Nucleic Acids Research, 1997, Vol. 25, No. 22 **4513–4522**

Characterization of multiple alternative RNAs resulting from antisense transcription of the PR264/SC35 splicing factor gene

Alain Sureau[+], Johann Soret[+], Cécile Guyon, Christophe Gaillard, Stéphanie Dumon, Monika Keller, Patricia Crisanti[1] and Bernard Perbal*

Laboratoire d'Oncologie Virale et Moléculaire, Unité de Formation et de Recherche de Biochimie, Université Paris 7 (D. Diderot), 2 Place Jussieu, 75005 Paris, France and [1]Faculté de Médecine Broussais-Hôtel Dieu, 75270 Paris, France

Received August 8, 1997; Revised and Accepted October 8, 1997

ABSTRACT

The PR264/SC35 splicing factor belongs to the family of SR proteins which function as essential and alternative splicing factors. Here, we report that the human PR264/SC35 locus is bidirectionally transcribed. Double *in situ* hybridization experiments have allowed simultaneous detection of sense and antisense RNA in human CCRF-CEM cells, suggesting that expression of the corresponding genes is not mutually exclusive. We have characterized three main classes of ET RNAs encoded by the opposite strand of the PR264/SC35 gene and containing PR264/SC35-overlapping sequences, PR264/SC35-non overlapping sequences or a combination of both. We show that their expression results from the use of alternative promoters, exons and polyadenylation signals. PR264/SC35-non overlapping ET mRNA species potentially encode two protein isoforms (449 and 397 amino acids) and are expressed from the PR264/SC35 promoting region. Northern blots and RNase protection analyses indicate that ET polyadenylated RNAs are differentially expressed in several human cell lines. Similar studies performed in the mouse have revealed that the bidirectional transcription of the PR264/SC35 locus is a conserved mechanism and that the open reading frame identified in a subset of human ET mRNAs is highly conserved (93% homology). Northern blot analyses performed with several murine tissues confirmed the differential expression of the ET gene and revealed that it is predominantly expressed in the testis.

INTRODUCTION

The PR264/SC35 splicing factor belongs to the evolutionarily well-conserved family of SR proteins (for reviews, see refs 1,2) which comprises at least nine members (3 and refs therein), designated SRp75, SRp55 (B52), p54, SRp40 (HRS), SRp30a (ASF/SF2), SRp30b (PR264/SC35), SRP30c, 9G8 and SRp20 (X16/RBP1). These RNA binding proteins have one or two copies of the RNA recognition motif (RRM) and contain regions rich in arginine and serine residues (RS domains). Like most of the SR proteins, PR264/SC35 is involved in constitutive splicing processes (4,5) and influences the selection of alternative splice sites in a concentration-dependent manner (5,6). Evidence that SR proteins such as ASF/SF2 and PR264/SC35 exhibit different substrate specificities suggests an essential and non-redundant *in vivo* function (7 and refs therein). This is supported by several studies showing that SRp55/B52 (8,9) and ASF/SF2 (10) splicing factors are essential for *Drosophila* development and required for cell viability, respectively.

We previously observed that short sequences (designated as ET sequences) complementary to PR264/SC35 transcripts are present at the 5′ end of a c-myb cDNA species isolated from both avian and human cells (11,12). Unlike numerous bidirectionally transcribed prokaryotic genes (13), little is known about the biological significance of antisense RNAs identified in eukaryotes. However, several functions have been proposed that stress the biological importance of natural antisense RNAs (14). Among the different examples of antisense RNA, some, like EB4 (15) and bFGF (16) antisense RNA, control the post-transcriptional level of the complementary RNA by regulating its stability. A similar role has been proposed for WT1 antisense transcripts (17). RNA–RNA duplexes would be either digested by double strand-specific RNAse (15) or modified by unwindase-deaminase and the resulting A→I conversion would trigger faster degradation of the reacting RNAs (16,18,19). A second class of antisense RNA has been shown to interfere with the processing of the sense RNA. Indeed, the splicing of c-erbA (20), N-myc (21) and probably p53 (22) transcripts is altered by the corresponding antisense RNAs. The third class consists of antisense RNAs that encode polypeptides. Among them, the c-erbAα antisense RNA species encodes a polypeptide belonging to the c-erbA family (23,24) and bFGF complementary transcripts encode a MutT-related protein (25,26). Such examples indicate that antisense RNA can exhibit both regulatory and coding capacities. Strikingly, complementary transcripts expressed from the c-erbAα gene as well as mRNA species transcribed from the α-1(IV)/α-2(IV) collagen (27), histidyl tRNA synthetase (28) and Surf1/Surf2 (29) bidirectional

*To whom correspondence should be addressed. Tel: +33 1 44 27 36 14; Fax: +33 1 30 64 18 65; Email: bernard.perbal@wanadoo.fr

[+]Present address: Laboratoires R. Latarjet, Centre Universitaire, 91405 Orsay Cedex, France

promoters code for structurally and/or functionally related products. This suggests that bidirectional transcription represents useful means to regulate and coordinate expression of related genes. For the other bidirectionally transcribed genes including c-myc (30), BCMA (31) and thymidylate synthase (32), the function of the antisense RNA is not yet understood.

As a first step to understanding the role of bidirectional transcription of the PR264/SC35 gene, we have characterized several classes of ET RNAs expressed in the PR264/SC35 opposite orientation and shown that expression of both sense and antisense complementary transcripts is not mutually exclusive. Our results also indicate that PR264/SC35-non overlapping sequences present in a subset of ET mRNA species contain a 1347 nucleotide (nt) open reading frame (ORF) which is highly conserved in the mouse. Finally, we have observed that ET expression is regulated in a tissue-specific way in both human and murine species.

MATERIALS AND METHODS

DDBJ/EMBL/GenBank accession numbers for human ET cDNAs are AF015183 (ET12), AF015184 (ET16), AF015185 (ET10), AF015186 (ET22), AF015187 (ET32), AF015188 (ET42), AF015189 (ET51) and AF015190 (ET61). Accession number for the murine ET8 cDNA is AF015191.

Cell cultures and tissues

HeLa and 293 (ATCC CRL1573) cells were grown in Dulbecco's modified Eagle's medium and in Eagle's minimum essential medium (Gibco BRL), respectively, supplemented with 10% fetal bovine serum. HL60 (ATCC CCL240) and CCRF-CEM (ATCC CCL119) cells were grown in RPMI 1640 medium (GIBCO/BRL) supplemented as recommended by the ATCC. The human thymus used in these studies was a surgical sample from a 1 month old girl. Murine tissues except thymus were from 11–14 week old adult mice. Thymus was from 5 day old mice.

Probes, RNA purification and analysis

The human double-stranded A, F and H probes were obtained by PCR amplification performed with the E830 genomic DNA subclone (25) and the H19 (5′-GCAGTGAGTCAGGCTGGG-AG-3′) and H62 (5′-TGAGGAAGTCTGGGTGGGAA-3′), H99 (5′-GTGAGTTGGAATCTGTCCTCC-3′) and H62 (5′-TGAGG-AAGTCTGGGTGGGAA-3′), H78 (5′-CGGGACGAGCAAG-CACAGC-3′) and H77 (5′-GTAAACGGGGCTGAGGGGAC-3′) oligonucleotide pairs, respectively. The B, E and G probes resulted from PCR amplification of ET12 sequences with H67 (5′-GATTGTCAGTGGCTTCGCCC-3′) and H60 (5′-CCACAA-GTTTGAAAGG CAGTG-3′), H73 (5′-GAACTTCGCCCTAA-ACCTGG-3′) and H64 (5′-AGATGGTGAGTGGAGAAGGC-3′), H21 (5′-CGAAGCGCCTGCGCACAGC-3′) and H22 (5′-GCGA-AGAGCAGTCAGGAACG-3′) primers, respectively. C and D probes were generated by PCR amplification of ET16 sequences with H79 (5′-GTCATCAGGAGCTTAAATAGG-3′) and H80 (5′-CTACAATGGCAACCACTGACG-3′), H100 (5′-GGACA-GCACAAGGAAACTGC-3′) and H101 (5′-CTGATATCTGA-GTTTTCCCTTGC-3′) oligonucleotide pairs, respectively. The 0.3 kilobase (kb) MuBH probe resulted from *Bam*HI–*Hin*dIII digestion of a 2.2 kb *Hin*dIII fragment of murine genomic DNA containing PR264/SC35 coding exons, first intron and promoter sequences (Ch.Gaillard and B.Perbal, unpublished data). RNA purification, selection of polyadenylated species and northern blot analyses were performed as previously described (33). The human and murine glyceraldehyde-3-phosphate dehydrogenase (GAPDH)-specific probes were used as an internal control to normalize for RNA amounts.

RNase protection analyses

The 1.34 kb *Hin*dIII–*Nhe*I fragment containing genomic sequences corresponding to coding exon 1 and 2 of the human PR264/SC35 gene was inserted into the pB KS(+) vector (Stratagene) and linearized at the *Nco*I site to generate the 0.38 kb HuRBP1 riboprobe. HuRBP2 was obtained by cloning of a 0.65 kb *Hin*dIII–*Sma*I fragment corresponding to PR264/SC35 exon 2 and flanking sequences into pB KS(+) and linearization at the *Hin*dIII site. The 0.47 kb *Hin*cII–*Hin*cII fragment containing sequences located 0.9 kb upstream from the PR264/SC35 gene transcription start site was inserted in the pB SK(+) vector and linearized at the polylinker *Eco*RI site to generate HuRBP3. The MuRBP1 riboprobe was generated by cloning of a 0.25 kb *Sma*I–*Hin*dIII fragment containing genomic sequences located 0.8 kb upstream from the murine PR264/SC35 first exon into the pB KS(+) vector and linearization at the polylinker *Bam*HI site. The 0.23 kb MuRBP2 riboprobe was obtained following insertion of a 0.52 kb *Nar*I–*Nar*I fragment containing murine PR264/SC35 exon 1 into pB KS(+) and linearization at an internal *Nco*I site. *In vitro* transcription was performed with T7 RNA polymerase (Promega). Samples of polyadenylated RNA were hybridized overnight at 50°C with 2 ng of [α-^{32}P]UTP-labelled probe. RNA–RNA hybrids were digested with RNase A (60 µg/ml) and RNase T1 (1500 U/ml) (Boehringer Mannheim) and analyzed in a 6% sequencing gel.

cDNA cloning, library screening and nucleotide sequencing

A human thymus oligo dT-primed cDNA library was constructed from 5 µg of polyadenylated RNA according to the manufacturer (Amersham). 500 000 plaques were screened with the ^{32}P-labelled A DNA probe. Dideoxynucleotide sequencing reactions were performed with [α-^{35}S] dATP and the T7 Pharmacia Sequencing kit. Sequence data treatments were performed as described elsewhere (34).

RT-PCR amplifications

First strand oligo dT-primed cDNA was synthesized from 5 µg of human thymus polyadenylated RNA with the Amersham cDNA synthesis kit. Four percent of the cDNA was used for amplification in the presence of 0.8 µM of each amplimer (H46: 5′-GAGACCGCAGCTTTAAAGGG-3′, H35: 5′-CTTGGACT-TGGACCTTCGTG-3′, H17: 5′-CGGCGGCTGTGGTGTGA-GT-3′, H60: 5′-CCACAAGTTTGAAAGGCAGTG-3′, H61: 5′-GATGGGCACTTCAAAGGAGG-3′ and H81: 5′-GGCAG-GATCTTCCAACAAGAGCAGG-3′), 500 µM dNTP, 50 mM Tris–HCl pH 9.2, 14 mM $(NH_4)_2SO_4$, 3 mM $MgCl_2$, 2% DMSO, 0.1% Tween 20 and 2.5 U Expand enzyme mixture (Boehringer Mannheim). Amplification procedure was as follows: 1 cycle of 2 min at 95°C; 10 cycles of 30 s at 95°C, 30 s at T_m – 8°C (T_m = 4[G+C] + 2[A+T]), 3 min at 68°C; 25 cycles of 30 s at 95°C, 30 s at T_m – 8°C, 3 min plus 20 s per cycle at 68°C; and 1 cycle of 7 min at 68°C.

Murine cDNA, synthesized from 2 µg of testis polyadenylated RNA, was amplified in the presence of 0.24 µM of the murine

M15 (5′-CTGCTGGCCTCTGAGCAAGCG-3′) and human H95 (5′-CCCATTCCACAGTGAAG-3′) amplimers.

RNA ligation mediated PCR (RLM-PCR)

Human thymus polyadenylated RNA (1 µg), treated with 1 U tobacco acid pyrophophatase (TAP, Epicentre) for 30 min at 37°C, was ligated to 600 ng of *in vitro* synthesized SK-T7 RNA oligoribonucleotide as previously described (35,36). Twenty percent of the ligation product was annealed to 40 ng of specific primer (H70: 5′-CGCCGGGATCCTATTTAAGCTCCTGATG-AC-3′ or H60: 5′-CCACAAGTTTGAAAGGCAGTG-3′) or 300 ng of a hexanucleotide random primer mixture by incubation for 5 min at 95°C and 45 min at 42°C in 1× reverse transcriptase buffer and 10 mM DTT. First strand cDNA synthesis was performed by addition of 200 U of Superscript reverse transcriptase (Gibco-BRL) and dNTP (0.5 mM) and incubation for 45 min at 42°C according to the manufacturer. The resulting first strand cDNA was used for PCR amplification in the presence of 250 ng SK primer, 200 ng specific primer (H70 or H60), 200 µM dNTP and 1.25 U *Taq* DNA Polymerase (Appligene). This procedure was followed by a nested PCR performed in the same conditions with SK primer and an internal 3′ oligonucleotide (H63: 5′-GGTGGGATCCGGGGACA-TTTTGGCTCAG-3′). PCR reactions were as follows: 1 cycle of 2 min at 94°C; 30 cycles of 1 min at 94°C, 2 min at 58°C, 3 min at 70°C; and 1 cycle of 7 min at 70°C.

Rapid amplification of cDNA ends (RACE-PCR)

RACE-PCR experiments were performed with the Marathon cDNA amplification kit (Clontech) according to the manufacturer's instructions. Double-stranded cDNA was amplified in the presence of 0.8 µM of each amplimer (H60: 5′-CCACAAGTTT-GAAAGGCAGTG-3′ or H80: 5′-CTACAATGGCAACCACT-GACG-3′ specific primers and AP1 (adaptor primer) as described for RT-PCR amplifications.

In situ hybridization

CCRF-CEM cells (3×10^5) were cytospun on siliconized slides, fixed in 4% paraformaldehyde (20 min) and dehydrated in a graded ethanol series. Oligonucleotide probes specific for ET (ETHS: 5′-CACGACAAGCGCGACGCTGAGGACGCTATG-3′) and GAPDH transcripts were tailed with digoxigenin-11-dUTP (Boehringer Mannheim). The PR264/SC35-specific oligonucleotide probe (PR2HS: 5′-CGACGAGGACTTGGACTTGGACC-TTCGTGC-3′) was tailed with biotin-16-dUTP (Boehringer Mannheim). GAPDH oligonucleotides were detected with alkaline phophatase-conjugated anti-digoxigenin antibodies (Boehringer Mannheim).

Double in situ labelling. Each slide was hybridized in 20 µl of hybridization mixture (Boehringer Mannheim) containing 100 ng of ETHS and PR2HS probes. Following overnight incubation at 37°C in a humid atmosphere, slides were processed as recommended (Boehringer Mannheim). Slides were incubated for 2 h with colloidal gold-conjugated anti-biotin antibodies (Aurion, diluted 1/150 in PBS, 0.5% BSA), washed in PBS (10 min) and in distilled water (5 × 1 min). Fixation was for 20 min in 2% glutaraldehyde, prior incubation for 2 h with phosphatase-conjugated anti-digoxigenin antibodies (diluted 1/2500). Slides were dipped in color substrate solution at room temperature and staining was stopped in 1 mM EDTA, 10 mM Tris–HCl pH 8.0 solution. Slides were rinsed in distilled water, treated with silver enhancer (BioCell) for 20 min and with distilled water for 1 min. Examination of the slides was performed with a Zeiss Axiophot microscope under direct light and/or epipolarization.

RESULTS

The human PR264/SC35 locus is bidirectionally transcribed

As a first step in the characterization of the transcripts expressed in the PR264/SC35 opposite orientation, RNase protection experiments were performed with polyadenylated RNA purified from human normal thymic tissue and cells of myelomonocytic (HL60) and epithelial (HeLa and 293) origin. Following hybridization with the HuRBP1 riboprobe (Fig. 1A), several fragments ranging from 170 to 115 nt in size were protected, although at different levels, with all RNA samples (Fig. 1B). This indicates that ET RNA species complementary to the first PR264/SC35 coding exon are expressed in HL60, HeLa, 293 and normal thymic cells. A similar experiment performed with the HuRBP2 riboprobe revealed that a major (185 nt) and two minor (170 and 155 nt) fragments were protected with thymic RNA (Fig. 1C). These fragments were undetectable in HeLa, HL60 and 293 samples. Five protected fragments (470, 370, 310, 230 and 180 nt) were detected following RNase protection analysis performed with the HuRBP3 riboprobe located upstream from the PR264/SC35 promoter sequences (Fig. 1D). No specific protection was observed upon use of an HuRBP3 complementary riboprobe (data not shown). Whether these protected RNA fragments may also correspond to the ET transcriptional unit was assessed by RT-PCR experiments performed with thymic RNA and the H17-H61 oligonucleotide primers (Fig. 4). Two amplified cDNAs (441 and 446 bp) corresponding to alternative ET RNAs differing by their splice donor sites were obtained (depicted as ET31 and ET32 in Fig. 4). Taken together, these observations indicate that ET transcription results in the expression of several RNA species overlapping both PR264/SC35 coding exons and extending upstream from the PR264/SC35 promoting region. Furthermore, our results suggest that ET expression may be regulated in a tissue-specific way.

We took advantage of the ET transcription extending outside the PR264/SC35 sequences to derive an ET-specific probe (probe A in Figs 1 and 4). Northern blot analyses performed with the A DNA probe allowed the detection of four major ET RNA species (3.2, 2.7, 2.4 and 2.1 kb) in the polyadenylated RNA fraction of thymic cells (Fig. 2). In HL60, CCRF-CEM and HeLa cells, the major ET transcripts were 3.0 and 2.4 kb in size (Fig. 2). Four transcripts of higher molecular weight (12.0, 11.0, 10.0 and 6.8 kb) were also revealed in CCRF-CEM human T lymphoma cells.

Simultaneous detection of ET and PR264/SC35 transcripts in human hematopoietic cells

Double labelling *in situ* hybridization experiments were performed in CCRF-CEM human T-lymphoma cells with the ETHS and PR2HS oligonucleotide probes complementary to ET and PR264/SC35 sequences, respectively (Fig. 1A). Digoxigenin end-labelled ETHS oligonucleotides were revealed as brownish-purple areas (Fig. 3A), while biotin end-labelled PR2HS oligonucleotides were revealed as light blue spots under direct

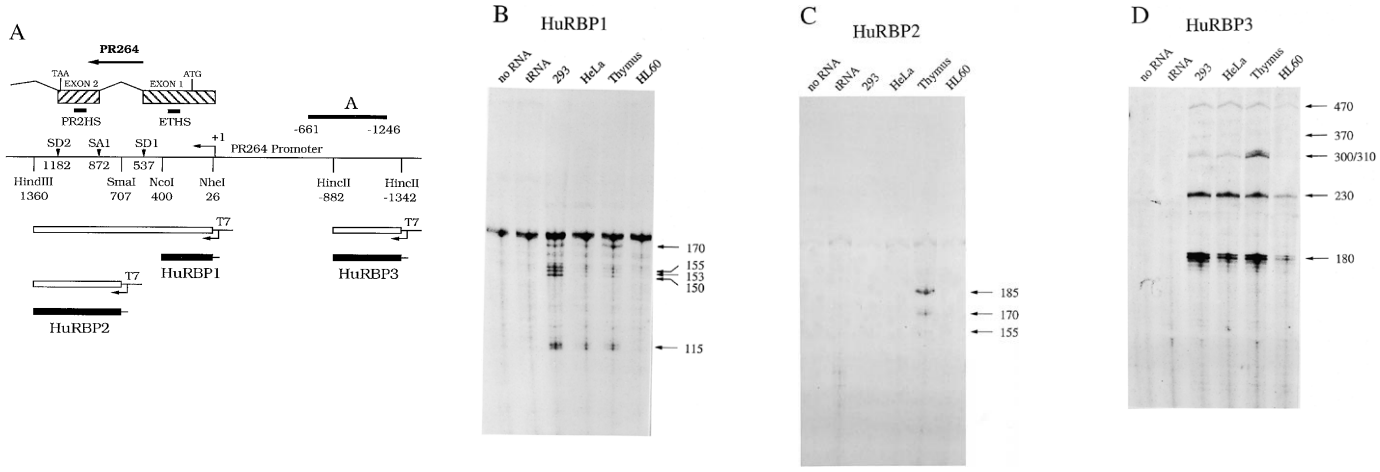

Figure 1. RNase protection analyses of human ET RNA species. (**A**) The structure of the RNA probes and their correspondence with the PR264/SC35 sequences. Pertinent restriction sites, PR264/SC35 transcription start site (+1), coding exons, splice donor (SD1, SD2) and acceptor (SA1) sites are indicated and landmark nucleotide numbers are mentioned. The A probe used for northern blot analysis and the oligonucleotide probes (PR2HS and ETHS) used in double *in situ* hybridization experiments are indicated. RNase protection experiments were performed with HuRBP1 (**B**), HuRBP2 (**C**) or HuRBP3 (**D**) probe and 5 µg of the indicated polyadenylated RNA. The specifically protected fragments are indicated by arrows. Sizes are in nucleotides. Protected fragments (180 nt with HuRBP1, 230 nt with HuRBP2) observed in control lanes ('no RNA' and 'tRNA') and resulting from autoprotection of the probe were not considered.

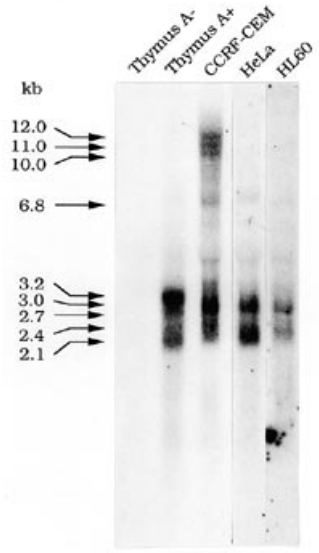

Figure 2. Northern blot analysis of human ET RNA species. Non-polyadenylated (thymus A^-, 20 µg/lane) and polyadenylated (thymus A^+, CCRF-CEM, HeLa, HL60, 20 µg/lane) RNA samples of human origin were hybridized with the ^{32}P-labelled DNA probe A. Transcripts detected in the different samples are indicated by arrows. Sizes are in kilobases.

light plus epipolarization or as brilliant blue gold particles under epipolarized illumination only (Fig. 3B).

Approximately 95% of the CCRF-CEM cells exhibited the specific coloration of both probes in their cytoplasm indicating that expression of ET and PR264/SC35 overlapping sequences is not mutually exclusive. In a few cells, the signal obtained with either the ETHS or the PR2HS probe was faintly detected, suggesting that expression of the corresponding gene was arrested or very low (Fig. 3C). Similar experiments performed with the human myelomonocytic HL60 cells also led to the codetection of ET and PR264/SC35 transcripts in the majority of the cells (data not shown).

Characterization of ET RNA species

Screening of a cDNA library constructed from human normal thymic polyadenylated RNA allowed us to isolate three cDNA clones, ET12 (1.15 kb), ET10 (1.94 kb) and ET16 (1.9 kb). Since 50 PR264/SC35 cDNAs were isolated following parallel screening of the non-amplified thymic library, the small number of ET clones likely reflected the low abundance of ET RNA in thymic cells. Comparisons of genomic and cDNA nucleotide sequences suggest that ET12 corresponds to an unspliced polyadenylated ET RNA species (Fig. 4). ET10 and ET16 cDNAs share a common overall structure but differ at the level of their 5′-proximal sequences and by the inclusion of alternative sequences in ET16 (delineated by the SD8 and SA5 splice sites in Fig. 4). These clones correspond to ET RNA species ending at a polyadenylation site distal to that identified with the ET12 cDNA and are constituted by at least five and six exons, respectively.

Potential ORFs of 449 and 397 amino acids (aa) were identified in the sequence of ET16 and ET10 cDNA clones, respectively. The ET10 ORF differs from that identified in the ET16 by the lack of 52 aa encoded by the most 3′-proximal alternative sequences represented in ET16. As shown in Figure 5, the ET16 and ET10 predicted protein isoforms, whose calculated molecular weight is 49 and 43 kDa, respectively, exhibit a high content of leucine, serine and phenylalanine residues. ET16 and ET10 predicted proteins share significant homologies (up to 34%) with five different ORFs from *Caenorhabditis elegans* genes whose respective functions are unknown.

Since the screening of the thymic cDNA library did not lead to the characterization of clones corresponding to ET transcripts overlapping the PR264/SC35 coding sequences, RT-PCR experiments were performed with thymic polyadenylated RNA species and different primer pairs, each containing an arbitrary amplimer located within the PR264/SC35 gene. Sequencing of the resulting products allowed us to identify additional cDNA species (ET21, ET41 and ET51 in Fig. 4) corresponding to ET RNAs complementary to the PR264/SC35 sequences. As already

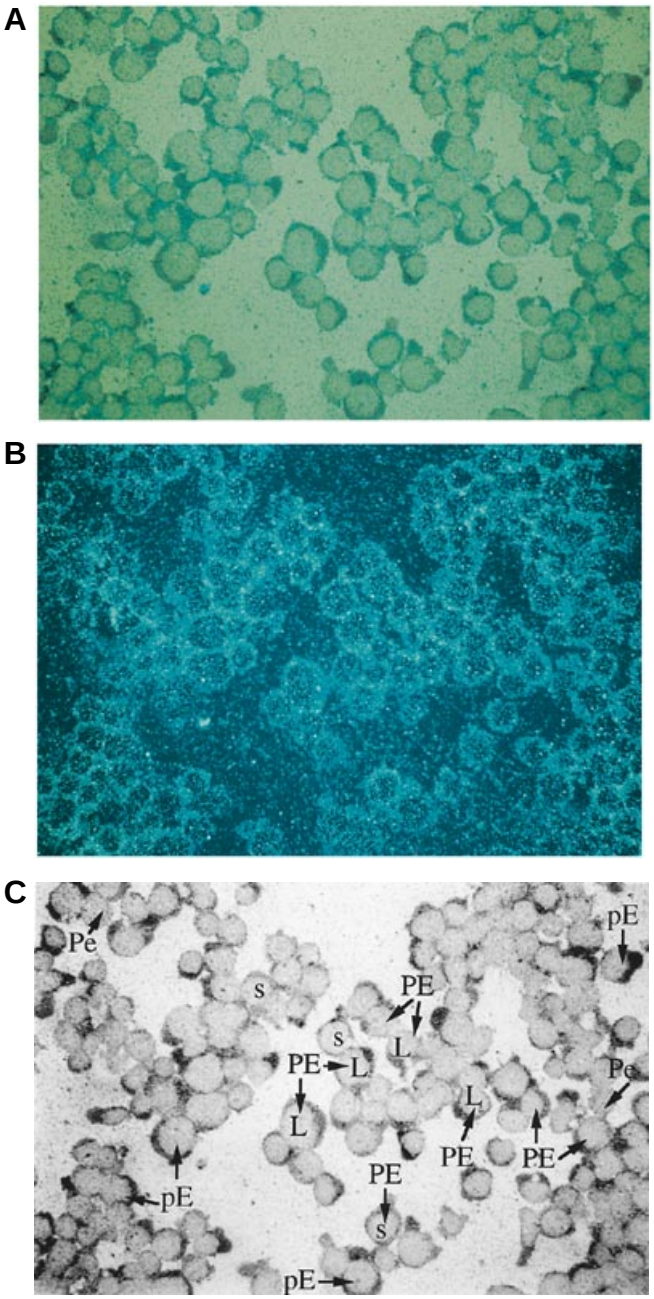

Figure 3. Simultaneous *in situ* hybridization detection of ET and PR264/SC35 transcripts in CCRF-CEM cells. (**A**) Double labelling observed under direct light plus epipolarized illumination (magnification, ×400) showing the brownish-purple coloration of the alkaline phosphatase precipitate (ET RNA detection) and the light blue gold particles (PR264/SC35 mRNA detection) in the cytoplasm of the cells. (**B**) Double labelling observed under epipolarized illumination only. PR264/SC35 transcripts are visualized by brilliant blue gold particles. (**C**) Laser reproduction of (A) showing examples of different cells which exhibit high levels of PR264/SC35- and ET-specific labelling (PE) or low levels of either PR264/SC35- or ET-specific labelling (pE and Pe, respectively). In cells exhibiting large cytosol areas (L), the blue colour of the gold grains appears to be attenuated by the alkaline phosphatase staining (A), as confirmed by the detection of the gold particles with epipolarized light only (B). In other cells containing smaller cytoplasmic area(s), the blue colour revealed with the combination of the two light sources (A) results from the concentration of gold grains. Since hybridization signals were not normalized, the amount of gold grains and the intensity of the brownish-purple coloration of the alkaline phosphatase precipitate cannot be compared. Control slides do not show the presence of endogenous phosphatase (data not shown). Positive and negative hybridization controls were checked with sense- and antisense-specific GAPDH probes (data not shown).

observed for ET31, ET21 and ET41 species contain at least two products resulting from the use of the SD2a (ET21, ET31, ET41) and SD2b (ET22, ET32, ET42) alternative splice donor sites (Fig. 4). These products, differing in size by 5 nt, likely account for the 150 and 155 nt fragments specifically protected with the HuRBP1 probe (Fig. 1B). The 5′-proximal sequences of ET21 and ET31 are complementary to the first PR264/SC35 coding exon, while those of ET41 overlap coding, non-coding and intronic sequences of the PR264/SC35 gene. The ET51 RT-PCR product corresponds to an unspliced ET transcript since we have verified that it does not result from amplification of residual genomic DNA (data not shown). An additional ORF (282 aa), identified in the two first exons of ET41 and ET42 cDNA clones, did not share any significant homology with proteins listed in the current databases.

These observations establish that the transcription proceeding in the orientation opposite to that of the PR264/SC35 gene gives rise to the expression of numerous ET RNA species resulting from the use of alternative splicing and polyadenylation signals.

Characterization of ET transcriptional start sites

To undertake the characterization of the transcription start sites corresponding to the different ET transcripts identified thus far, RLM- and RACE-PCR experiments were performed with human thymic polyadenylated RNA. Sequencing of the resulting products revealed that the potential start sites of ET mRNAs corresponding to the ET16 cDNA species (ET81 and ET101, Fig. 4) are spread over 110 nt located within the PR264/SC35 promoter region. In as much as the ET16 5′-end is located in the same region, these results suggest that this clone represents a complete cDNA. The ET61 cDNA corresponds to a 3′-truncated version of ET21. However, it is difficult to ascertain whether its 5′-end defines a genuine transcription start site since only one clone was obtained.

These experiments also led us to identify additional cDNA species: ET71 differing from ET10 by the use of an SD3 instead of SD4 splice donor site and ET91 differing from the ET16 cDNA by the retention of the SD5-SA2 intron. This intron was also identified in ET111 cDNAs which could belong to either the ET12/ET51 or ET91 cDNA species. The latter hypothesis seems, however, more likely since the potential start sites of the RNAs corresponding to ET91 and ET111 cDNA species are clustered within a 170 nt region of the PR264/SC35 promoter.

Taken together, these results strongly suggest that ET transcripts are expressed from alternative promoters and that the PR264/SC35 promoting region exhibits a bidirectional transcriptional activity responsible for the expression of non-overlapping ET RNAs corresponding to the ET16, ET91 and, possibly, ET10 and ET71 cDNA species.

Analysis of the different ET cDNAs also indicates that their structure accounts for most of the major fragments protected with the HuRBP3 probe (Fig. 1D). Indeed, the 180 nt fragment likely corresponds to the SA2-SD6 sequences identified in cDNAs such as ET16 and ET10 (Fig. 4). The 230 nt and 300–310 nt fragments are in good agreement with the size of sequences delineated by the 3′-end of the probe and the SD6 splice site (such as in ET91), and by the 3′-end of the probe and the poly A1 site (such as in ET12), respectively. The 470 nt fragment resulting from protection of the full length probe likely reflects the existence of an unspliced ET transcript ending at a distal polyadenylation site. These observations suggest that the ET cDNAs characterized thus far are representative of most of the ET transcription products.

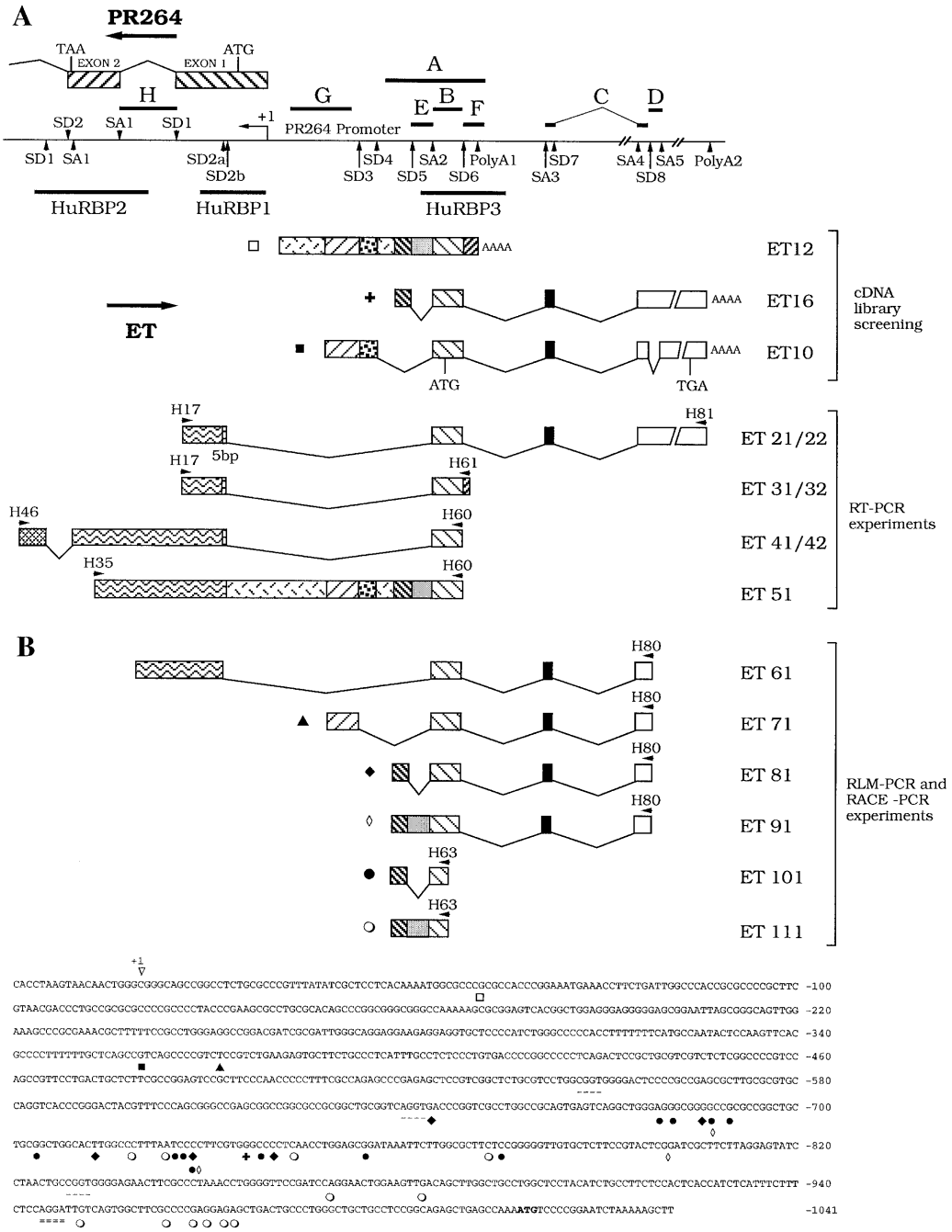

Figure 4. Schematic structure of the human ET cDNAs. (**A**) Oligonucleotides used in RT-PCR experiments are indicated at both ends of the corresponding products by horizontal arrowheads. The splice donors (SD), splice acceptors (SA) and polyadenylation signals (poly A) used to generate the various ET RNAs are in agreement with consensus sequences. The PR264/SC35 transcription start site (+1), coding exons, splice donor (SD1, SD2) and acceptor (SA1) sites are indicated. ET16 and ET10 translation initiation and stop codons are mentioned. Probes used for RNase protection experiments, northern blot hybridizations and cDNA library screening are positioned on the ET-PR264/SC35 locus. ET and PR264/SC35 transcriptional orientation are indicated by arrows. (**B**) Characterization of human ET RNA 5′-ends by RLM- and RACE-PCR experiments. The structure of the PCR products is drawn. Corresponding 5′-ends are indicated under the PR264/SC35 promoter sequence by symbols defining each class of cDNA. Specific oligonucleotides used in these experiments are represented at the 3′-end of the cDNA by arrowheads. The PR264/SC35 transcription start site defining the sequence numbering is indicated by +1. The 5′-end of the ET10, ET12 and ET16 cDNAs are represented by the corresponding symbol. ET splice donors SD3, SD4, SD5 are underlined and splice acceptor SA2 is underlined twice. The ET16 and ET10 translation initiation codon is in bold.

Differential expression of the human ET gene

In an attempt to identify more precisely the ET RNA species, northern blot analyses of polyadenylated RNA, purified from human thymic and HeLa cells, were performed with DNA probes specific for common or alternative regions of the different ET cDNAs (Fig. 4). Hybridization with the B probe, specific for all the ET cDNAs characterized thus far, led, in both cases, to the detection of five major RNA species, 2.1, 2.4, 2.7, 3.0 and 6.8 kb in size (Fig. 6). It is noteworthy that the fuzzy hybridization

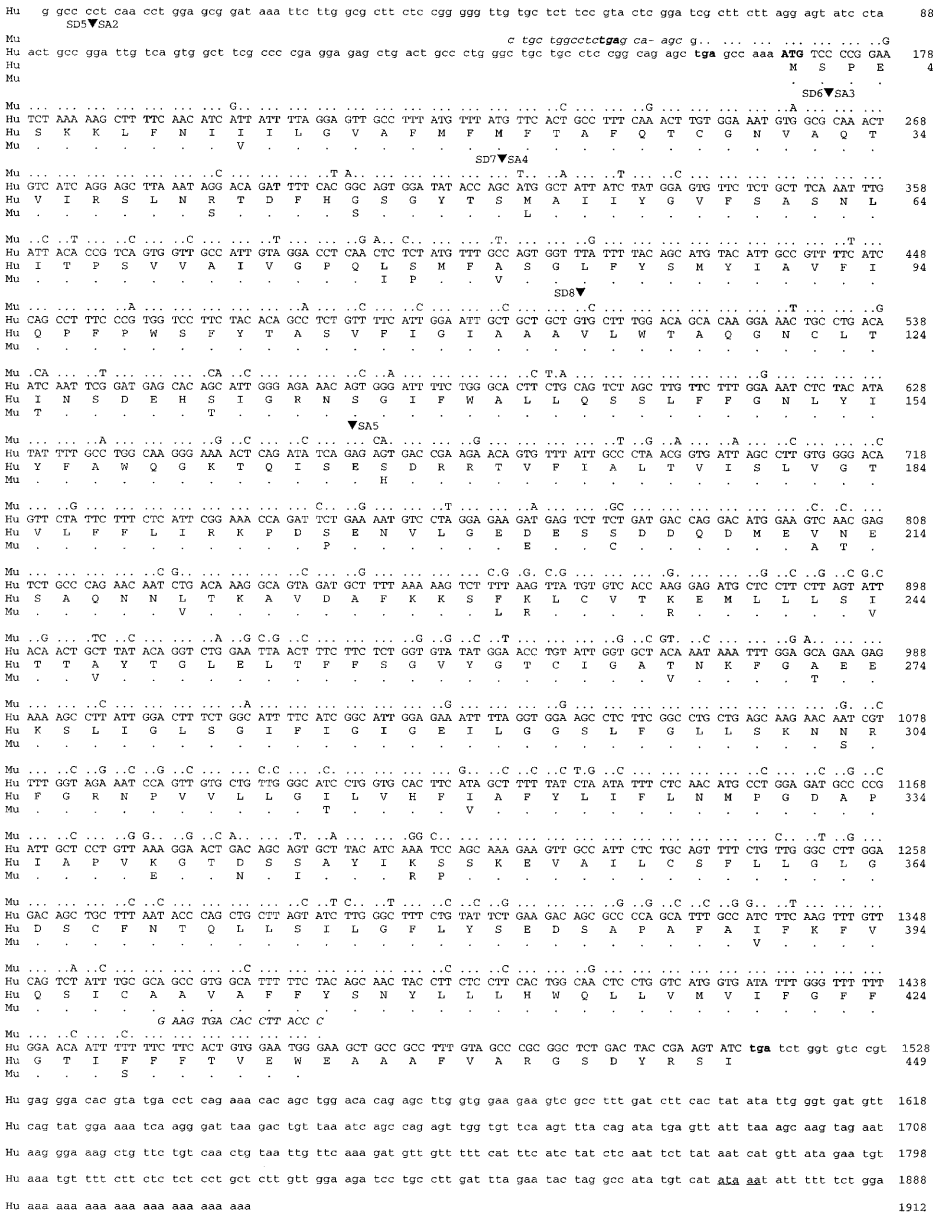

Figure 5. Nucleotide sequence of the human ET16 cDNA and deduced amino acid sequence of the predicted ET16 protein. The murine (Mu) ET8 cDNA nucleotide sequence and deduced amino acid sequence are aligned with the corresponding human sequences (Hu). Murine M15 and human H95 oligonucleotides used in ET8 RT-PCR amplification are indicated in italic. Only the divergent nucleotides and amino acids are indicated. Coding sequences are in capital letters. The translation initiation codon is in bold. In frame translation stop codons are in bold lowercase letters. Arrowheads indicate splice donor and acceptor sites. The polyadenylation signal is underlined. The 3′ alternative sequences missing in ET10 cDNA are delineated by SD8 and SA5 splice sites.

signals likely reflect the detection of ET transcripts slightly differing in size as a result of heterogeneous transcription start sites and/or the use of short alternative exons.

Following hybridization with the C and D probes specific for the common and alternative potential coding sequences, respectively, differences were observed between the relative expression levels of the ET RNA species detected in thymic and HeLa cells. Indeed, the 2.1 kb transcripts were predominantly detected in thymic cells, while the major species in HeLa cells was 2.7 kb in size. These results confirm that the ET gene is differentially expressed in HeLa and normal thymic cells. Furthermore, the similar hybridization patterns obtained with the C and D probes indicate that ET transcripts lacking the SD8-SA5 alternative sequences detected by the D probe (Fig. 4) are likely expressed at a low level in HeLa and normal thymic cells.

The differential expression of the ET gene in HeLa and normal thymic cells was clearly observed when northern blot analyses were carried out with the E, F and G probes specific for alternative sequences represented in a subset of ET cDNA species. Both the E and G probes preferentially revealed a 2.7 and a 3.2 kb RNA species in HeLa and normal thymic cells, respectively (Fig. 6). Interestingly, the F probe corresponding to alternative sequences located immediately downstream from the SD6 splice site revealed a major 3.2 kb ET transcript in thymic cells. No ET RNA

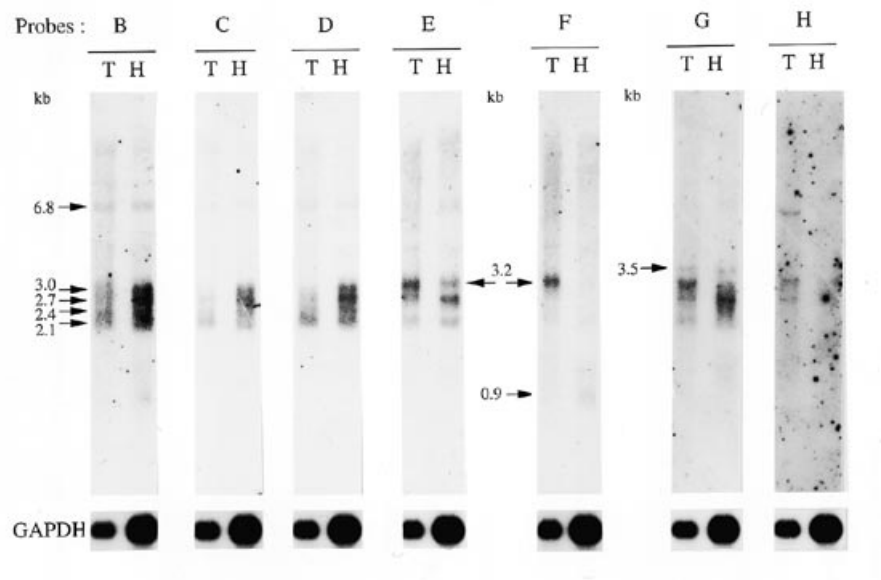

Figure 6. Characterization of the alternative ET RNA species. Northern blot analyses of polyadenylated human thymic (T) and HeLa (H) RNA samples (10 µg/lane) were performed with the indicated ^{32}P-labelled probes. The size of the detected transcripts is indicated in kilobases. GAPDH was used as an internal control to normalize for variations in RNA amounts.

was clearly detected in HeLa cells with the same probe. These observations suggest that the use of the proximal alternative polyadenylation site (poly A1) is regulated in a tissue-specific way.

As a means to identify the ET RNA species overlapping the PR264/SC35 transcribed sequences, HeLa and thymic polyadenylated RNA samples were also analysed with the H probe (Fig. 4) corresponding to the PR264/SC35 first intron. As shown in Figure 6, three ET mRNAs (2.7, 3.2 and 6.8 kb) were faintly detected in thymic cells. These transcripts were hardly detectable in HeLa cells despite the apparently higher RNA amount transferred on the blot. These observations, which are in good agreement with the results of RNase protection experiments performed with the HuRBP2 riboprobe, indicate that PR264/SC35-overlapping ET RNA species are preferentially expressed in thymic cells but very weakly represented as compared to PR264/SC35-non overlapping ET transcripts.

Conservation and tissue-specific expression of ET sequences in the mouse

To determine whether the bidirectional transcription of the PR264/SC35 locus is an evolutionarily conserved mechanism, polyadenylated RNAs purified from different adult murine tissues were analyzed by northern blot hybridization with the MuBH probe (Fig. 7A). As shown in Figure 7B, a 2.3 kb and a minor 3.9 kb mRNA species were detected, although at different levels, in all tested tissues. The higher signals were obtained with testis and kidney RNA while both species were faintly detected in spleen, adrenal gland and thymic samples. Interestingly, an abundant (2.5 kb) and two minor (7.4 and 0.8 kb) mRNA species were exclusively observed in testis.

Whether these transcripts correspond to ET RNAs was assessed by RNase protection experiments performed with the MuRBP1 riboprobe derived from the MuBH fragment. As shown in Figure 7C, two major (100 and 105 nt) and one minor (250 nt) protected fragments were obtained with testis mRNA, while no specific protection was observed with the MuRBP1 complementary riboprobe (data not shown). Interestingly, the size of the protected fragments is in agreement with that expected from murine ET RNA species corresponding to the human ET10, ET16 (major fragment) and ET12 (minor fragment) cDNAs. Similar observations were made with a second probe (MuRBP2) derived from the first PR264/SC35 coding exon (Fig. 7D). Again, the size of the major protected fragment (145–150 nt) is in accordance with that expected from murine ET RNA species corresponding to the human ET21/22 and ET41/42 cDNAs.

Taken together, these observations indicate that the bidirectional transcription leading to the expression of ET transcripts containing PR264/SC35-overlapping and non-overlapping sequences is a conserved mechanism and that ET expression is regulated in a tissue-specific way. Moreover, our results suggest that the overall structure of ET transcription products is conserved in both human and murine species.

In order to further investigate the conservation level of murine ET sequences, RT-PCR experiments were performed with testis polyadenylated RNA and amplimers delineating most of the potential ORF identified in the human ET16 cDNA. Sequence analysis of the resulting products (Fig. 5) indicates that the human and murine sequences are 87% homologous. Furthermore, comparison of the predicted proteins encoded by the two cDNAs reveals a much higher homology (93%).

DISCUSSION

The studies presented here demonstrate that the human and murine PR264/SC35 loci are subjected to bidirectional transcription leading to the expression of the PR264/SC35 alternative splicing factor and ET transcripts whose biological role remains to be established.

Our results indicate that the ET gene is expressed in the form of several polyadenylated RNA species resulting from the use of alternative splice and polyadenylation sites. Furthermore, RLM- and RACE-PCR experiments have suggested that the PR264/SC35

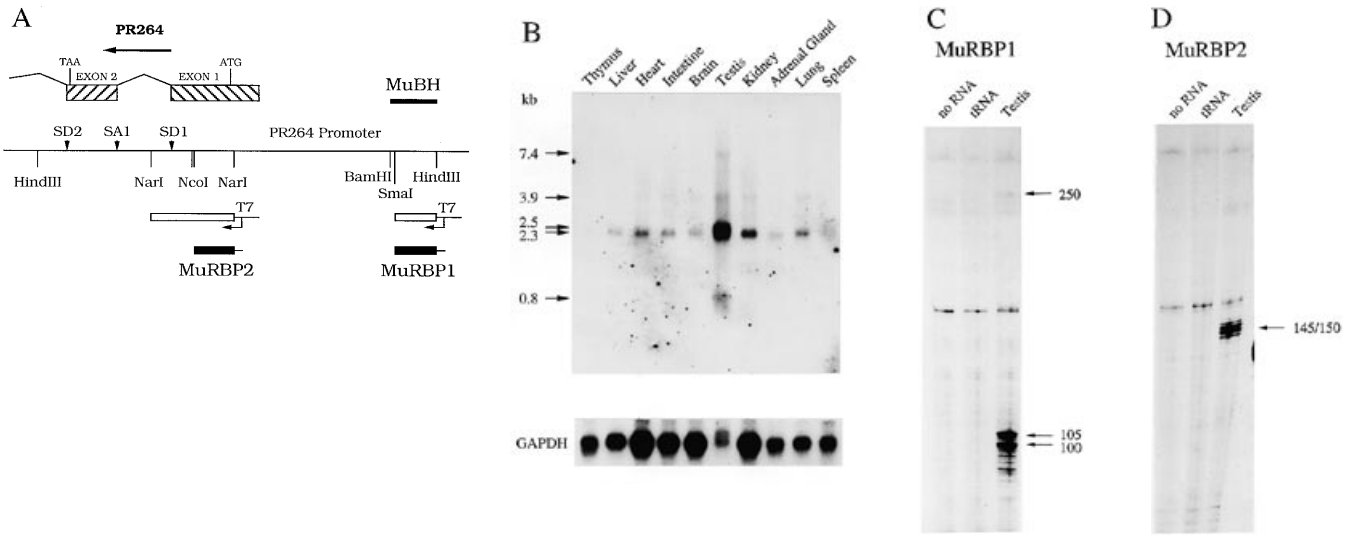

Figure 7. Conservation and expression of ET sequences in murine tissues. (**A**) The correspondence of the DNA and RNA probes with the murine genomic PR264 sequences. The pertinent restriction sites are mentioned. The PR264 coding exons, splice donor (SD1, SD2) and acceptor (SA1) sites are indicated. (**B**) Northern blot analysis of polyadenylated RNA samples from murine tissues (5 µg/lane) was performed with the ^{32}P-labelled MuBH DNA probe. Sizes are in kilobases. GAPDH was used as an internal control to normalize for variations in RNA amounts. RNase protection experiments were performed with MuRBP1 (**C**) or MuRBP2 (**D**) RNA probe and 3 µg murine testis polyadenylated RNA. The specifically protected fragments are indicated by arrows. Sizes are in nucleotides. Protected fragments (160 nt with HuRBP1 and HuRBP2) observed in control lanes ('no RNA' and 'tRNA') were not considered.

promoting region is responsible for the expression of a subset of ET transcripts. Consistent with this, preliminary results obtained with the luciferase reporter system have suggested that the PR264/SC35 promoter region might exhibit a bidirectional transcriptional activity (V.Pierron, J.Soret and B.Perbal, unpublished results).

A striking feature of the bidirectional promoters described thus far is the absence of a consensual TATA box in either orientation (37–39). In most cases, GC clusters, being the hallmark of house-keeping gene promoter regions (40), have been shown to drive transcription in both orientations (27,29). We have previously reported that the PR264/SC35 promoter region contains a canonical TATA box and that PR264/SC35 transcription is initiated at a major start site (41). Analysis of the nucleotide sequences located upstream from this start site has revealed that the PR264/SC35 promoting region also contains GC clusters which could represent regulatory elements involved in the expression of ET RNA species corresponding to the ET16, ET81 and ET91 cDNAs. This hypothesis is supported by our findings that the potential start sites mapped by RLM- and RACE-PCR experiments are heterogeneous and spread over a 300 nt region, two features generally observed with transcripts expressed from GC box-containing promoters (42). The presence of a consensual TATA box in the PR264/SC35 promoter raises interesting questions concerning the regulation and coordination of bidirectional transcription. Indeed, analysis of the α1 (IV) and α2 (IV) collagen gene promoter has revealed that a TATA box directs transcription to start mainly downstream from it, while it can simultaneously decrease the initiation of transcription in the opposite direction in a bidirectional promoter (27). Along this line, the TATA box governing PR264/SC35 expression could hinder transcription in the ET orientation and account for the low expression level of ET RNA species expressed from the bidirectional promoter.

We have characterized three main classes of ET transcripts. In the first class, ET RNA species contain PR264/SC35-overlapping sequences and are polyadenylated at the proximal site identified in this study (poly A1 in Fig. 4). The promoting region governing expression of these transcripts remains to be localized. The second class consists of ET mRNA species expressed from the PR264/SC35 bidirectional promoter and ending at the distal polyadenylation site (poly A2 in Fig. 5). The third class corresponds to ET transcripts sharing common features with RNAs of the first (PR264/SC35-overlapping sequences) and second class (PR264/SC35-non-overlapping sequences and polyadenylation site).

Since *in situ* hybridization experiments have allowed the simultaneous detection of ET and PR264/SC35 transcripts in CCRF-CEM and HL60 human hematopoietic cells, the existence of two classes of ET RNA species containing PR264/SC35-overlapping sequences raises the question of whether such transcripts could act as riboregulators of PR264/SC35 expression. Numerous studies undertaken in prokaryotes (13) and in higher eukaryotes (14) have revealed that the transcription of complementary RNA species constitutes a powerful means of regulating gene expression. Since formation of RNA–RNA duplexes in several eukaryotic systems was shown to result in alterations of splicing (20,21) or stability (15,16) of the reacting transcripts, expression of several alternative ET transcripts complementary to different regions of PR264/SC35 mRNAs might be of significance.

Transfection experiments performed in our laboratory, however, indicate that overexpression of ET RNA species corresponding to the ET41 cDNA does not result in a significant alteration of the PR264/SC35 protein level in HeLa cells (data not shown). Given that the ET gene is expressed through numerous alternative mRNAs, these preliminary observations could indicate that not all forms of ET transcripts exhibit a regulatory role. Another hypothesis could be inferred from the work of Wang *et al.* (10) showing that artificial modifications of the ASF/SF2 alternative splicing factor level in avian DT40 cells are counteracted by compensatory mechanisms affecting expression of the endogenous gene. In the light of these results, one can envisage that a marked alteration of the PR264/SC35 expression level cannot be

achieved in HeLa cells in response to overexpression of PR264/SC35-overlapping exogenous ET sequences. On the other hand, it is also possible that ET RNA species complementary to PR264/SC35 transcripts are not involved in the regulation of PR264/SC35 expression. Indeed, northern blot analyses of the ET (this study) and PR264/SC35 (data not shown) expression in different murine tissues did not reveal the existence of an obvious inverse correlation between the respective levels of both RNA species. Such a lack of correlation between *in vivo* levels of sense and antisense RNAs has also been reported for the c-erbA gene (43) despite previous *in vitro* splicing experiments showing that an excess of Rev-erbAα sequence complementary to that of the THRα5 transcript blocked splicing of this RNA (20). This suggests that the biological significance of bidirectional transcription is more complex than expected from *in vitro* observations.

Analyses of different human cell lines and murine tissues have revealed that expression of ET RNA species is regulated in a tissue-specific way at both the qualitative and the quantitative levels. This is especially obvious in murine testis cells which specifically express an abundant 2.5 kb ET RNA as well as two minor transcripts of 7.4 and 0.8 kb. The generation of multiple and unique transcripts in the testis is not limited to ET but has been reported for a number of other genes (reviewed in ref. 44). Whether this preferential expression reflects a role of ET products in the regulation of testis-specific biological processes remains to be established.

Sequence analyses of the human ET cDNAs characterized in this study have allowed us to identify two ORFs encoding potential proteins of 28 kDa (ET41 ORF) and 49–43 kDa (ET16 and ET10 ORF, respectively). The predicted product encoded by the ET41 cDNA showed no homology with proteins listed in the current databases and we have not determined whether the corresponding ORF is conserved in the mouse. On the other hand, the potential protein isoforms encoded by the ET10 and ET16 cDNAs share significant homologies with the translation products of five *C.elegans* genes likely belonging to the same family. RT-PCR experiments have allowed us to characterize a murine ET16 homolog and to show that the corresponding protein is highly conserved between the two species. Such a conservation suggests that the proteins encoded by ET16 and ET10 mRNA species play a specific role. In the light of several studies showing that bidirectionally transcribed genes such as c-erbA (23,24), α-1IV/α-2IV collagen (27), histidyl tRNA synthetase (28) and Surf1/Surf2 (29) encode structurally and functionally related products, it will be interesting to examine whether ET proteins are involved in splicing processes.

Further studies should allow the characterization of ET-encoded proteins and determine the role, if any, of the ET transcripts in the regulation of PR264/SC35 expression.

ACKNOWLEDGEMENTS

Thanks are due to Drs J.Stevenin and D.Lawrence for critical reading of the manuscript. We are also grateful to Pr. B.Roizman, Dr J.Stevenin, Dr C.Martinerie, Dr M.Laurent and G.Chevalier for helpful discussions and comments. These studies were done in the Laboratoire d'Oncologie Virale et Moléculaire at the Institut Curie, Orsay, France and were supported by grants from Centre National de la Recherche Scientifique (UMR 146), Association pour la Recherche contre le Cancer and Ligue Nationale Contre le Cancer (Comité National and Comité du Cher).

REFERENCES

1 Fu, X.D. (1995) *RNA*, **1**, 663–680.
2 Manley, J.L. and Tacke, R. (1996) *Genes Dev.*, **10**, 1569–1579.
3 Zhang, W.J. and Wu, J.Y. (1996) *Mol. Cell. Biol.*, **16**, 5400–5408. .
4 Fu, X-D, and Maniatis, T. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 1725–1729.
5 Fu, X-D, Mayeda, A;, Maniatis, T.and Krainer, A. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 11224–11228.
6 Caceres, J.F., Stamm, S., Helfman, D.M. and Krainer, A.R. (1994) *Science*, **265**, 1706–17091.
7 Gallego, M.E., Gattoni, R., Stevenin, J., Marie, J. and Expert-Bezançon, A. (1997) *EMBO J.*, **16**, 1772–1784.
8 Kraus, M.E. and Lis, J.T. (1994) *Mol. Cell. Biol.*, **14**, 5360–5370.
9 Ring, H.Z. and Lis, J.T. (1994) *Mol. Cell. Biol.*, **14**, 7499–7506.
10 Wang, J., Takagaki, Y. and Manley, J.L. (1996) *Genes Dev.*, **10**, 2588–2599.
11 Vellard, M., Soret, J., Viegas-Pequignot, E., Galibert, F., Van Cong, N., Dutrillaux, B. and Perbal, B. (1991) *Oncogene*, **6**, 505–514.
12 Sureau, A. and Perbal, B. (1991) *C. R. Acad. Sci. Paris*, **312**, 323–328.
13 Inouye, M., (1988) *Gene*, **72**, 25–34.
14 Krystal, G.W. (1992) in Erickson,R.P. and Izant,J.P. (eds) *Gene Regulation: Biology of Antisense RNA and DNA*. Raven Press, New York, USA.
15 Hildebrandt, M. and Nellen, W. (1992) *Cell*, **69**, 197–204.
16 Kimelman, D. and Kirschner, M.W. (1989) *Cell*, **59**, 687–696.
17 Malik, K.T., Wallace, J.I., Ivins, S.M. and Brown, K.W. (1995) *Oncogene*, **11**, 1589–1595.
18 Wagner, R.W., Smith, J.E., Cooperman, B.S. and Nishikura, K. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 2647–2651.
19 Bass, B.L. and Weintraub, H. (1988) *Cell*, **55**, 1089–1098.
20 Munroe, S.H. and Lazar, M.A. (1991) *J. Biol. Chem.*, **266**, 22083–22086.
21 Krystal, G.W., Armstrong, B.C. and Battey, J.F. (1990) *Mol. Cell. Biol.*, **10**, 4180–4191.
22 Khochbin, S. and Lawrence, J.J. (1989) *EMBO J.*, **8**, 4107–4114.
23 Miyajima, N., Horiuchi, R., Shibuya, Y., Fukushige, S., Matsubara, K., Toyoshima, K. and Yamamoto, T. (1989) *Cell*, **57**, 31–39.
24 Lazar, M.A., Hodin, R.A., Darling, D.S. and Chin, W.W. (1989) *Mol. Cell. Biol.*, **9**, 1128–1136.
25 Volk, R., Koster, M., Poting, A., Hartmann, L. and Knochel, W. (1989) *EMBO J.*, **8**, 2983–2988.
26 Li, A.W., Too, C.K. and Murphy, P.R. (1996) *Biochem. Biophys. Res. Commun.*, **223**, 19–23.
27 Heikkila, P., Soininen, R. and Tryggvason, K. (1993) *J. Biol. Chem.*, **268**, 24677–24682.
28 O'Hanlon, T.P., Raben, N. and Miller, F.W. (1995) *Biochem. Biophys. Res. Commun.*, **210**, 556–566.
29 Lennard, A.C. and Fried, M. (1991) *Mol. Cell. Biol.*, **11**, 1281–1294.
30 Spicer, D.B. and Sonenshein, G.E. (1992) *Mol. Cell. Biol.*, **12**, 1324–1329.
31 Laabi, Y., Gras, M.P., Brouet, J.C., Berger, R., Larsen, C.J. and Tsapis, A. (1994) *Nucleic Acids Res.*, **22**, 1147–1154.
32 Dolnick, B.J. (1993) *Nucleic Acids Res.*, **21**, 1747–1752.
33 Sureau, A. and Perbal, B. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 932–936.
34 Dessen, P., Fondrat, C., Valencien, C. and Mugnier, C. (1990) *Comput. Appl. Biosci.*, **6**, 355–356.
35 Fromont-Racine, M., Bertrand, E., Pictet, R. and Grange, T. (1993) *Nucleic Acids Res.*, **21**, 1683–1684.
36 Liu, X. and Gorovsky, M.A. (1993) *Nucleic Acids Res.*, **21**, 4954–4960.
37 Farnham, P.J., Abrams, J.M. and Schimke, R.T. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 3978–3982.
38 Lavia, P., Macleod, D. and Bird, A. (1987) *EMBO J.*, **6**, 2773–2779.
39 Johnson, P. and Friedman, T. (1990) *Gene* (Amsterdam), **88**, 207–213.
40 Smale, S.T. and Baltimore, D. (1989) *Cell*, **57**, 103–113.
41 Sureau, A., Soret, J., Vellard, M., Crochet, J. and Perbal, B. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 11683–11687.
42 Melton, D.W. (1987) *J. Cell Sci.*, **88**, 267–270.
43 Jannini, E.A., Mitsuhashi, T. and Nikodem, V.M. (1992) *Biochem. Biophys. Res. Commun.*, **184**, 739–745.
44 Eddy, E.M., Welch, J.E. and O'Brien, D.A. (1993) in de Kretser,D. (ed.) *Molecular Biology of the Male Reproductive System*. Academic Press Inc., New York, NY, pp. 181–231.