* II fragment consistent with methylation of sites 14−17 and restriction at site 18. In the MCT analysed both the 1.1 and 2.0kb *Hpa* II fragments were present, indicating that site 14 is partially methylated. In parallel *Msp* I digests of the four DNA samples, bands of 0.55 and 1.1kb were observed in all cases, confirming that the differences observed were not due to polymorphisms. In order to analyze the methylation state of sites 15−17 a 0.56kb *Msp* I fragment from intron 2 was used as a hybridisation probe (probe III, Figure 7B). In liver and lymphocytes a 2.0kb *Hpa* II fragment was again observed, consistent with methylation of sites 14−17. In BEN DNA two *Hpa* II fragments, of 0.6 and 0.75kb, were found. This indicates that site 17 is partially methylated, the 0.6kb fragment presumably arising from restriction at sites 17 and 18 (when site 17 is unmethylated) and the 0.75kb fragment arising from restriction at sites 15 (or 16) and 18 (when site 17 is methylated). As summarised in Figure 5C, the *Msp* I sites in intron 2 are predominantly unmethylated in the calcitonin-expressing cell-line BEN. Sites 15 and 16 are both marked as being unmethylated in BEN DNA, though it should be noted that partial methylation of either of these sites would not be detected by this analysis since the sites are so close together. In MCT a combination of the patterns seen in liver and lymphocyte and in BEN was again observed. Taken together with the results obtained with probe II it appears that there are two distinct patterns of methylation in the MCT analysed; one of these is the liver/lymphocyte pattern of complete methylation of sites 14−17 and the other is the predominantly unmethylated pattern observed in BEN. Thus in Figure 5C MCT is shown as having a combination of the two methylation patterns described. Overall these results indicate that both eutopic expression of the calcitonin/α-CGRP gene (in MCT) and ectopic expression in a lung tumour cell-line are associated with demethylation of intron 2 sequences. The presence of some methylation of intron 2 in the MCT sample analysed may be due to contamination of the tumour by cells of non-C-cell origin; these cells, if they are not expressing the calcitonin/α-CGRP gene, would be expected to exhibit the intron 2 methylation pattern found in liver and lymphocytes.

Methylation and calcitonin/α-CGRP gene transcription

The sequence data and methylation studies reported here suggest that at least three components may contribute to tissue-specific regulation of human calcitonin/α-CGRP gene expression. Firstly, the CpG island at the 5′ end of the gene may affect the accessibility of the promoter to transcription factors. Since CpG islands are associated with all housekeeping genes so far characterised (15,33), it has been suggested that these islands allow access of transcription factors to the promoters (16). The activation of tissue-specific genes may require the establishment of methylation-free regions around the promoter (35). It may be the case that tissue-specifically expressed genes which already possess CpG islands, such as the calcitonin/ α-CGRP gene, are actively repressed in non-expressing tissues (16). Previously Baylin *et al.* (36,37) have reported studies on the methylation of the calcitonin/α-CGRP gene in a variety of tissue and tumour types. Although they map fewer *Msp* I sites than reported here their partial map is equivalent to the map in Figure 5C if the sites they report correspond to clusters of sites in Figure 5 (see legend to Figure 5). They find, in agreement with us, that sites at the 5′ end of the gene are unmethylated in normal tissues. Interestingly, they show that in small-cell lung carcinoma cell-lines the 5′ region of the gene frequently becomes methylated. In the context of our analysis this implies that in tumours CpG islands can become methylated. This phenomenon may be related to the ectopic activation of calcitonin/α-CGRP gene expression in small-cell lung carcinoma.

Secondly,the demethylation of sequences in intron 2 associated with calcitonin/α-CGRP gene expression may be a requirement for the expression of the gene. Demethylation of CpG dinucleotides correlates with the activation and concomitant alteration of chromatin structure of many tissue-specific genes (38).The demethylation of specific sequences of DNA near some genes has been associated with binding of nuclear factors to those sequences (39). Thus sequences within the body of the calcitonin/α-CGRP gene may be involved in the activation and maintenance of calcitonin/α-CGRP gene expression. Delineation of these sequences will at first require a more exact analysis of methylation changes by using genomic sequencing. Interestingly, similar methylation patterns were observed in both MCT and in the BEN cell-line. This suggests that activation of calcitonin/α-CGRP gene expression may occur by similar mechanisms in both normal and ectopic situations.

Thirdly, the interaction of transcription factors with promoter sequences may contribute to regulation of calcitonin/α-CGRP gene expression. Several transcription factor recognition sequences, including the octamer and two cAMP-responsive elements, are present in the promoter of the calcitonin/α-CGRP gene (see Figure 3). Of particular interest is the possibility that ectopic expression of the calcitonin/α-CGRP gene, which has been documented for many carcinomas (40,41), might be due to oncogenic alterations in transcription factors. Recent discoveries that the cellular homologues of some oncogenes are transcription factors (42) suggests that tumourigenesis arising in part through mutation in transcription factors might also lead to ectopic expression of other genes recognised by such factors. The contribution of consensus sequence elements in the promoter to normal and ectopic calcitonin/ α-CGRP expression is currently being assessed.

Conclusion

The human calcitonin/α-CGRP gene locus has been analysed in detail. Restriction mapping of cosmid clones revealed a novel *Pvu* II RFLP. The entire exon-intron structure of the gene and 1.8kb of 5′ flanking region have been sequenced. This revealed a large CpG island which extends from at least −1.8kb to the centre of intron 1. This island is not uniformly CpG-rich. Instead CpG dinucleotides are clustered around exon 1 and around −1.5kb. However these two regions do not appear to be distinct CpG islands. *Msp* I sites within the island were unmethylated irrespective of calcitonin/α-CGRP gene expression, whereas methylation of *Msp* I sites within intron 2 correlated with expression of the gene. These results suggest that activation of calcitonin/α-CGRP gene expression may be related to both the presence of a CpG island and demethylation of internal regions of the gene.

ACKNOWLEDGEMENTS

We thank Dr. D. Kioussis for the HPB-ALL cos202 library and for advice on cosmid library screening, and Miss P. Harrington for excellent technical assistance. This work was supported by the Cancer Research Campaign. R.V.T. was supported by an MRC Training Fellowship.

*To whom correspondence should be addressed

Present addresses: +Division of Molecular Medicine, Clinical Research Centre, Watford Road, Harrow, Middlesex HA1 3UJ and §Department of Biotechnology, Research Division I, ICI Pharmaceuticals, Mereside, Alderley Park, Macclesfield, Cheshire SK10 4TG, UK

REFERENCES

1. Amara, S.G., Jonas, V., Rosenfeld, M.G., Ong, E.S. and Evans, R.M. (1982) Nature *298*, 240−244.
2. Amara, S.G., Evans, R.M. and Rosenfeld, M.G. (1984) Mol. Cell Biol. *4*, 2151−2160.
3. Rosenfeld, M.G. Amara, S.G. and Evans, R.M. (1984) Science *225*, 1315−1320.
4. Nelkin, B.D., Rosenfeld, K.I., de Bustros, A., Leong, S.S., Roos, B.A. and Baylin, S.B. (1984) Biochem. Biophys. Res. Commun. *123*, 648−655
5. Steenbergh, P.H., Hoppener, J.W.M., Zandberg, J., Lips, C.J.M. and Jansz, H.S. (1985a) FEBS Lett *183*, 403−407.
6. Steenbergh, P.H., Hoppener, J.W.M., Zandberg, J., Visser, A., Lips, C.J.M. and Jansz, H.S. (1986) FEBS Lett. *209*, 97−103.
7. 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.
8. 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.
9. Minivielle, S., Cressent, M., Lasmoles, F., Julliene, A., Milhaud, G. and Moukhtar, M.S. (1986) FEBS Lett. *203*, 7−10
10. Steenbergh, P.H., Hoppener, J.W.M., Zandberg, J., Cremers, A.F.M., Jansz, H.S. and Lips, C.J.M. (1985b) in Calcitonin (A. Pecile, ed.) pp23−31 (Elsevier Science Publishers).
11. Craig, R.K., Schifter, S., Broad, P.M., Riley, J.H., Edbrooke, M.R. and Marshall, I. (1986) in Neuroendocrine Perspectives 5 (Elsevier) Muller, E.E. and MacLeod, R.M., eds..
12. Alevizaki, M., Shiraishi, A., Rassool, F.V., Ferrier, G.J.M., MacIntyre, I. and Legon, S. (1986) FEBS Lett *206*, 47−52.
13. Hoppener, J.W.M., Steenbergh, P.H., Zandberg, J., Adema, G.J., Geurts van Kessel, A.H.M., Lips, C.J.M. and Jansz, H.S. (1988) FEBS Lett *233*, 57−63.
14. 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.
15. Bird, A.P. (1986) Nature *321*, 209−213.
16. Bird, A.P. (1987) Trends Genet. *3*, 342−347.
17. Kioussis, D., Wilson, F., Daniels, C., Leveton, C., Taverne, J. and Playfair, J.H.L. (1987) EMBO J. *6*, 355−361
18. Grosveld, F.G., Dahl, H.H.M., de Boer, E. and Flavell, R.A. (1981) Gene *13*, 227−237.
19. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA *74*, 5463−5467.
20. Sanger, F., Coulson, A.R., Barrel, B.G., Smith, A.J.H. and Roe, B.A. (1980) J. Mol. Biol. *143*, 161−178.
21. Broad, P.M., Schifter, S. and Craig, R.K. (1987) Br. J. Cancer *155*, 175−177.

22. Southern, E. (1975) J. Mol. Biol. *98*, 503–517.
23. Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P. (1977) J. Mol. Biol. *113*, 237–251.
24. Feinberg, A.P. and Vogelstein, B.A. (1983) Anal. Biochem. *132*, 6–13.
25. Botstein, D., White, R.L., Skolnick, M. and Davis R.W. (1980) Am. J. Hum. Genet. *32*, 314–331.
26. Hoppener, 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.
27. Solomon, E. (1984) Nature *309*, 111–112.
28. Fearon, E.R., Feinberg, A.P., Hamilton, S.H. and Vogelstein, B. (1985) Nature *318*, 377–380
29. Scrable, H.J., Witte, D.P., Lampkin, B.C. and Cavenee, W.K. (1987) Nature *329*, 645–647.
30. Craig, R.K. Hall, L., Edbrooke, M.R., Allison, J. and MacIntyre, I. (1982) Nature *295*, 345–347.
31. Le Moullec, J.M., Jullienne, A., Chenais, J., Lasmoles, F., Guiliana, J.M., Milhaud, G. and Moukhtar, M.S. (1984) FEBS Lett. *167*, 93–97
32. Riley, J.H., Edbrooke, M.R. and Craig, R.K. (1986) FEBS Lett. *198*, 71–79.
33. Gardiner-Garden, M. and Frommer, M. (1987) J. Mol. Biol. *196*, 261–282.
34. Ham, J., Ellison, M.L. and Lumsden, J. (1980) Biochem. J. *190*, 545–550.
35. Murray, E.J. and Grosveld F. (1987) EMBO J. *6*, 2329–2335.
36. Baylin, S.B., Hoppener, J.W.M., de Bustros, A. Steenbergh, P.H., Lips, C.J.M. and Nelkin, B.D. (1986) Cancer Res. *46*, 2917–2922.
37. Baylin, S.B. Fearon, E.R., Vogelstein, B., de Bustros, A., Sharkis, S., Burke, P.J., Staal, S.P. and Nelkin, B.D. (1987) Blood *70*, 412–417.
38. Cedar, H. (1988) Cell *53*, 3–4.
39. Becker, P.B., Ruppert, S. and Schutz, G. (1987) Cell *51*, 435–443.
40. Baylin, S.B. and Mendelsohn,, G. (1980) Endocrine Reviews *1*, 45–77.
41. Silva, O.L., Becker, K., Primack, A., Doppman, J. and Snider, R. (1974) N. Eng. J. Med. *290*, 1122–1124.
42. Bohmann, D., Bos, T.J., Admon, A., Nishimura, T., Vogt, P.K. and Tjian, R. (1987) Science *238*, 1386–1392.
43. Breathnach, R. and Chambon, P. (1981) Ann. Rev. Biochem. *50*, 349–383.
44. Dynan, W.S. and Tjian, R. (1983) Cell *33*, 669–680.
45. Falkner F.G. and Zachau H.F. (1984) Nature *310*, 71–73.
46. Tsukada, T., Fink, J.S., Mandel, G. and Goodman, R.H. (1987) J. Biol. Chem. *262*, 8743–8747.
47. Angel, P. Imigawa, M., Chiu, R., Stein, B., Imbra, R.J. Rahmsdorf, H.J., Jonat, C., Herrlich, P. and Karin, M. (1987) Cell *49*, 729–739
48. Kariya, Y., Kato, K., Hayashizaki, Y., Himeno, S., Tarui, S. and Matsubara, K. (1987) Gene *53*, 1–10.

This article, submitted on disc, has been automatically converted into this typeset format by the publisher.