The EMBO Journal vol.9 no.3 pp.857-868, 1990

An unusual structure of a putative T cell oncogene which allows production of similar proteins from distinct mRNAs

T.Boehm, J.M.Greenberg[1], L.Buluwela[2], I.Lavenir, A.Forster and T.H.Rabbitts

Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

[1]Permanent address: University of Minnesota Medical School, Box 159 UMHQ, Harvard Street at East River Road, Minneapolis, MN 55455, USA

[2]Present address: Department of Biochemistry, Charing Cross and Westminster Medical School, Fulham Palace Road, London W6 8RF, UK

Communicated by T.H.Rabbitts

We previously identified a putative T cell oncogene on chromosome 11 near a translocation t(11;14)(p15;q11) in a human T cell tumour. The gene is transcribed from distinct promoters which have unrelated sequences, which occur within close but distinct methylation-free islands and which allow cell specific production of mRNA. The alternative first exons each contain a protein initiation codon from which two species of protein can be made, differing by only a single amino acid. The protein sequence is highly conserved between man and mouse (98%) and the same single codon difference between alternative first exons is also conserved. This is, therefore, a new form of eukaryotic gene organization from which similar proteins can be made from distinct mRNA species.

Key words: alternative initiation codon/chromosome translocation/dual promoter/exon structure/T cell receptor

Introduction

Chromosomal translocations are a common feature of lymphoid malignancies and these frequently involve antigen receptor genes (viz. immunoglobulin or T cell receptor genes in B and T cells respectively). This common association of antigen receptor genes and chromosome abnormalities is thought to reflect mistakes in the normal gene rearrangement of the receptor genes (Croce, 1987; Rabbitts *et al.*, 1988). Genes adjacent to the chromosome breakpoints on the reciprocal chromosomes are of considerable interest, not only because they are potential new oncogenes, but also because they represent a class of gene particularly susceptible to alteration of transcriptional activity as a consequence of the new mutant chromosomal context resulting from the abnormality. The study of the normal state of these genes, therefore, presents some interesting problems in genetic organization and control. Since the T cell receptor genes have been cloned, it has been possible to study genes associated with these chromosomal lesions. A number of chromosomal translocations and inversions have been examined and in the majority of cases the TCRA/D locus (chromosome band 14q11) is involved (Rabbitts *et al.*, 1988). So far none of these T cell abnormalities have been found to involve known proto-oncogenes, with the exception of t(8;14)(q24;q11) which joins TCRA with the c-*myc* gene (Shima *et al.*, 1986; Bernard *et al.*, 1988; Finger *et al.*, 1988). Indeed, a number of these breakpoint regions have not yielded evidence of transcription units (Boehm *et al.*, 1989) and thus it is possible that these loci represent proto-oncogenes which are related, not to cell growth regulation *per se*, but to cellular differentiation (Knudson, 1985) and therefore are expressed in a developmentally regulated way. One such candidate gene was first identified in association with the TCR δ gene via a translocation t(11;14)(p15;q11) in the cell line RPMI8402 (Boehm *et al.*, 1988a), which we provisionally named the 11p15 gene to await a more meaningful designation. The breakpoint in RPMI8402 has also been studied recently by McGuire *et al.* (1989). In the current paper, we present a full analysis of the 11p15 locus with reference to its expression characteristics. We show that the 11p15 gene is unique, in comparison to other known genes, in that two virtually identical proteins can be made from distinct, alternatively spliced mRNA species, in which alternative first codon exons encode alternative protein initiation codons, but similar protein sequences. This represents a novel form of eukaryotic gene organization.

Results

The 11p15 gene product is highly conserved between species

Our previous preliminary examination of a transcription unit on chromosome 11 near to the translocation t(11;14)-(p15;q11) breakpoint showed that several probes from the region detected various sized RNA species in the translocation bearing T cell line RPMI8402 (Boehm *et al.*, 1988a; and unpublished data). cDNA clones corresponding to these RNAs were isolated from a λgt10 library of RPMI8402 RNA using initially the probe, pA27 (this probe is indicated in Figure 5) and subsequently one of the cDNA clones. The nucleotide and derived protein sequences of one cDNA clone (pCRT21, illustrated in Figure 4) are shown in Figure 1. This clone derives from the major 1.4 kb mRNA species in RPMI8402 cells and encodes a protein of 156 amino acids (17.7 kd). The derived protein sequence has two methionine residues at the N-terminus and contains 15 cysteine residues, in keeping with the recently published sequence (McGuire *et al.*, 1989).

As shown below, the human cDNA probe detects a related mRNA in some mouse and rat cells. The corresponding cDNA (pN2A) was studied from a mouse neuroblastoma cell line Neuro2A (CCL131) to allow a comparison of possible functional domains in the protein. The mouse nucleotide and protein sequence are compared to the human sequence in Figure 1. This comparison reveals a startling conservation of the putative proteins in both species, there being only three amino acid differences between the two forms shown in Figure 1 (98% identity). This remarkable

© Oxford University Press

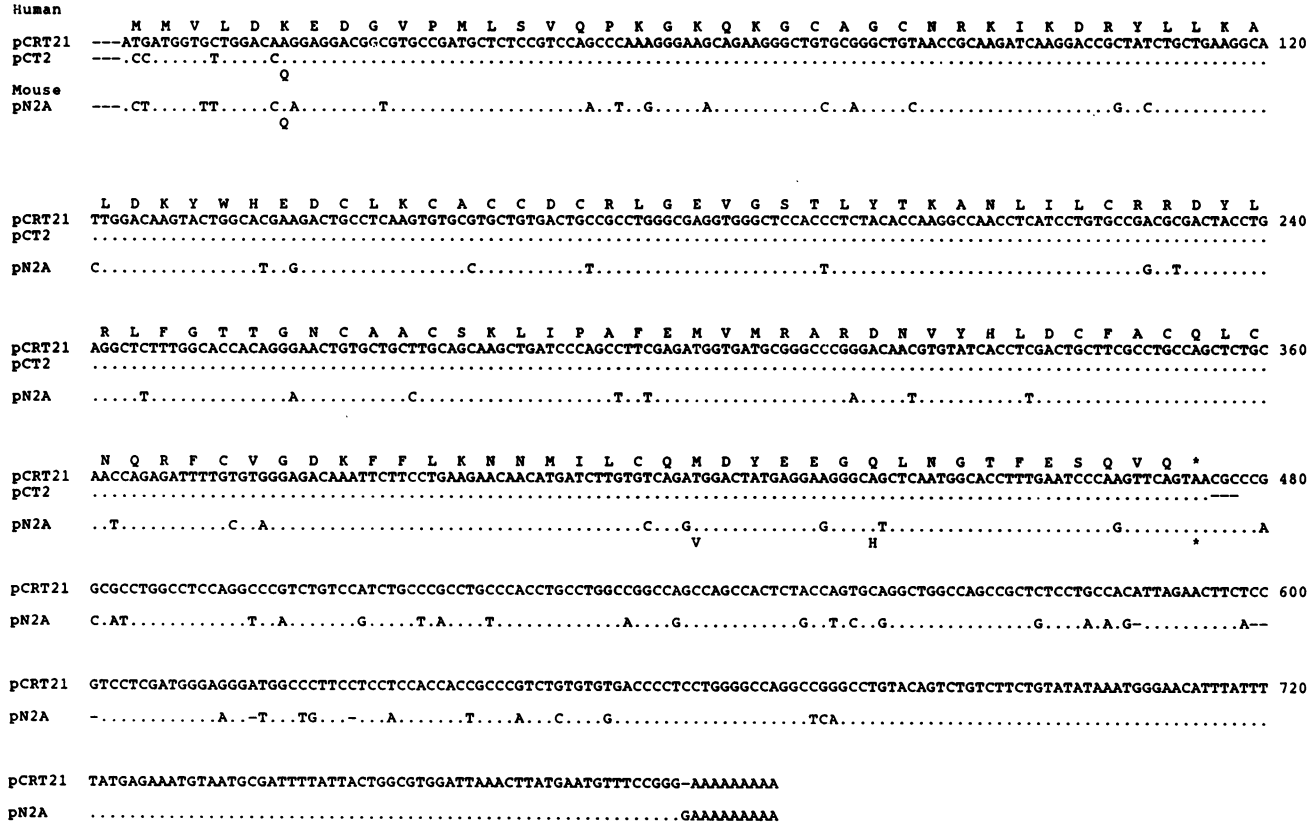

Fig. 1. Sequences of cDNA clones corresponding to the 11p15 mRNA. cDNA clones were isolated from a λgt10 library of RPMI8402 RNA or mouse neuroblastoma N2A (CCL131) RNA. The nucleotide sequence of one human clone, pCRT21, is shown from the presumed methionine initiation codon to the poly(A) tail (this clone actually extends into the 5′ untranslated region up to residue 979 of the promoter 1a sequence shown in Figure 9A). A second human clone, pCT2, is also indicated from this protein start (this clone starts at residue 235 of the promoter 1 sequence in Figure 9A) up to the translation termination codon. Note: the derived sequence of pCT2 actually contains an erroneous A residue after the third residue of exon 2; we were able to demonstrate that this residue is a reverse transcription error (see Materials and methods) and is not shown in the figure for clarity. The mouse cDNA clone, pN2A, is shown from the relevant position at the start of the protein up to the poly(A) tract. It actually extends 127 nucleotides further upstream (see Figure 9B). Derived protein sequences are shown in the single letter code; the full translation is shown for pCRT21 but only codon differences are indicated for the other two clones.

conservation at the protein level is, of course, also reflected in the nucleotide sequence, which is particularly striking in the 3′ untranslated region in which homology between species is not maintained generally. The conservation of the 11p15 gene between the species does not allow the highlighting of any distinct regions as functional domains, but is presumably indicative of a gene with an important basic function.

Unexpectedly, the sequence of a second human cDNA clone indicated a variant N-terminus for the protein (pCT2, Figure 1). Compared with the clone pCRT21, their sequences differed by a single codon (excluding the double methionine in pCRT21). Interestingly, the sequences of the mouse pN2A and the human pCT2 clones were identical in this region (Figure 1). This suggested that either a genetic polymorphism affected the 11p15 transcripts in RPMI8402 or that there is an alternative RNA splicing choice in the gene which allows the two forms to be derived. The coincidence of the mouse sequence with one form of the human mRNA suggests the latter.

Expression of the 11p15 gene in different cell types

We previously showed that a 4 kb RNA could be detected in RPMI8402 [which carries the t(11;14)(p15;q11)] using an 11p15 probe close to the translocation junction (designated pB1, see Figure 5) (Boehm *et al.*, 1988a). The nature of the 11p15 transcripts was further investigated by Northern filter hybridization of RNA of the T cell lines RPMI8402 and JM, together with RNA isolated from BALB/c mouse brain, thymus and spleen. Several different sized RNA species were detected using cDNA probe pCRT3.8 (Figure 2) in total cytoplasmic RNA from RPMI8402. These RNAs included the 4 kb species previously detected (Boehm *et al.*, 1988a) and the predominant 1.4 kb RNA species corresponding to the size of the mature mRNA. A relatively faint signal was observed in the JM RNA but there was no detectable 11p15 RNA in the mouse thymus or the other tissues (a faint background hybridization can be seen with the 28S rRNA under the low stringency hybridization conditions used). Several of this heterogeneous set of RNA species do not, however, appear to be polyadenylated (Figure 2), most notably a 3.5 kb RNA and a small RNA species seen in the RPMI total RNA. The predominant polyadenylated mRNA is the 1.4 kb species, and the 4 kb species is most likely an incompetely spliced form (further evidence for which is given below in the analysis of the cDNA clones) and not the result of heterogeneous transcription (McGuire *et al.*, 1989).

Figure 2 shows that RPMI8402 RNA contains a very much higher level of the 1.4 kb mRNA than the JM RNA

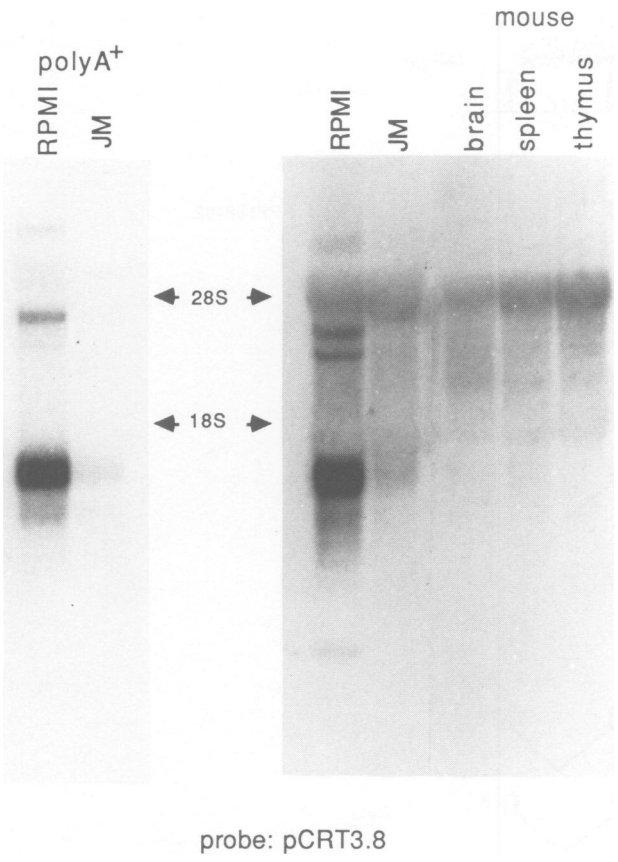

Fig. 2. Northern hybridization of human T cell and mouse tissue RNA with 11p15 cDNA clone pCRT3.8. 10 μg of total RNA or 1 μg poly(A)$^+$ RNA [selected on oligo(dT) columns] from the indicated sources were fractionated on 1.4% agarose after glyoxylation and transferred to Hybond N. The filter was hybridized and washed as described (Boehm *et al.*, 1988a) with the cDNA clone pCRT3.8 (which is illustrated in Figure 4). The age of the mouse from which the tissue RNAs were made was 4 weeks. The sizes of the two rRNA species are shown. The other T cell lines tested, and found to be negative for the presence of the 11p15 mRNA were CCRF-HSB2, HUT78, HUT102, CCRF-CEM, HPB-MLT, Peer, SUP-T1, Per 255, MOLT4, MOLT16, MOLT17 and EL4. Negative B cell lines were Daudi, WEHI 231, ACVA-1 and 6 lymphoblastoid cell lines.

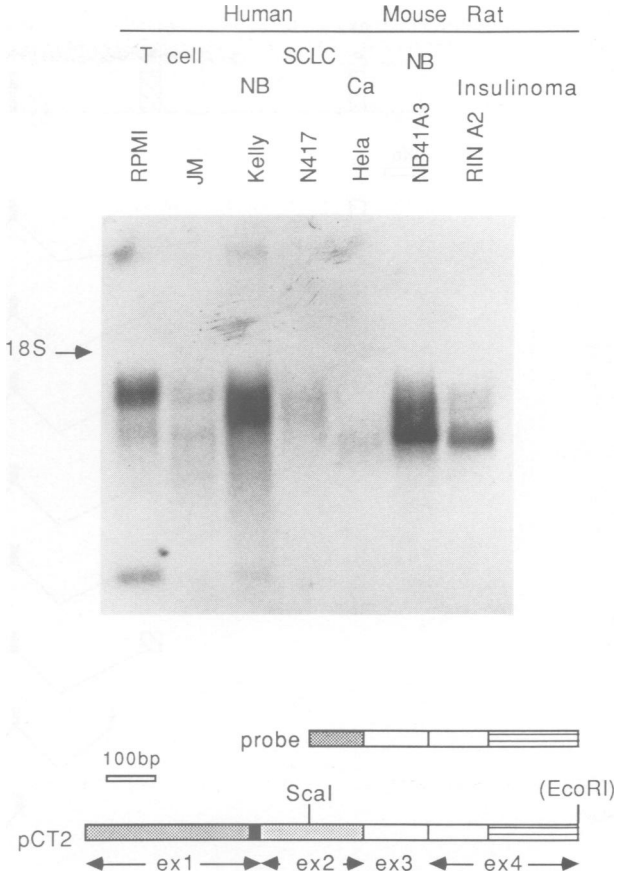

Fig. 3. 11p15 RNA in cell lines of various lineages. 10 μg of total RNA from the indicated sources were fractionated on 1.4% agarose after glyoxalation and transferred to Hybond N. The filter was hybridized and washed as described (in legend to Figure 2) with the probe shown; this was a *ScaI–EcoRI* fragment of the cDNA clone pCT2 which encompasses part of exon 2, all of exon 3 and part of exon 4 (see Figure 4). Three other human and one other mouse (Neuro 2A) neuroblastomas were positive for 11p15 RNA, as were four additional small cell lung carcinomas (SCLC). Ca refers to carcinoma. Various other cell lines were studied and found to be negative in this Northern assay: rhabdomyosarcomas CRL1598, CCL136, HTB82; melanomas M21, C32, A375, HT144; breast carcinomas MCS7, BT-474; osteogenic sarcoma CRL 1543; Wilms' tumours CRL1441, HTB48; carcinomas CCL 105 (undifferentiated); SVK14 (skin), PC3 (prostate).

which may be attributable to the different stages of T cell maturity (as judged by surface phenotype characteristics) of the cells or might be the consequence of the chromosomal translocation. No other T cell line tested showed any evidence for transcription of the 11p15 gene, including cells (eg. CCRF.HSB2) of similar surface phenotype as RPMI8402 (see legend to Figure 2 for a full list of lymphoid cell lines tested). In addition we have failed to detect the homologous mRNA in rat thymus RNA (1 month old Fischer Rat) and in mouse thymus RNA prepared from day 18 or newborn mice (in addition to that shown in Figure 2). No evidence for 11p15 mRNA was found in any B cell line tested nor in mouse spleen RNA. Thus it seems that T cells do not generally express the 11p15 gene and the lack of detectable expression in whole thymus RNA argues that, at most, a small subset of thymus cells express the gene.

Since we could not generally find 11p15 mRNA in T cells, alternative sites of transcription were sought using a panel of human and rodent cell lines for Northern hybridization. The list of cells used is given in the legend to Figure 3. Only cell lines of putative neural crest derivation ('neuro-endocrine' lineage) were found to express high levels of the mRNA (except for two carcinoma lines, including HeLa); a representative selection of these detected mRNAs are shown in the Northern hybridization in Figure 3. The 11p15 mRNA expression in the 'neuroendocrine' cell lines includes three human and two mouse neuroblastoma lines, five human small lung cell carcinoma (SCLC) lines and a rat insulinoma.

Two forms of mRNA expression from the 11p15 gene

A significant feature of the hybridization pattern shown in Figure 3 is the differences observed in the sizes of mature 11p15 mRNA detected in the various RNA populations. The hybridization of the neuroblastoma RNA (both human and mouse) appears to occur in two discrete bands (~1.4 and 1.2 kb) whereas the translocation bearing cell line has mainly the 1.4 kb species, with very small amounts of what seems to be the 1.2 kb RNA. The hybridization of the rat insulinoma RNA, on the other hand, seems to reflect the

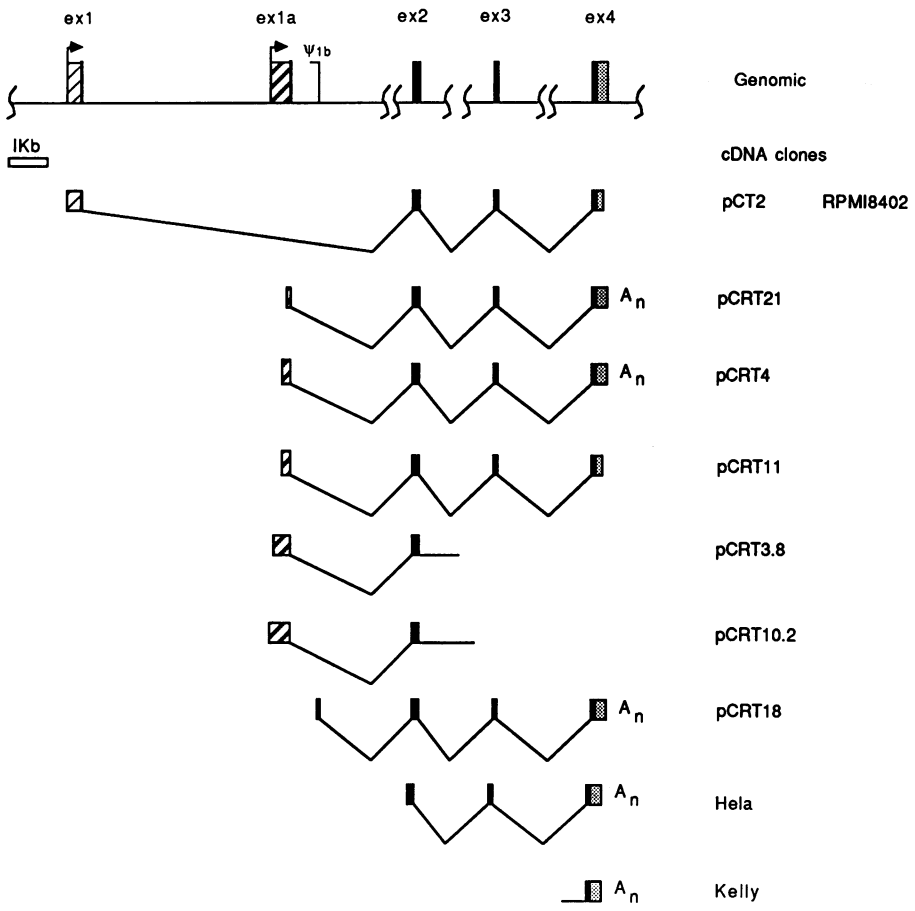

Fig. 4. Sequence organization of human cDNA clones corresponding to the 11p15 gene. At the top is a sketch of the 11p15 gene with its constituent exons (introns are not to scale). Below is the make-up of a series of cDNA clones isolated from RPMI8402 (pC clones) or from HeLa (carcinoma) or Kelly (neuroblastoma). A_n represents the presence of poly(A) at the 3′ ends of the cDNA. The splicing arrangements for the derivation of each clone are indicated. Note that, according to the arrangement of donor sites, splicing from exon 2 directly to exon 4 is formally possible but this has not been observed in any of the cDNAs analysed. Another apparently aberrant clone is pCRT18, which is a single example of a clone containing a pseudo-exon (ψ1b) and which probably results from a transcriptional read-through from a promoter and use of a cryptic splice site downstream of exon 1a. The exact start sites for cDNAs from RPMI8402 and HeLa are marked in Figures 6 and 9A. The cDNA from Kelly contains exon 4 and 460 nucleotides of intron sequences.

presence of predominantly the smaller species, and in HeLa RNA a weak signal is seen which again emanates from a smaller RNA species. There are various ways in which different size RNA molecules can be generated from one gene, including differential splicing, differential polyadenylation or alternative promoter usage. The presence of two forms of cDNA which differ at their 5′ ends, however, argues for the latter explanation. In addition, there is a clear difference in the size of the two RNAs and these RNAs are rather specific in size as seen, for example, in the c-*myc* mRNA (Hamlyn and Rabbitts, 1983).

Analysis of the 11p15 gene structure reveals two alternative first exons

The cDNA clones which were obtained from the RPMI8402 library showed a complex pattern of sequence arrangement suggesting an unusual exon structure of the 11p15 gene. cDNA clones were also examined from the human neuroblastoma line Kelly and the human adenocarcinoma HeLa and their exon content and structure examined. In conjunction with genomic cloned DNA, the organization of the 11p15 gene was determined. Figure 4 shows the organization of these various cDNA clones beneath a sketch of the gene. The detailed restriction map of the 11p15 human gene is given in Figure 5, together with various probes used to create the map, and the exon boundaries in the genomic DNA (shown in Figure 5) were determined from sequencing of genomic subclones (Figure 6).

These studies showed that the gene has five exons, of which exons 1 and 1a are different choices depending on the usage of two alternative transcription promoters (see below). Seven cDNA clones have been studied in detail from the RPMI cDNA library. One of these (pCT2, Figure 4) contains an exon designated exon 1 spliced to exons 2, 3 and 4. Five other examples contain the beginning of (pCRT10.2) or part of an alternative exon 1 designated 1a (pCRT21, pCRT4, pCRT11 and pCRT3.8) linking to exons 2, 3 and 4 in three cases or exon 2 and then intronic sequences in two cases (Figure 4). A single cDNA clone isolated from HeLa mRNA is apparently partial as it contains only exons 2, 3 and 4 and a clone from the neuroblastoma Kelly is aberrant since it only contains exon 4 and intronic sequences upstream of it (Figure 4).

These analyses of cDNA and genomic sequences establish the organization of the mRNA for the 11p15 gene. The 11p15 gene in humans is ~48 kb in length containing 5 exons. There are two promoters (see below) in the gene and the two first exons (exons 1 and 1a) are 4.7 kb apart. A long

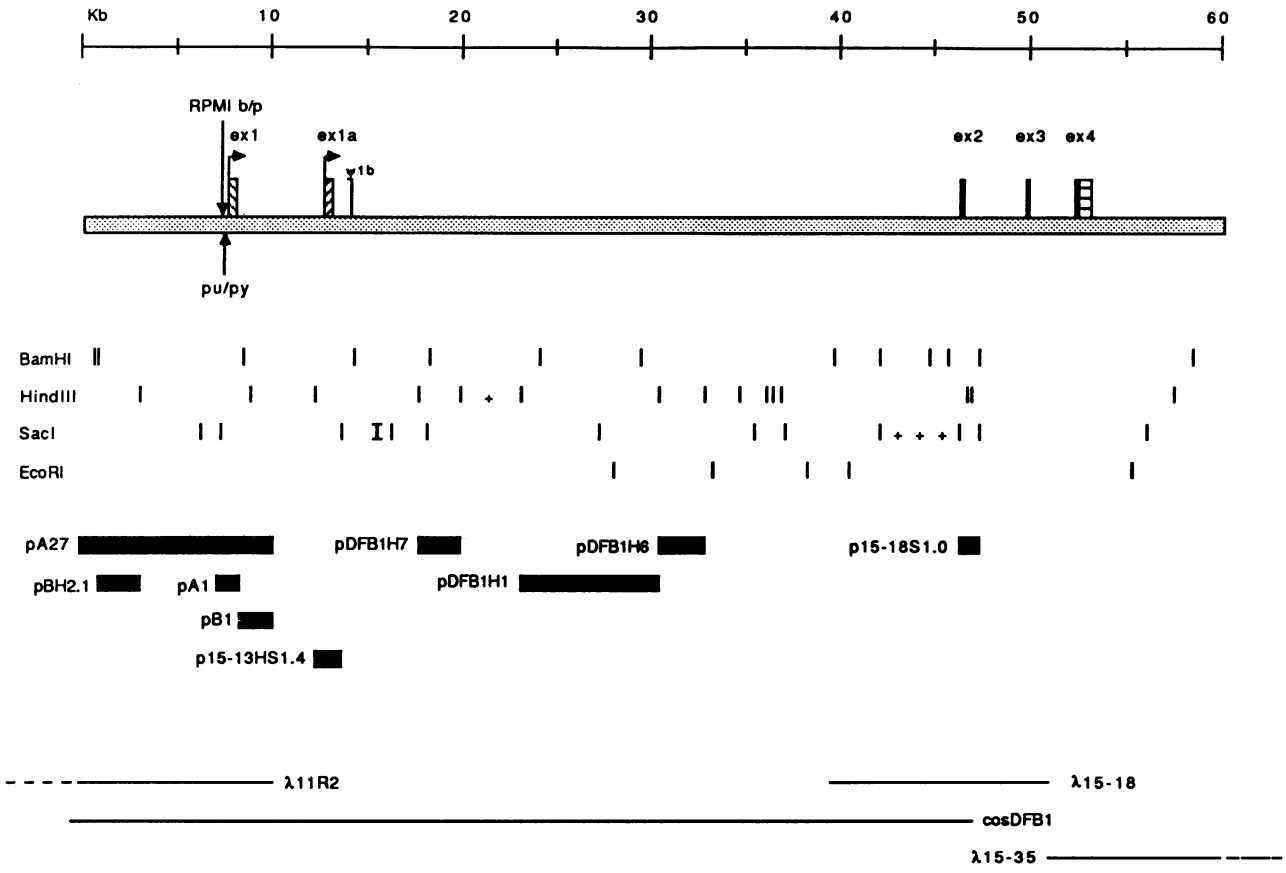

Fig. 5. Detailed organization of the human 11p15 gene. The drawing represents the human 11p15 gene with its exons located (ψ1b is not a real exon but occurs in one aberrantly spliced cDNA, see Figure 4). The position of the translocation breakpoint (b/p) in the T cell tumour line RPMI8402 is indicated by the downward arrow and the location of the tract of alternating purine–pyrimidine residues, previously noted (Boehm *et al.*, 1988a), is shown by the upward arrow. The two transcription promoters, 1 and 1a, are also shown. Beneath the drawing are detailed restriction sites and the location of various non-repetitive probes used to make the map and in other experiments (+ represents additional restriction sites, the precise locations of which could not be mapped). A polymorphic *Sac*I site is indicated by 'I'. A 0.3 kb *Pst*I fragment (clone pDFB1SS4P0.3; not indicated) can be used to assess this polymorphism. Co-dominant inheritance was observed in one family. The frequency is 0.5 for both alleles. A selected number of recombinant clones used to derive the map are shown at the bottom.

intron of ~33 kb separates exon 1a and exon 2; exon 2 and exon 3 are 4 kb apart and exon 3 is 2.3 kb from the last exon 4. Comparison of obtained sequences shows that there are no base mutations in the expressed alleles of RPMI8402 compared with germ line sequences.

Dual promoters control 11p15 gene expression

The possible independent origins of the 1.2 kb mRNA species were examined by Northern hybridization of RPMI8402 or N417 RNA using exon 1, 1a or 2 probes (Figure 7). Probes for exons 1a and 2 detect RNA in both cell lines while the exon 1 probe only detects RNA in N417. The exon 2 probe, corresponding to the major part of the protein coding region, detects the 1.4 kb RNA in RPMI8402 as shown with the cDNA probe which comprised exons 2, 3 and 4 (see Figure 2). In N417 RNA, on the other hand, this probe detects mainly the 1.2 kb RNA. The exon 1 probe also detects the 1.2 kb mRNA in N417 cells but no hybridization was found in RPMI8402 RNA. On the other hand, the exon 1a probe detects RNA in both cell lines, but both display only the 1.4 kb mRNA (the quantity of this RNA is greater in RPMI than in N417 as would be predicted from the ratio of RNAs detected with the exon 2 probe). Therefore, taken together with the cDNA clone sequence organization (Figure 4), the two RNA species must be the result of different promoter activities (designated promoter 1 and 1a); further, while RPMI8402 seems to lack activity from promoter 1, N417 appears to have preferential activity from promoter 1.

These results indicate that the 1.4 kb mRNA derives from a transcriptional promoter upstream of exon 1a which is active in both cell lines and that the 1.2 kb species derives from a promoter upstream of exon 1. This Northern analysis thus confirms the distinct identity and sequence composition of the two mRNA species. In addition functional tests for the activity of the two promoters were carried out. Promoter 1 was assessed in β galactosidase reporter constructs in transgenic mice, principally utilizing a construct carrying an 8 kb *Xho*I–*Nru*I fragment of chromosome 11 germ line sequences (Greenberg *et al.*, 1990). An *in vitro* functional test for the activity of promoter 1a was done by linking a genomic fragment carrying the 5′ end of exon 1a (*Bam*HI–*Bss*HII fragment, see Figures 5 and 10) to a promoterless bacterial chloramphenicol acetyl transferase (CAT) gene. In Figure 8, the CAT activities of cell extracts prepared from RPMI8402 cells after transfection with various plasmids are compared. The promoterless CAT gene gave no enzyme activity after transfection into RPMI cells (pCAT poly(A)4) but the

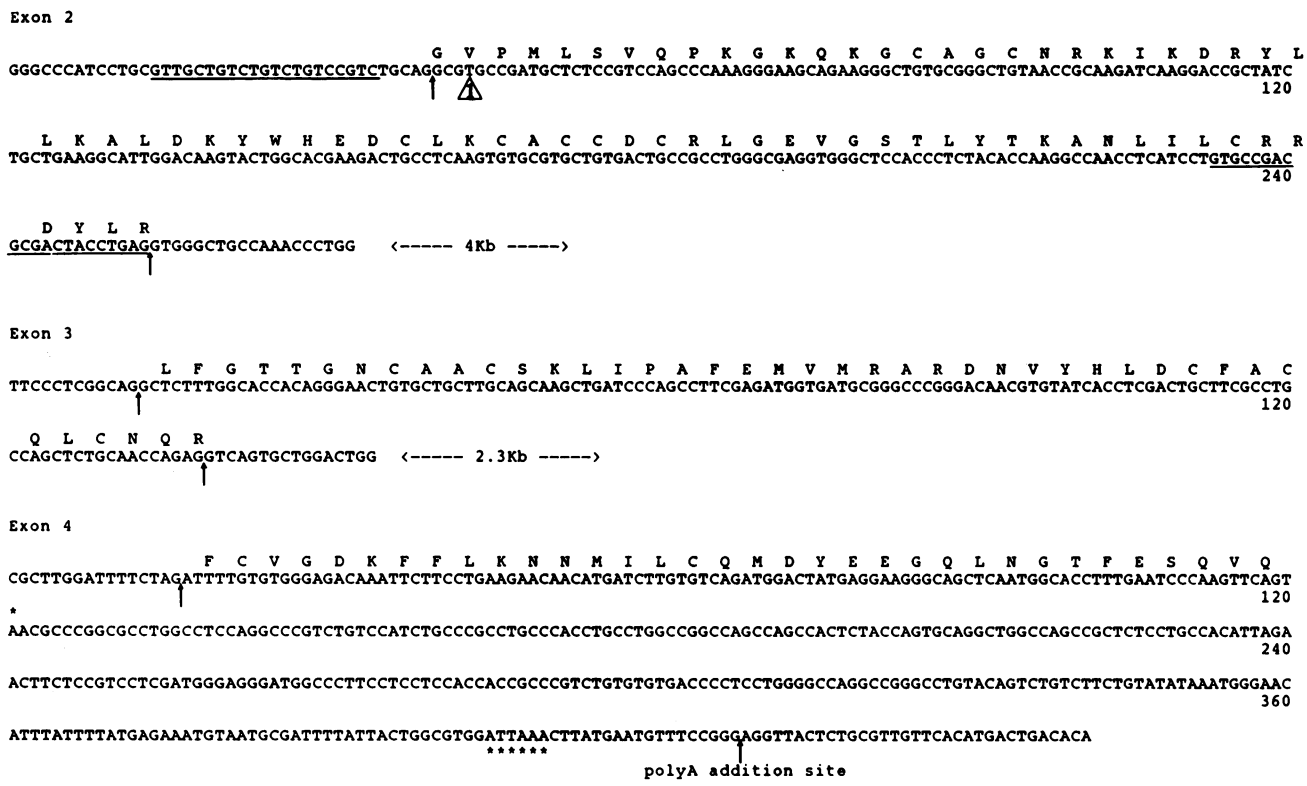

Fig. 6. Nucleotide sequence of the genomic exons for the human 11p15 gene. The genomic sequences are given with a small amount of adjacent intronic sequence to establish the exon boundaries within the cDNA clones. Splice sites are arrowed and the distance between exons is marked. The protein translation of the exons is given in the single letter code and the poly(A) addition signal and the polyadenylation site defined by the cDNAs are shown. The exon 2 PCR probe used in Figure 7 was made using oligonucleotides corresponding to underlined residues in the intron 5′ to exon 2 and the end of exon 2. Note that the genomic sequences of the alternative first exons are given in Figure 9A. (1) denotes the start of the cDNA obtained from the HeLa cDNA library.

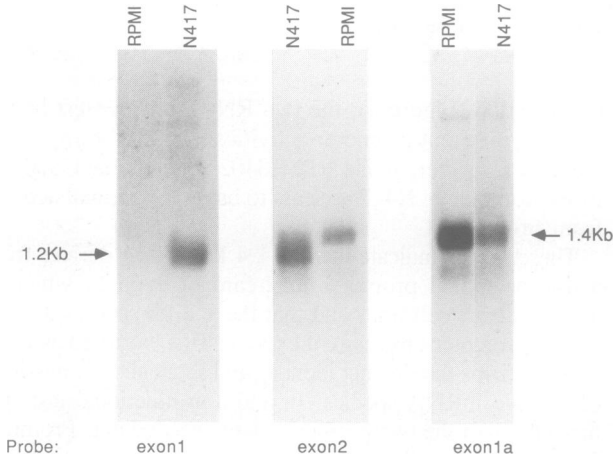

Fig. 7. Northern hybridization to demonstrate the independence of the two 11p15 gene promoters. Poly(A)$^+$ selected RNA from RPMI8402 (1 μg) or N417 (3 μg) were separated on 1.4% agarose after glyoxylation, transferred to nylon filters and hybridized to the indicated probes: the sizes of the two mRNA species were determined in independent experiments. The exon 1 probe corresponds to nt 235–670 of the promoter 1 sequence in Figure 9A, the exon 1a probe corresponds to nt 404–826 of the promoter 1a sequence in Figure 9A, and the exon 2 probe was the PCR product prepared with the primers indicated in Figure 6 (see Materials and methods for details).

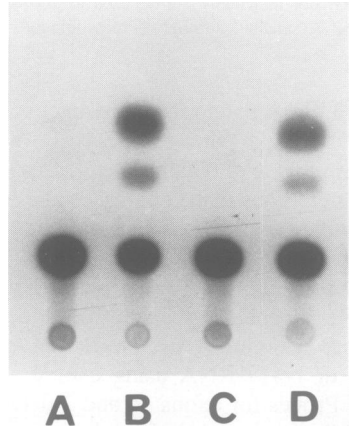

Fig. 8. *In vivo* activity of promoter 1a assayed by production of CAT. The *in vivo* promoter activity of sequences downstream of exon 1 and upstream of exon 1a was determined by linking the 4.7 kb *Bam*HI–*Bss*HII fragment to a promoterless CAT gene and assayed after transfection into RPMI8402 cells. (**A**) No extract (control), (**B**) positive control (pBsptCAT3); (**C**) negative control [pCATpoly(A)4]; (**D**) promoter 1a construct (pCAT1a). For derivation of constructs see Materials and methods.

construct placing the *Bam*HI–*Bss*HII fragment 5′ to the CAT gene (pCAT1a) showed clear evidence of CAT activity (comparable with a control CAT clone). Thus, sequences

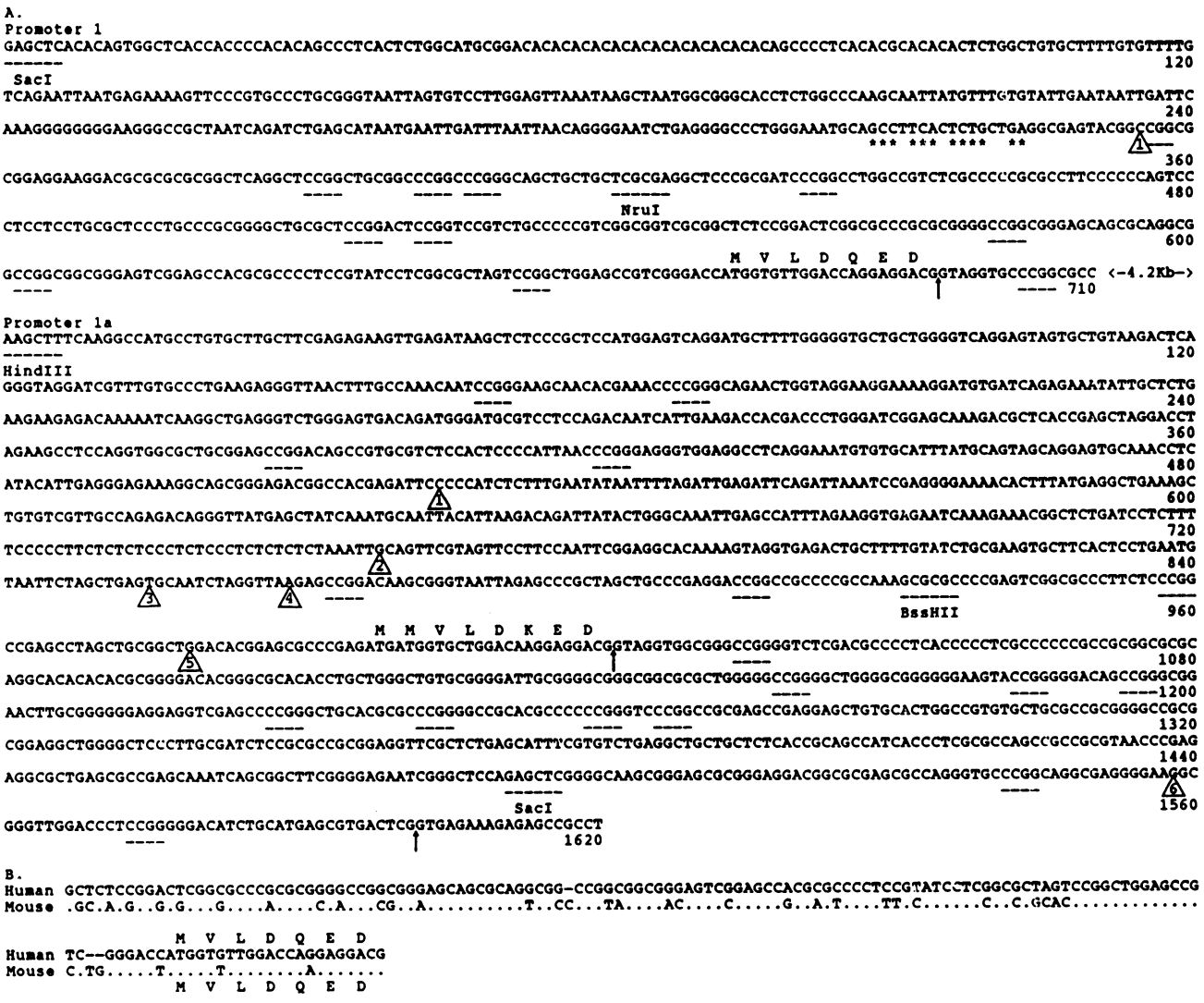

Fig. 9. Genomic sequences of the alternative 11p15 gene promoters. (**A**) The human genomic sequences of promoters 1 and 1a are shown with the alternative exons 1 indicated with their derived translations and donor splice sites. The sequence homologous to the 'initiator' element (Smale and Baltimore, 1989) in promoter 1 is starred. The restriction sites for *Sac*I and *Hin*dIII are indicated to allow alignment to the maps in Figures 5 and 10. *Bss*HII and *Nru*I sites underlined were utilized in the production of the CAT constructs (Figure 8) and β-galatosidase fusion constructs for transgenic mice respectively (Boehm *et al.*, in preparation). *Hpa*II sites (5'-CCGG) are underlined. The start of the cDNAs derived from each promoter region are indicated, as is the start of the read through clone, pCRT18 with the cryptic donor splice site used in this clone marked. Numbers below a nucleotide residue denote the start sites of various cDNAs obtained from the RPMI8402 cDNA library. Promoter 1: 1, start of pCT2; promoter 1a: 1, start of pCRT10.2; 2, start of pCRT3.8; 3, start of pCRT4; 4, start of pCRT11; 5, start of pCRT21; 6, start of pCRT18. (**B**) Comparison of promoter 1 sequences in human and mouse. The sequences obtained in cDNA clone pN2A are compared to the relevant part of the human promoter 1 sequence up to the last residue in human exon 1a. Gaps were introduced to maximize homology. About 74% of nucleotides are identical in the stretch compared.

located downstream of exon 1 and upstream of exon 1a show independent promoter activity, as predicted from the Northern (Figure 7) and cDNA (Figure 4) data. These independent experiments, therefore, confirm that there are two independent promoters in this gene.

Promoters 1 and 1a thus are separate transcriptional entities which have different activities in different cell types. The derived nucleotide sequences of the two promoter regions (Figure 9) are quite different from each other (no useful alignment of the sequences could be obtained) and neither appears to contain recognizable control elements, except promoter 1 which shows homology to the recently described 'initiator' sequence which is involved in the transcription of the terminal transferase and other genes (Smale and Baltimore, 1989). No such motif occurs at promoter 1a. These results suggest that the two promoters do not share transcriptional control mechanisms. This conclusion is reinforced by the conservation of promoter 1 sequences between human and mouse (Figure 9B).

Protein initiation codons in the alternative first exons

The promoter activities described above and sequence of the various cDNA clones (Figure 1) indicate that the N-terminus of the 11p15 protein is encoded alternatively by either exon 1 or by exon 1a. This conclusion was confirmed by obtaining the human genomic sequence of the promoter/exon 1 and of promoter/exon 1a (Figure 9A). The boundaries of the postulated exon 1 and exon 1a are indeed marked by donor

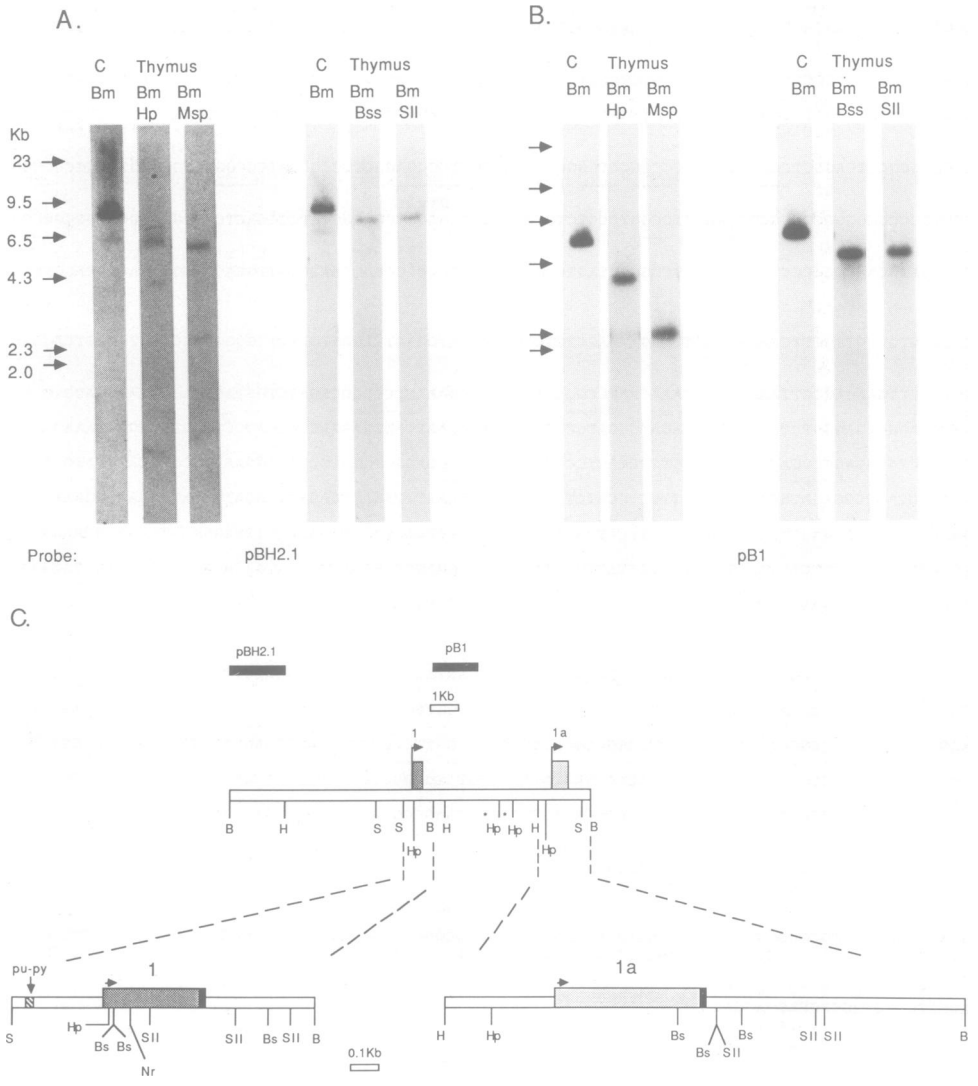

Fig. 10. Hybridization analysis of the methylation status of the two 11p15 promoter regions. (**A**) and (**B**). Genomic filter hybridizations of human DNA (10 μg/lane) with the indicated probes (the locations of the two probes are shown in C). C is human dermal fibroblast DNA and the thymus DNA comes from a 19 week foetus. Bm = *Bam*HI; Hp = *Hpa*II; Msp = *Msp*I; Bs = *Bss*HII; SII = *Sac*II. Sizes were estimated by co-electrophoresis of λDNA cut with *Hin*dIII. (**C**) Representation of the promoter region from the 11p15 gene. The upper diagram shows the relationship between the two promoters and the two probes used in (A) and (B). B = *Bam*HI; H = *Hin*dIII; S = *Sac*I; Hp = *Hpa*II (the two *Hpa*II sites indicated with stars are discussed in the text). The lower diagrams are enlarged versions of the two promoters. The shaded areas correspond to the regions found in the longest cDNA clone derived from each promoter region. The purine–pyrimidine tract previously described is shown in promoter 1 (pu-py). Only the first *Hpa*II sites are indicated (cf. to Figure 9A for other sites). No *Sac*II or *Bss*HII sites other than those indicated occur in the *Sac*II–*Bam*HI fragment containing promoter 1 and the *Bam*HI fragment containing promoter 1a. The presence of additional *Hpa*II sites between the two *Hpa*II sites marked with * has not been investigated. S = *Sac*I; Hp = *Hpa*II; Bs = *Bss*HII; Nr = *Nru*I; SII = *Sac*II; B = *Bam*HI.

splice sites and the coding part of the exon is 7 or 8 codons for exons 1 and 1a respectively (Figure 9); each exon contains its own ATG initiation codon. The derived sequence of these two short exons in man differ only in one internal residue (lysine–glutamine interchange) and exon 1a has two consecutive ATG codons. Interestingly, the mouse cDNA clone derived from the neuroblastoma Neuro2A possesses the exon 1 sequence which is identical to its human counterpart and which is consistent with transcription from the equivalent promoter 1 in this mouse line (Figures 1 and 9B). We have used a PCR based cDNA cloning procedure with single sided specificity (Belyavsky *et al.*, 1989) to obtain the sequence of the mouse exon 1a. The result of this experiment shows that the protein coding capacity is identical to the human exon 1a sequence (Figure 11C).

Thus the N-terminal seven or eight residues of the 11p15 protein are encoded separately by two distinct exons, which are controlled by two transcriptional promoters. Even so, the final protein made from either transcription start will be virtually the same in man and in mouse (i.e. except for the single amino acid). Therefore, the dual promoter/coding exon organization described for the 11p15 gene is thus far unique as a mechanism for protein production.

The dual promoters exist in distinct methylation free islands

The discrepancy in amounts of the two mRNA species in the cell lines tested indicated that there is cell specific activity

of the promoters (Figures 3 and 7) and, further, the sequence divergence of the two promoter regions suggests that they are separately controlled. The nucleotide sequences of the two regions include several rare cutting restriction enzymes sites which are the hallmark of methylation free islands which occur in the promoter regions of many genes (Bird, 1987). The methylation status of the dual promoters was assessed, therefore, in genomic DNA made from human thymus. Figure 10C shows a map of the promoter region with an indication of the positions of *Sac*II and *Bss*HII (other rare cutters, e.g. *Eag*I, are also present but are not indicated). Two probes were used to separately assess the methylation of the *Sac*II, *Bss*HII and *Hpa*II sites of both promoters (Figure 10C).

The probe pBH2.1 detects a 7.5 kb *Bam*HI genomic fragment. This fragment is reduced to ~7 kb by digestion with either *Bss*HII or with *Sac*II (Figure 10A), showing the methylation free status of the sites illustrated in Figure 10C. Similarly, pB1 detects a 5.8 kb *Bam*HI fragment (Figure 10B) which is reduced to ~4.8 kb with either *Bss*HII or *Sac*II. The two promoters are thus methylation free as judged by these two methylation dependent enzymes. Methylation also affects *Hpa*II digestion of genomic DNA, but *Msp*I, which has the same recognition sites as *Hpa*II, is unaffected by methylation. Since the recognition site (CCGG) occurs more frequently than those of *Bss*HII or *Sca*II, it was possible to test whether the whole region is a methylation free island or whether the two promoters are independent islands. Digestion of thymus DNA with mixtures of *Bam*HI and *Hpa*II or *Msp*I reduces the size of the *Bam*HI fragment to about the same size (although the *Hpa*II digest is somewhat more heterogeneous). This corresponds to digestion, within the *Bam*HI fragment, of the first *Hpa*II site within exon 1 (Figure 10C). Clearly, it is not possible to comment on the other *Hpa*II sites in the region but, together with the *Bss*HII and *Sac*II digestion, it is probable that this is a methylation free region.

The probe pB1, on the other hand, detects a 4.0 kb *Hpa*II + *Bam*HI fragment which corresponds to digestion at the *Hpa*II site just upstream of the exon 1a (see Figure 10C). Thus the promoter 1a is methylation free. In addition, digestion with *Msp*I and *Bam*HI shows that the methylation free regions of the dual promoters are in fact separate because this digestion allows the detection of a 2.5 kb fragment compared with the 4.0 kb *Hpa*II + *Bam*HI fragment. Thus the first CCGG in the *Bam*HI fragment detected by pB1 is between the two promoter regions (indicated in Figure 10C) and this site is methylated (because it is not cut by *Hpa*II in thymus DNA). Furthermore the same is true of the adjacent sites because *Hpa*II does not recognize any of these until the promoter proximal one. The region at least stretching from the indicated *Hpa*II sites (starred in Figure 10C) to the exon 1a site is, therefore, methylated. (Note: other *Hpa*II sites in this stretch can be assessed from the promoter sequence in Figure 9A.) The dual promoters of the 11p15 gene are, therefore, separate but adjacent methylation free islands which is, in itself a novel situation and presumably relates to the apparent differences in promoter activity.

Discussion

A novel gene structure designed to yield two similar forms of a protein

Previously a number of situations have been described in which alternative splicing within a gene gives rise to one protein (for review, see Andreadis *et al.*, 1987). The two relevant examples are the α-amylase and myosin light chain genes, as these also have alternative promoters. These are diagrammatically compared with the 11p15 gene in Figure 11A. The α-amylase gene (top of Figure 11A) is transcribed from two tissue specific promoters which give rise to alternative first exons (Young *et al.*, 1981). However, the protein coding capacity of this gene starts in exon 2 so that the coding part of the alternative mRNAs is the same, irrespective of the first exon attached. A different situation is seen in the example of the myosin light chain gene (Nabeshima *et al.*, 1984) where a single gene has two promoters (centre of Figure 11B) which gives rise to two major forms of mRNA; these mRNA molecules have different exon organization and content, from which the two differing forms of protein can result. In this case, the alternative methionine codons are within the alternative first exons, the first being virtually in the middle of exon 1 and the second being exactly at the 3′ end of exon 2. Together with programmed alternative splicing, this results in two quite different proteins.

The 11p15 gene and its protein products represent a third, distinct transcription−translation organization in which two similar proteins are made from a choice of two mRNAs transcribed from distinct promoters (Figure 11A, bottom). The alternative protein coding starts at identical positions within either exon 1 or 1a in contrast to the situations of amylase or myosin (Figure 11A). The 11p15 protein can exist in both humans and mouse, in two virtually identical forms since the single amino acid change is between two hydrophilic residues (glutamine or lysine) near the N-terminus of the protein. At present it is unknown, whether this fairly conservative substitution alters the function of the protein. Unlike the other dual promoter genes thus far described, the 11p15 gene is a single gene which gives rise to two mRNAs with different first exons from which essentially identical proteins can be synthesized. Therefore, it is possible that the protein inside the cell will be functionally identical irrespective of the promoter used for its mRNA synthesis. However, the fact that the glutamine/lysine substitution is conserved between man and mouse might indicate a functionally important change.

Why are two similar forms of protein made from dual promoters?

The synthesis of two mRNA species from the 11p15 gene and the ability to produce two similar forms of protein superficially resembles that found, for instance with the amylase gene (Young *et al.*, 1981). However, in the present case there is a very different organization leading to this situation (Figure 11A). Nonetheless, it is conceivable that the reason for the presence of alternative promoters is similar, viz. the need for independent control of the production of the same protein in different cells. Our data show that the two promoters responsible for the two mRNAs are differentially expressed, at least as judged by mRNA levels (although we cannot exclude post-transcriptional mechanisms), in cells of quite different embyrological origin (i.e. T cells and neural crest derivatives) and recent experiments using a human promoter 1 fragment in transgenic mice have indicated that the central nervous system is a major site of transcriptional activity of this gene (Greenberg *et al.*, 1990). The nucleotide sequence of the

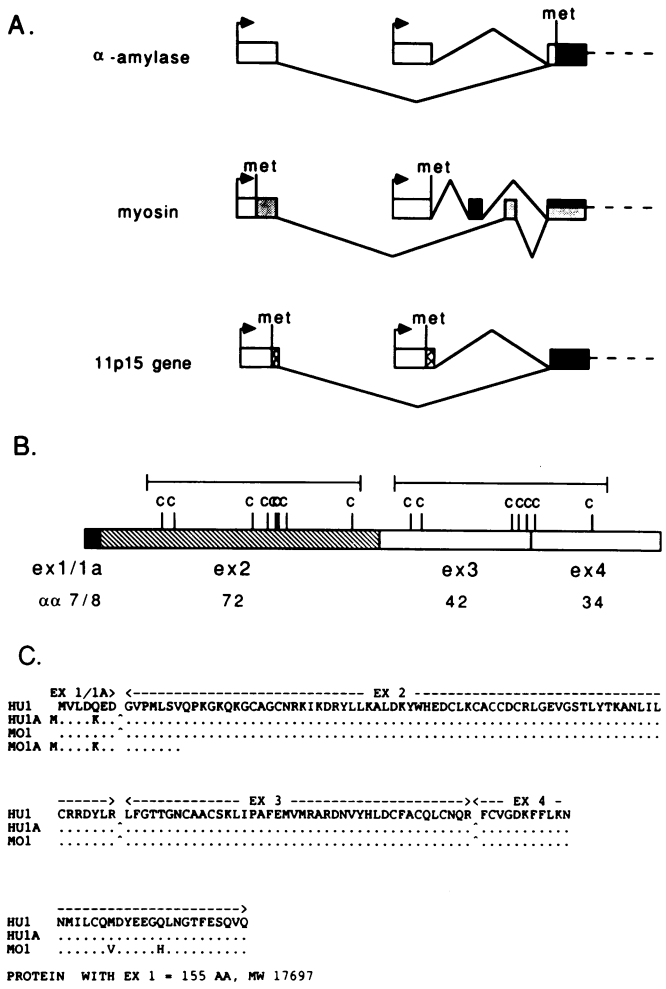

Fig. 11. Structural comparison of the 11p15 gene and its protein products. (**A**) A diagrammatic comparison of the exon organization of the 11p15, α-amylase and myosin genes. The various, relevant exons are indicated and the locations of the dual promoters in each gene are arrowed. The position within the gene of the methionine initiation codons are given (met) and the non-coding regions are as open boxes. RNA splicing alternatives are indicated. (**B**) Diagrammatic representation of the 11p15 derived protein. The length of each exon in codons is shown with its delineation. The positions of cysteine residues within the putative protein are indicated and the apparently duplicated regions are overlined. (**C**) Comparison of putative protein sequences derived from the two forms of human and mouse cDNA clones. The full derived sequence, in the single letter code, from the clone pCRT21 is shown and the exon boundaries are indicated by arrow heads. Identity between residues in this sequence and the second human sequence or the mouse sequences are given as dots. (The exon boundaries of the mouse are not known because only cDNA sequence was obtained.) The nucleotide sequence of mouse exon 1a as determined by PCR-based cDNA synthesis (see Materials and methods) is as follows: 5′-CTGGACAC-GGAGCGCCCGAGATGATGGTGCCTGGACAAGGAGGACG-3′. The two tandem ATG initiation codons are underlined. Only the deduced protein sequence is given in the figure. The obtained 5′ untranslated sequence is identical to the human sequence, very similar to the situation with exon 1 (see Figure 9B).

dual promoters of the 11p15 gene are rather different from each other, e.g. the 'initiator' sequence is found in promoter 1 but not in promoter 1a. However, the sequences for promoter 1 in human and mouse are very similar (Figure 9B); a comparison between promoter 1a sequences from human and mouse could not be made because of the short region of 5′ untranslated sequence in the mouse cDNAs (legend to Figure 11). In the human DNA, the two promoters lie within separate methylation free islands which might provide one clue to the differential usage of the dual promoters and certainly this is an interesting area for study of the control of transcription of this gene.

The 11p15 gene encodes a highly conserved protein in man and mouse

The deduced protein encoded by the 11p15 gene indicates a 155 amino acid polypeptide. Figure 11B shows a comparison of the two forms of human and mouse protein. Based on the present data, it seems certain that the mouse genome also has dual exons 1 for this gene, because the single amino acid difference between exons 1 and 1a is conserved across species. Hydropathy plots of the protein make it seem unlikely to be a membrane protein and it is presumably therefore a soluble, intra-cellular protein. There are no obvious sequence similarities to known proteins, except some homology with the EF-hand Ca^{2+} binding family of proteins (Kretsinger, 1980). However, this is unlikely to be of functional significance because of the divergence of crucial residues.

The presence of 15 cysteine residues and their location within the protein indicates a metal binding protein. This potential metal binding feature is in common with a number of other proteins, e.g. metallothionein (Durnam, 1980) and the c-*myc* protein (van Straaten and Rabbitts, 1987). In both

species examined there are 15 cysteine residues. As illustrated in Figure 11B, these appear to have resulted from an evolutionary duplication giving rise to two tandem sets (overlined in Figure 11B). Interestingly the second, but not the first, group of cysteine residues is split by an RNA splice site. The extremely conserved nature of the deduced protein sequence in man and mouse is intriguing and certainly indicates a protein with some important basic function.

Materials and methods

Nucleic acid biochemistry

Procedures for DNA and RNA extraction, poly(A)$^+$ selection of RNA, Southern and Northern blotting, hybridization experiments and DNA sequencing procedures have all been described previously (Boehm *et al.*, 1988a,b, 1989; Buluwela *et al.*, 1989).

Genomic libraries and probes

The genomic structure of the 11p15 locus was determined initially using cDNAs as probes and in subsequent rounds, selected unique genomic fragments for chromosomal walking experiments to screen human genomic libraries made in λ2001 (Lefranc *et al.*, 1986) and a *Hin*dIII partial digest library in the Lorist 6 vector made from primary human dermal fibroblast DNA as described (Boehm *et al.*, 1988b). Overlapping λ and cosmid clones were isolated and used to derive the map shown in Figure 5. Probes used for hybridization experiments are indicated on the maps given with the appropriate figures. Exon specific probes were derived as follows: an *Eco*RI/*Nco*I fragment (corresponding to nt 235–670 of promoter 1 sequence in Figure 9A) isolated from pCT2 was used as an exon 1 specific probe; an *Eco*RI/*Nar*I fragment (nt 404–826 of promoter 1a sequence in Figure 9A isolated from pCRT10.2) was prepared as an exon 1a-specific probe. An exon 2 specific probe was made by using subclone p15-18S1.0 (Figure 5) as template and the two exon 2 primers described below in a polymerase chain reaction.

Determination of exon 2 sequence in RPMI 8402

As described in the legend to Figure 1, the cDNA clone pCT2 showed an extra A residue after position 31 (see Figure 1). Since that sequence change was not observed in any other cDNA, we determined the sequence of exon 2 by polymerase chain reaction using RPMI 8402 DNA as template. The following primers were used (linkers in the lower case letters): forward 5′-tttgaattcGTTGCTGTCTGTCTGTCCGTC-3′ derived from the 3′ end of intron 2 (see Figure 6); reverse 5′-tttaagcttCTCAGGTAGTCGCGTC-GGCAC-3′ derived from the 3′ end of exon 2 (see Figure 6). Pcr conditions were as follows: denaturation, 2 min at 94°C; annealing, 2 min at 55°C; extension, 2 min at 70°C. The PCR product was purified by agarose gel electrophoresis, cut with *Hin*dIII and *Eco*RI and cloned into mp18/19 vectors and sequenced. Forty-seven individual clones were sequenced and all found to contain wild-type sequences (i.e. no A at the position of the pCT2 cDNA sequence). We thus conclude that the extra residue in pCT2 represents a reverse transcription artefact.

cDNA libraries

The RPMI 8402 cDNA library has been described previously (Baer *et al.*, 1988). A cDNA library from Neuro 2A (CCL 131) was made in λgt10 using poly(A)$^+$ selected cytoplasmic RNA and a cDNA synthesis kit from Amersham. The HeLa cDNA library in λgt10 was a gift from D.R.Gewert and I.Kerr, the Kelly cDNA library in λgt10 was kindly provided by J.M.Bishop and J.Ibson.

Initial screening of the RPMI 8042 library was carried out with pA27 (Figure 5) to obtain clone pCT2 (Figure 4). This cDNA was used to rescreen the RPMI 8402 library and as a probe to screen the Neuro2A, HeLa and Kelly cDNA libraries.

A PCR based cDNA cloning procedure with single sided specificity was employed to determine the sequence of mouse exon 1a. To this end, the first strand synthesized using oligo(dT) as primer was G-tailed with terminal transferase as described (Belyavsky *et al.*, 1989). This tailed first strand cDNA was amplified using a universal C primer: 5′-AAGAATTCC-CCCCCCCCCCC-3′ (Belyavsky *et al.*, 1989) and two back primers specific for mouse exon 2 (primer 1, 5′-TAGAAGCTTACAGCCCTTCTGTT-TCCCCTTAGGT-3′ and primer 2, 5′-TAGAAGCTTGTCCAGTGCC-TTCAGCAGGTAC-3′). The *Hin*dIII and *Eco*RI cloning sites at the 5′ ends of amplification primers are underlined. Amplification conditions were as follows: 94° for 2 min, 45° for 2 min and 65° for 2 min for 30 cycles. Individual amplification products were cloned into M13 vectors and their consensus sequences determined.

Transfection and CAT assay

The 1.6 kb *Bam*HI–*Hin*dIII fragment of pSV2CAT (Gorman *et al.*, 1982) was blunt ended with Klenow fragment of DNA polymerase I and cloned into the *Eco*RV site of Bluescript (Stratagene, San Diego, USA) to give pBsptCAT3. This vector gives high CAT activity in all cell lines tested, presumably via readthrough from plasmid promoters; hence it was used as a positive control in the present experiments. To eliminate CAT activity from the vector, a 1.6 kb *Sac*I fragment from pZA (Hall *et al.*, 1983) containing the 3′ end of lacZ and the SV40 polyadenylation site was cloned into the *Sac*I site of pBSptCAT3 to give pCATpoly(A)4. This plasmid was used as a starting material for subcloning various fragments from the 11p15 locus. The 4.7 kb *Bam*HI–*Bss*HII fragment containing exon 1a lacking the donor splice site (Figure 5) was blunt ended as above and cloned into the *Sma*I site of pCATpoly(A)4 to give pCAT1a.

Transfection by electroporation (Neumann *et al.*, 1982) of plasmid constructs was carried out using the Genepulser apparatus (Bio-Rad) using 1×10^7 cells in 200 μl phosphate buffered saline containing 20 μg of super-coiled plasmid DNA. For RPMI 8402, the transfection conditions were 0.2 KV and 960 μF at an inter-electrode distance of 0.4 cm, resulting in time constants of ~50 μs (Döffinger *et al.*, 1988). CAT assays were carried out after 48 h as described (Gorman *et al.*, 1982).

Acknowledgements

We wish to thank Drs D.R.Gewert and I.Kerr for the HeLa cDNA library and Drs J.M.Bishop and J.Ibson for the provision of the Kelly library. Also we would like to thank Sven Petterson for invaluable help with the isolation of pre-natal mouse thymus and Aaron Klug for discussions on the putative protein structure. Cell lines were gratefully received from Drs J.Minowada, P.H.Rabbitts, U.Kees, B.O.Boehm, and the ATCC. J.M.G. is the recipient of a Special Fellowship from the Leukaemia Society of America.

References

Andreadis,A., Gallego,M.E. and Nadal-Ginard,B. (1987) *Annu. Rev. Cell Biol.*, **3**, 207–242.

Baer,R., Forster,A., Lavenir,I. and Rabbitts,T.H. (1988) *J. Exp. Med.*, **167**, 2011–2016.

Belyavsky,A., Vinogradova,T. and Rajewsky,K. (1989) *Nucleic Acids Res.*, **17**, 2919–2932.

Bernard,O., Larsen,C.-J., Hampe,A., Mauchauffe,M., Berger,R. and Mathieu-Mahul,D. (1988) *Oncogene*, **2**, 195–200.

Bird,A.P. (1987) *Trends Genet.*, **3**, 342–347.

Boehm,T., Baer,R., Lavenir,I., Forster,A., Waters,J.J., Nacheva,E. and Rabbitts,T.H. (1988a) *EMBO J.*, **7**, 385–394.

Boehm,T., Buluwela,L., Williams,D., White,L. and Rabbitts,T.H. (1988b) *EMBO J.*, **7**, 2011–2017.

Boehm,T., Mengle-Gaw,L., Kees,U.R., Spurr,N., Lavenir,I., Forster,A. and Rabbitts,T.H. (1989) *EMBO J.*, **8**, 2621–2631.

Buluwela,L., Forster,A., Boehm,T. and Rabbitts,T.H. (1989) *Nucleic Acids Res.*, **17**, 425.

Croce,C.M. (1987) *Cell*, **49**, 155–156.

Döffinger,R., Pawlita,M. and Sczakiel,G. (1988) *Nucleic Acids Res.*, **16**, 11840.

Durnam,D.M., Perrine,F., Gannon,F. and Palmiter,R.D. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 6511–6515.

Finger,L.R., Huebner,K., Cannizzaro,L.A., McLeod,K., Nowell,P.C. and Croce,C.M. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 9158–9162.

Gorman,C.M., Moffat,L.F. and Howard,B.H. (1982) *Mol. Cell. Biol.*, **2**, 1044–1051.

Greenberg,J.M., Boehm,T., Sofroniew,M.V., Keynes,R.J., Barton,S.C., Norris,M.L., Surani,M.A., Spillantini,M.-G. and Rabbitts,T.H. (1990) *Nature*, in press.

Hall,C.V., Jacob,P.E., Ringold,G.M. and Lee,F. (1983) *J. Mol. Appl. Genet.*, **2**, 101–109.

Hamlyn,P.H. and Rabbitts,T.H. (1983) *Nature*, **304**, 135–139.

Knudson,A.G. (1985) *Cancer Res.*, **45**, 1437–1443.

Kretsinger,R.H. (1980) *CRC Crit. Rev. Biochem.*, **8**, 119–174.

Lefranc,M.-P., Forster,A., Baer,R., Stinson,M.A. and Rabbitts,T.H. (1986) *Cell*, **45**, 237–246.

McGuire,E.A., Hockett,R.D., Pollock,K.M., Bartholdi,M.F., O'Brien,S.O. and Korsmeyer,S.J. (1989) *Mol. Cell. Biol.*, **9**, 2124–2132.

Nabeshima,Y., Fuji-Kuriyama,Y., Muramatsu,M. and Ogata,K. (1984) *Nature*, **308**, 333–338.

Neumann,E., Schaefer-Ridder,M., Wang,Y. and Hofschneider,P.H. (1982) *EMBO J.*, **1**, 841–845.
Rabbitts,T.H., Boehm,T. and Mengle-Gaw,L. (1988) *Trends Genet.*, **4**, 300–304.
Shima,E.A., LeBeau,M.M., McKeithan,T.W., Minowada,J., Showe,L.C., Mak,T.W., Minden,M.D., Rowley,J.D. and Diaz,M.O. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 3439–3443.
Smale,S.T. and Baltimore,D. (1989) *Cell*, **57**, 103–113.
van Straaten,J.P. and Rabbitts,T.H. (1987) *Oncogene Res.*, **1**, 221–228.
Young,R., Hagenbüchle,O. and Schibler,U. (1981) *Cell*, **23**, 451–458.

Received on October 30, 1989; revised on December 11, 1989