

Alternative Splicing of the First Intron of the Steroid Receptor RNA Activator (SRA) Participates in the Generation of Coding and Noncoding RNA Isoforms in Breast Cancer Cell Lines

FLORENT HUBE,^{1,2,4} JIMIN GUO,^{1,4} SHILPA CHOONIEDASS-KOTHARI,^{1,4} CHARLTON COOPER,^{1,4}
MOHAMMAD K. HAMEDANI,^{1,4} ALEXANDER A DIBROV,^{1,4} ANNE A.A. BLANCHARD,^{2,3}
XUEMEI WANG,^{1,4} GEORGE DENG,^{1,2} YVONNE MYAL,^{2,3} and ETIENNE LEYGUE^{1,4}

ABSTRACT

The Steroid Receptor RNA Activator 1 (SRA1) has originally been described as a noncoding RNA specifically activating steroid receptor transcriptional activity. We have, however, identified, in human breast tissue, exon-1 extended SRA1 isoforms containing two initiating AUG codons and encoding a protein we called SRAP. We recently reported a decreased estrogen receptor activity in breast cancer cells overexpressing SRAP, suggesting antagonist roles played by SRA1 RNA and SRAP. SRA1 appears to be the first example of a molecule active both at the RNA and at the protein level. No data are currently available regarding the mechanisms possibly involved in the generation of coding and noncoding functional SRA1 RNAs. Using 5'-Rapid Amplification of cDNA Extremities (5'-RACE), we have herein identified several putative transcription initiation sites surrounding the second methionine codon and used to generate coding SRA1 transcripts. In the process, we also identified an alternatively spliced noncoding SRA1 transcript still containing an intron-1 sequence. Using targeted RT-PCR approaches, we confirmed the presence in breast cancer cell lines of SRA1 RNAs containing a full as well as a partial intron-1 sequence and established that the relative proportion of these RNAs varied within breast cancer cell lines. Using a "minigene" strategy, we also showed that artificial RNAs containing the SRA1 intron-1 sequence are alternatively spliced in breast cancer cell lines. Interestingly, the splicing pattern of the minigene products parallels the one of the endogenous SRA1 transcripts. Altogether, our data suggest that the primary genomic sequence in and around intron-1 is sufficient to lead to a differential splicing of this intron. We propose that alternative splicing of intron-1 is one mechanism used by breast cancer cells to regulate the balance between coding and functional noncoding SRA1 RNAs.

INTRODUCTION

THROUGH ITS MITOGENIC ACTION on breast epithelial cells, estrogen not only controls the growth and the development of the normal mammary gland, but also promotes breast tumorigenesis and breast cancer progression (for a review, see Jensen and Jordan, 2003). This estrogenic action is mainly mediated through two estrogen receptors (ERs) alpha and beta (Green *et al.*, 1986; Mosselman *et al.*, 1996), that belong to the steroid/thyroid/retinoic acid receptors superfamily and act as li-

gand-dependent transcription factors (Evans, 1988). Over the last few years, it has become apparent that the balance between coactivators and corepressors ultimately controls steroid receptor action in a given tissue (Shibata *et al.*, 1997; Jordan and Morrow, 1999). This balance is modified during breast tumorigenesis, and a search for possible means to readjust it has therefore started worldwide with the hope to develop novel therapeutic strategies (McKenna *et al.*, 1999a, 1999b; McKenna and O'Malley, 2001; Gao *et al.*, 2002; Gao and Nawaz, 2002).

In 1999, Lanz *et al.* (1999) identified three human transcripts

Departments of ¹Biochemistry and Medical Genetics, ²Pathology, ³Physiology, and ⁴Manitoba Institute of Cell Biology (MICB), University of Manitoba, Winnipeg, Manitoba, Canada.

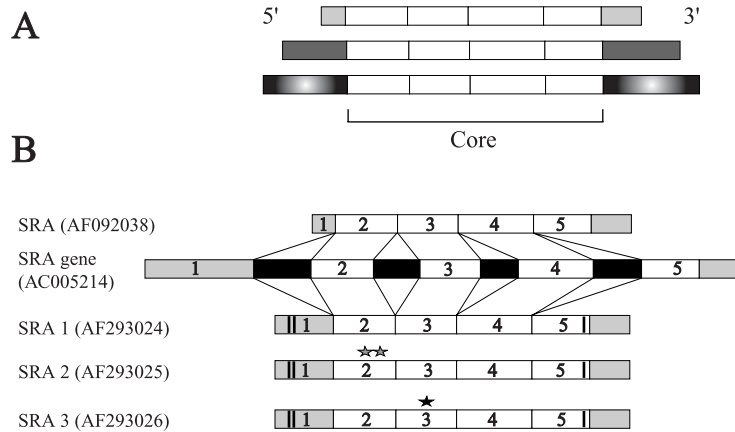


FIG. 1. Schematic representation of SRA-mRNAs and gene structure. **(A)** Three SRA1 cDNAs sharing a central core region but different in their 5' and 3' extremities have been originally characterized (Lanz *et al.*, 1999). **(B)** The original non-coding SRA1 (AF092038) and genomic sequence (AC005214) have been aligned together with the three SRA1 sequences we cloned from normal breast cells. Vertical bars: initiating AUG (in exon 1) and stop codon (in exon 5). Gray stars: point mutations. Black star: point mutation followed by three bases insertion. The core is indicated in white and introns in black.

(Fig. 1A), which shared a central core region, differed in their 5' and 3' terminal extremities, and defined a very peculiar coactivator, the Steroid Receptor RNA Activator (SRA; Lanz *et al.*, 1999). Indeed, these transcripts, unable to code for any detectable protein *in vitro* or *in vivo*, specifically activated steroid receptors as RNA molecules, as opposed to all other previously described coregulators that act as proteins. It has now been shown that SRA1 RNAs modulate steroid receptor activity embedded in a ribonucleo-protein complex containing multiple proteins, such as the corepressor Sharp, the coactivator SRC-1, or the AF-1 specific activator p72/p68 protein (Lanz *et al.*, 1999; Shi *et al.*, 2001; Watanabe *et al.*, 2001; Deblois and Giguere, 2003). The original SRA1 sequence (and the only human one put in Genbank by Lanz *et al.*, #AF092038) is fully contained within five separated exon-like regions of the chromosome 5 genomic sequence (Fig. 1B). The core, which is necessary and sufficient for the noncoding RNAs to potentiate the activity of steroid receptors *in vitro* (Lanz *et al.*, 1999) and to stimulate mammary cell growth *in vivo* (Lanz *et al.*, 2003) encompasses exon-2 to a portion of exon 5. In 2000, we cloned from normal breast tissue three novel SRA1 mRNAs (AF293024, AF29325, and AF293026; Fig. 1B), mainly identical to the original SRA1 sequence but presenting an extended 5' region. This region contained two ATG codons encoding the first methionines of putative 236/224 amino acid open reading frames. These coding variants, which contain an intact core region, are also expected to coactivate steroid receptors. These SRA1 variants as well as their corresponding SRA1 proteins (SRAP) have now been detected in human breast cancer cell lines, human breast tumor tissues, as well as in the skeletal muscle of various species (Leygue *et al.*, 1999; Emberley *et al.*, 2003; Chooniedass-Kothari *et al.*, 2004, 2006). Even though SRAP function remains currently under investigation, our preliminary data have already shown that the expression of coding SRA1 RNA and its corresponding SRAP decreases the activity of estrogen receptor in MCF-7 breast cancer cells (Chooniedass-Kothari *et al.*, 2006). Moreover, we also observed that ER positive breast cancer patients expressing detectable levels of SRAP in their primary tumor were less likely to die from the disease, suggesting that the protein itself could be a new prognostic marker for breast cancer (Chooniedass-Kothari *et al.*, 2006). To our knowledge, SRA1 is the first, and so far only, molecule ever described to act at the RNA level and at

the protein level. No data are therefore currently available on how such bi-faceted system could be regulated. Herein, we have addressed this question.

Using 5'-Rapid Amplification of cDNA Extremities (5'-RACE) and reverse-transcription polymerase chain reaction (RT-PCR) strategies, we investigated possible mechanisms regulating the balance between coding and noncoding SRA1 RNAs in breast cancer cells.

MATERIALS AND METHODS

In silico prediction of transcription initiation sites (TSS)

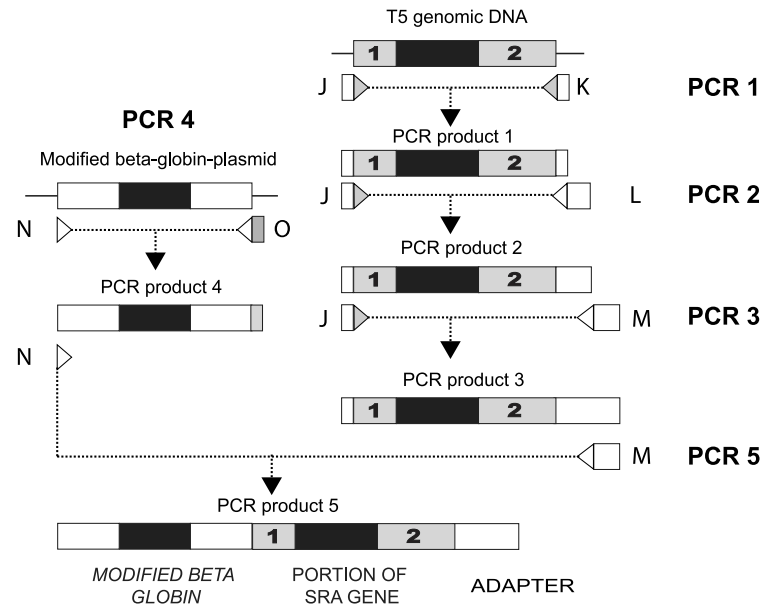
Three online algorithms have been used to identify the putative transcription initiation sites of SRA1 transcripts. The input sequence, taken from the SRA1 genomic sequence (Genbank AC005214), corresponded to 1500 bp upstream of intron-1 + intron-1 + exon-2, totaling 1781 bp. Algorithms included DragonPF (http://research.i2r.a-star.edu.sg/promoter/promoter1_5/DPF.htm, expected TSS sensitivity 80%), McPromoter (<http://genes.mit.edu/McPromoter.html>, cutoff score -0.05) and NNPP2.2 (http://www.fruitfly.org/seq_tools/promoter.html, cutoff score 95%). CpG island prediction was performed using CpGplot program (<http://www.ebi.ac.uk/emboss/cpgplot/>, observed/expected ratio >60 , Percent G+C >30 , length >50).

The database of transcription start sites (DBTSS, <http://dbtss.hgc.jp/>) is based on precise experimentally determined 5'-end clones, that is, sequencing of capped-mRNAs.

Human breast cancer cell lines

BT-20 (cat# HTB-19), MDA-MB-468 (cat# HTB-132), MDA-MB-231 (cat# HTB-26), MCF10A (cat#CRL-10317), ZR-75 (cat# CRL-1500), T47D (cat# HTB-133), and MCF7 (cat# HTB-22) breast cancer cell lines were obtained from the American Type Culture Collection (ATCC, Manassas, VA). T5 breast cancer cell line was kindly provided by Dr. LC. Murphy (Coutts *et al.*, 1996). All the cells were grown up to 70% confluence, harvested, and cell pellets stored at -70°C , as previously described (Emberley *et al.*, 2003). Total RNA and genomic DNA were extracted from cell pellets using Trizol™

FIG. 2. Minigene construction. The insert part of the minigene has been constructed as described in the Materials and Methods section. Briefly, we first amplified (PCR 1) the genomic DNA from T5 cells using an upper primer (J) overlapping the 3' end of the beta-globin construct and the 5' end of the coding SRAP sequence, and a lower primer (K) containing the end of SRA1 exon 2. Two additional PCRs (PCR2 and PCR3), performed using J-L and J-M, respectively allowed the addition of the adapter region. The beta globin part of the construct was amplified (PCR 4) from a previously constructed plasmid (Dominski and Kole, 1991; Xie *et al.*, 2005) using an upper beta globin specific primer (N) and a lower hybrid primer (O) containing a 3'-extremity specific for the end of beta-globin exon and a 5'-extremity corresponding to the beginning of SRAP coding region. The coamplification of the mixed products 3 and 4 allowed the production of a full insert subsequently cloned in the expression vector.



reagent (Gibco BRL, Grand Island, NY) according to the manufacturer's instructions.

5'-RACE

Total RNA (2 μ g) from T5, MCF-7, and MDA-MB-468 was used as starting substrate to perform 5'-RACE using the SMARTTM RACE cDNA amplification kit (Clontech, Mountain View, CA) following the manufacturer's instructions. The SRA1 specific lower primer used consisted in SRAL1 (sequence, Table 1), annealing with the end of the core sequence and previously shown to effectively amplify SRA1 isoforms in breast cancer cells (Emberley *et al.*, 2003).

RT-PCR, triple-primer-PCR (TP-PCR)

One microgram of total RNA from the various cell lines was reverse transcribed in a final volume of 25 μ l using Moloney Murine Leukemia Virus (M-MLV) reverse transcriptase and random hexamers and analyzed as previously described (Emberley *et al.*, 2003). One microliter of reverse-transcription mixture was amplified in a final volume of 30 μ l, in the presence of 60 mM Tris-HCl (pH 8.5), 15 mM $[(\text{NH}_4)_2\text{SO}_4]$, 1.5 mM MgCl_2 , 0.2 mM dATP, 0.2 mM dTTP, 0.2 mM dGTP, 0.2 mM dCTP, 4 ng/ μ l of each primer (pairs or three primers for RT-PCR and TP-PCR, respectively), 1 unit of *Pfu* DNA polymerase (Sratagene, La Jolla, CA) and 10 nM α -³²P dCTP. Each PCR consisted of a 5-min preincubation step at 94°C followed by 30

TABLE 1. SEQUENCE OF PRIMERS USED

Primer	Sequence	Used in
SRAL1	AGTCTGGGGAACCGAGGAT	5' Race, Fig. 5
A	GCTGCCCGCTGGCCAAGCGGA	Fig. 5
B	GCCAAGCGGAAGTGGAGAT	Fig. 5
C	GGAAGTGGAGATGGCGGAGCTG	Fig. 5
D	ATGGCGGAGCTGTACGTGAAGC	Fig. 5
E	TACGTGAAGCCGGGCAACAA	Fig. 5
F	TACGTGAAGCCGGGTGAGCG	Fig. 5
G	GCCAAGCGGAAGTGGAGAT	Fig. 6
H	CCCCAGTATAAGCTAACAGT	Fig. 6
I	GACGTCTCCAATGCCTGTT	Fig. 6
J	GCCACACTGAGTGAGATGACGCGCTGCCCC	Fig. 2
K	GAGGAGAGGGTTGGGGATAGGCTTGCCCTCTGGGGGATCCATCCTGGGGTG	Fig. 2
L	GCCTTGGCGTAGAATCGAGACCGAGGAGGGTTGGGGATAGGCTTGC	Fig. 2
M	CTTTGGCGTAGAATCGAGAC	Fig. 2
N	GTGCACCTGACTCCTGAGGAGAA	Fig. 2
O	GGGGCAGCGGTCATCTCACTCAAGTGTGGC	Fig. 2
P	GTGCACCTGACTCCTGAGGAGAA	Fig. 8
Q	CTCTGGGGATCCATCCTGGGGTG	Fig. 8

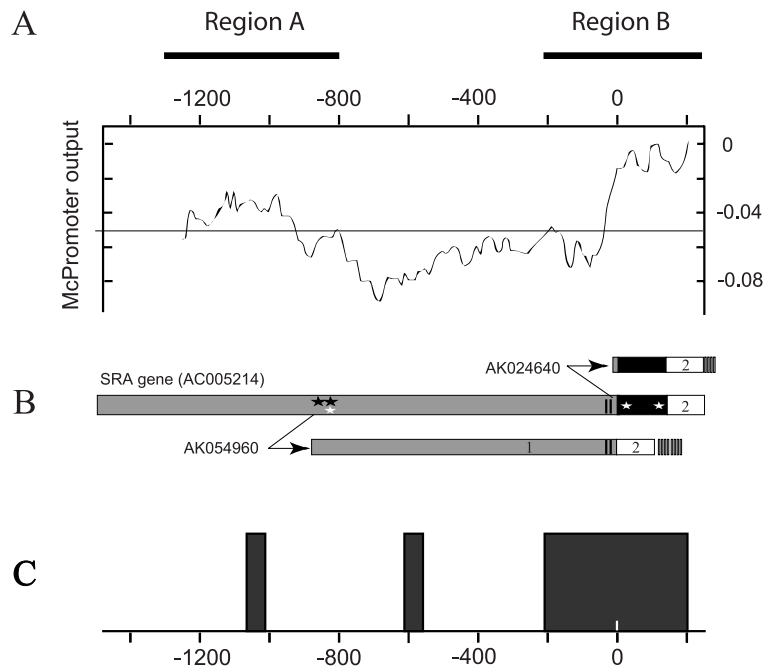


FIG. 3. Prediction of two promoter regions in the SRA1 gene. (A) One thousand seven hundred eighty-one base pairs corresponding to 1500 bp of exon 1, 155 bp of intron-1, and 126 bp of exon-2 were analyzed for putative transcription start sites (TSS) using three programs using cutoffs indicated in the Materials and Methods section. Sequence is numbered using the junction exon-1/intron-1 as 0. (Top panel) McPromoter prediction. The horizontal bar depicts the threshold above which the likelihood of TSS in the region considered is higher. Two promoter regions (A, B) are predicted. (B) The corresponding portion of the SRA1 gene sequence (AC005214) is depicted with exon-1 in gray, intron-1 in black, and exon-2 in white. Positions of the two ATG codons are indicated by vertical bars. Dragon promoter finder and NNPP2.2 predictions for TSS are indicated by white and black stars in the genomic sequence, respectively. The exact true initiation start site of two transcripts (AK024640 and AK054960) is indicated by arrows. (C) The positions of CpG islands identified using the CpG plot program on the same 1781 bp are indicated by dark boxes.

cycles of amplification (30 sec at 94°C, 30 sec at 60°C, and 30 sec at 72°C). The sequences of primers used are detailed in Table 1.

Radiolabeled PCR products were then separated on polyacrylamide gels as previously described (Leygue *et al.*, 1999). Following electrophoresis, the gels were dried and exposed 30 min to a Molecular ImagerTM-FX Imaging screen (Bio-Rad, Hercules, CA). Exposed screen was then scanned using a Molecular ImagerTM-FX (Bio-Rad), which allows the subsequent quantification of each observed signal.

For further sequencing, PCR products were excised from the dry gel, reamplified in the absence of radioactivity, and subcloned using TOPO TA cloning[®] kit according to the manufacturer's instructions (Invitrogen, Carlsbad, CA).

Minigene construction (Fig. 2)

One microgram of genomic DNA from T5 cells was amplified as described above in the absence of radiolabeled nucleotide and using primers J-K (PCR1; Fig. 2 and Table 1). The 384-bp PCR product 1 was separated on ethidium bromide-stained gel, cut out under UV, and purified using the Wizard[®] PCR preps DNA Purification System (Promega, Madison, WI) according to the manufacturer's instructions. This purified PCR product was then diluted (1/1000) and used as substrate for the subsequent PCR2 performed using J-L primers (Fig. 2 and Table 1). Similarly, the resulting PCR product 2 (407 bp) was reamplified using a J-M primer pair to generate the PCR product 3 (456 bp). A modified plasmid (kindly provided by Dr. Xie) containing human beta-globin exon sequences around a modified constitutive intron (Dominski and Kole, 1991; Xie *et al.*, 2005) was then used as a target for amplification with primers N-O to obtain the PCR product 4 (414 bp). The two products, 3 and 4, overlapping in their 5' and 3' region, respectively, were purified, diluted, and mixed together. This mix

was directly used as a template for the final amplification with N-M, allowing the generation of the PCR product 5 (840 bp). This latter product was subsequently purified and subcloned using the pcDNA4/HisMax TOPO[®] TA expression kit (Invitrogen) according to the supplier's instructions. Integrity of the final construction has been checked by sequencing (University Core DNA Services, Calgary, AB, Canada).

Transfection experiments

Transient transfections using the newly synthesized minigene were performed using lipofectamineTM reagent as described by the manufacturer. Twenty-four hours following transfection, cells were lysed and total RNA extracted, reverse transcribed, and amplified as described earlier using primers P-Q (Table 1).

Quantification of signals

TP-PCR signals from four independent experiments were quantified using a Molecular ImagerTM-FX (Bio-Rad) as previously described (Leygue *et al.*, 1996). For each experiment and each cell line, the relative amount of fully spliced and intron-1 retained SRA1 was first expressed as the percentage of the total signal measured following amplification (individual percentage). The total signal corresponds to the sum of the signals of each band observed. The average percentage of intron-1 in T5 cells was then calculated (average T5) from the results of four independent experiments. For each experiment and each cell line (including T5), the individual percentage corresponding to intron 1 was then divided by this T5 average to obtain a normalized value. For each experiment and each cell line (including T5), these normalized percentages were subsequently averaged to obtain the global intron-1 expression (in arbitrary unit). The standard deviations were calculated. The differences between cell lines were tested using the Student's *t*-test (two

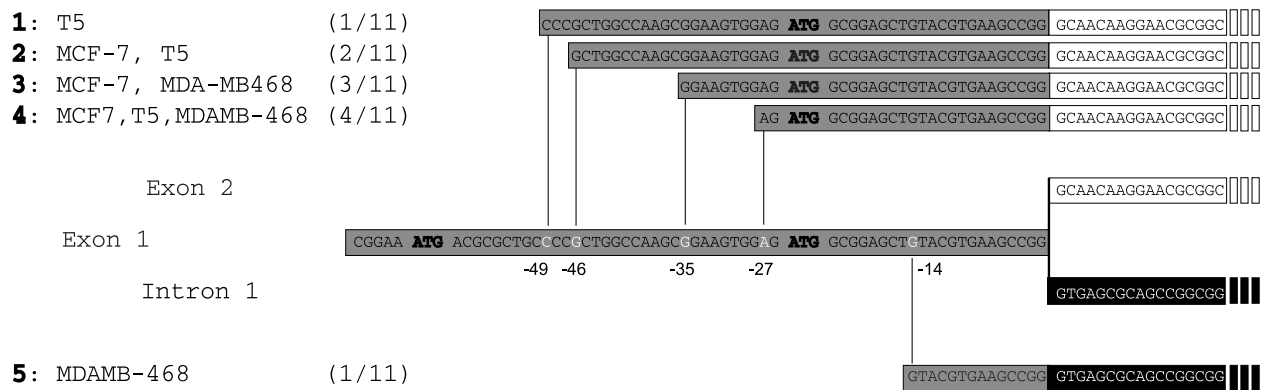


FIG. 4. Identification of putative transcription start sites located within the region B and used in breast cancer cell lines. Total RNA, extracted from T5, MCF-7, and MDA-MB-468 breast cancer cell lines, was analyzed by 5'-RACE as described in the Materials and Methods section. Eleven clones have been sequenced and five putative transcription start sites identified (1–5). For each site, the cells where the transcripts came from, the frequency of occurrence (x out of 11), the exact position on exon-1 sequence (corresponding nucleotide in white), as well as the 5' sequence are indicated. Numbers are given relatively to the junction exon-1 (gray box)/intron-1 (black box). Exon-2 sequences are boxed in white. The two initiating ATG codon are bolded.

tails, equal variance assumed). A similar procedure was used to express the signal of each minigene PCR product in MDA-MB-468 versus T5 cells.

RESULTS

Putative and experimentally determined transcription SRA1 RNAs initiation sites

As outlined earlier in the text, the main difference between the original noncoding and the coding SRA1 isoforms laid within an extended exon 1 sequence (Fig. 1). This observation raised the possibility that these transcripts might therefore be initiated at distinct positions within the SRA1 promoter. We used a series of softwares predicting putative transcription initiation sites (Fig. 3) with cutoffs detailed in the Materials and Methods section. We analyzed 1781 bp covering exon-1, intron-1, and exon-2 (Fig. 3). For position reference, we used the junction between exon-1 and intron-1 as position 0. Two regions (Fig. 3A), region A (–1300/–800) and region B (–200/+200), are predicted by McPromoter to contain putative TSS. Such prediction is corroborated by both NNP2.2 and Dragon-promoter finder, which predict TSS at –839/–872 and –833/+32/+132, respectively (Fig. 3B). Interestingly, these two regions also contain CpG islands, often associated with genomic sequences highly involved in transcription initiation (Fig. 3C).

Only two SRA1 RNAs sequences are currently present in the database of transcription start sites (DBTSS), which contains only experimentally checked integral 5' end cDNA sequences (Fig. 3B). The first sequence (AK054960), isolated from cerebellum, starts at position –883. It therefore contains the two ATG codons (position –61 and –25, respectively) and is able to code for the SRAP. The second (AK024640), cloned from endothelial cells of human coronary artery, is initiated at position –12. Not only it does not contain either of the methionine codons, but it also retains intron-1 sequences (155 bp long),

which introduces a shift in the SRAP reading frame. This RNA is therefore unable to lead to the production of SRAP.

5'-RACE identification of transcription initiation sites in three breast cancer cell lines

To interrogate the 5'-extremity of SRA1 transcripts present in breast cancer cells, total RNA was extracted from three breast cancer cell lines (MCF-7, T5, and MDA-MB-468) and analyzed by 5'-RACE as described in the Materials and Methods section. Eleven clones have been sequenced (three, three, and five clones from T5, MDA-MB-468, and MCF-7, respectively) and five possible initiation sites identified, clustered in a region located at the 3' end of exon-1 (Fig. 4). Four of these sites (1–4, positions –49, –46, –35, and –27, respectively), corresponding to fully spliced RNAs (exon 1/exon-2 junction seen), are located upstream of the second ATG codon, suggesting that transcripts initiated in the region A can also be coding and fully spliced. Interestingly, one clone identified in MDA-MB-468, and starting downstream of the second ATG codon (site 5, position –14) corresponded to a messenger containing a full retention of intron 1, as seen for AK024640. That observation was of particular interest, as the retention of intron-1 makes such transcript unable to code for the SRAP but keeps it functional at the RNA level, as the core sequence (exon-2 to exon-5) is preserved.

Coding/noncoding SRA1 transcripts population in two breast cancer cell lines

To characterize the population of SRA1 transcripts detectable in breast cancer cells, a series of radiolabeled RT-PCR was performed using different upstream primers and a unique downstream primer recognizing the end of the core sequence. As shown in Figure 5A, the upper primers (A–F, position –54, –42, –35, –25, and –14) spanned the region surrounding the second ATG codon and are shown in the 5'-RACE experiments to overlap several TSS. Primer E was overlapping the exon-1/exon-2 junction, whereas primer F spanned the exon-1/intron-

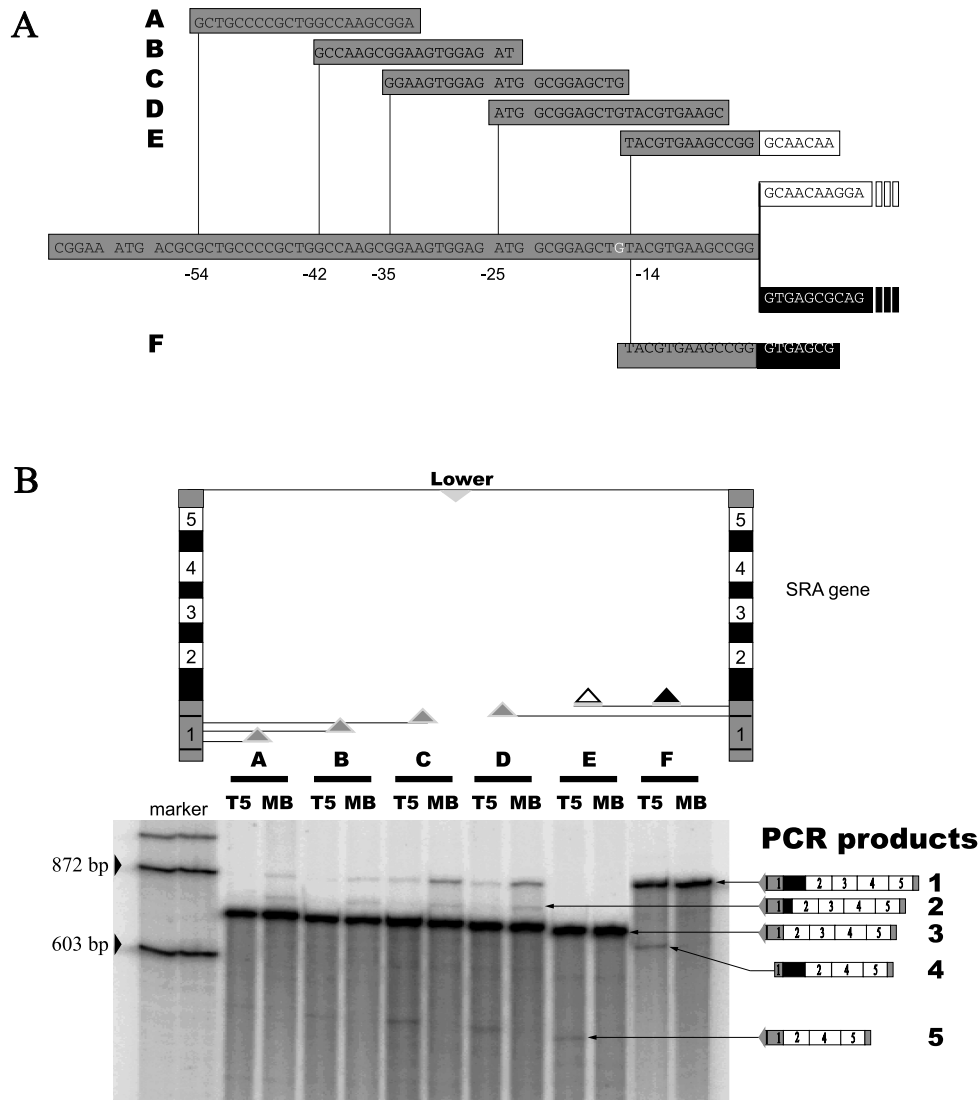


FIG. 5. Identification of SRA1 transcripts in T5 and MDA-MB-468 breast cancer cell lines. Total RNA, extracted from T5 and MDA-MB-468 (MB) breast cancer cell lines, was analyzed by RT-PCR and radiolabeled PCR products separated on acrylamide gels as described in the Materials and Methods section. (A) The sequences and exact positions of the upstream primers used are shown. Numbers are given relatively to the junction exon-1 (gray box)/intron-1 (black box). Exon-2 sequence is boxed in white. The two initiating ATG codon are bolded. The white G corresponds to the position of the transcription start site 5 corresponding to the intron-1 retained noncoding SRA1 RNA identified by 5'-RACE (Fig. 4). Note that primers E and F overlap the junction exon-1/exon-2 and exon-1/intron-1, respectively. (B) The upper panel depicts the position of the primers used. The lower panel shows the radiography of the gel. PCR products have been sequenced and are schematically depicted on the right.

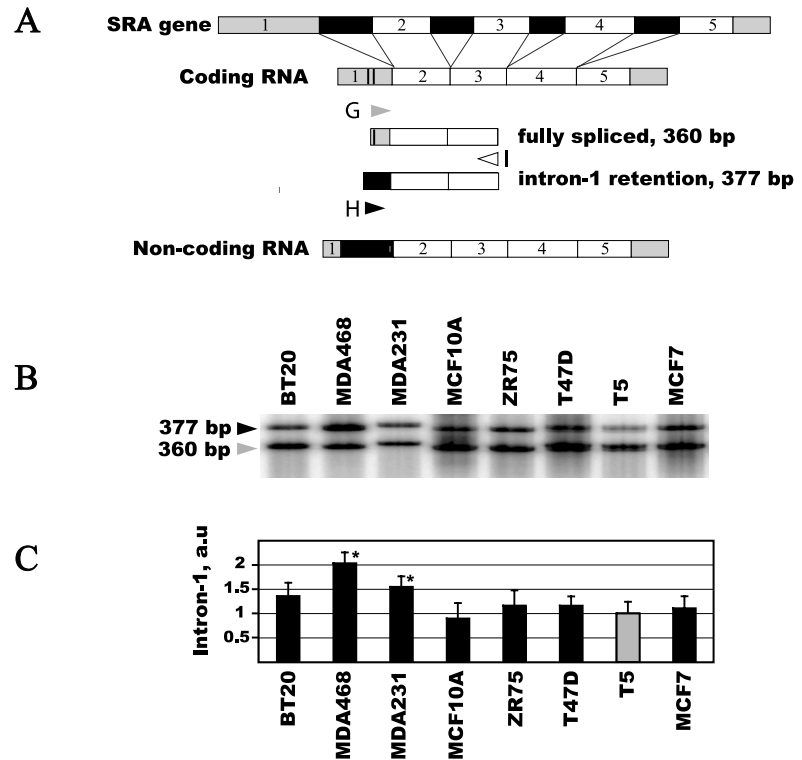
1 boundary. Total RNA from two different cell lines (T5 and MDA-MB-468) was extracted and reverse-transcribed before performing PCR as described in the Materials and Methods section (Fig. 5B). The PCR products (1-5) indicated by arrows have been cloned and sequenced, and the corresponding RNA composition depicted at the bottom right of Figure 4B. Beside the expected fully spliced and coding SRA1 RNA (PCR product 3), four additional species were reproducibly observed in the cells analyzed.

The top PCR product (PCR product 1) corresponds to a full intron-1 retention and appears, relative to the normally spliced transcript (PCR product 3), more highly expressed in

MDA-MB-468 than in T5 cells. It should be stressed that this transcript is still detectable when using primers A-C, which are all upstream of the original site 5 (-14, Fig. 4) shown to generate such a transcript.

An additional transcript (PCR product 2) containing only the 3' part of intron-1 (60 last nucleotides, position +96 to +155), but otherwise normally spliced (all other introns removed), is also detectable in MDA-MB-468 cells and at a lower level in T5 cells. This transcript could act as coactivator as it contains a full core, but does not encode a SRAP. Indeed, the 60 inserted bases bring a premature stop codon to the SRAP open reading frame initiated at the ATG codons. As the fully intron-1 in-

FIG. 6. TP-PCR amplification of coding and noncoding SRA1 mRNAs in breast cancer cells lines. **(A)** Principle: three primers are used during the PCR amplification. The lower primer (white, I) is common to the two sequences, whereas the upper G (gray) and H (black) primers are specific for exon-1 (gray) and intron-1 (black), respectively. Following PCR, the relative proportion of the signals obtained is proportional (but different) to the original relative amounts of the starting cDNAs (Leygue *et al.*, 1996). **(B)** Total RNA was extracted from BT-20, MDA-MB-468 (MBA468), MDA-MB-231 (MBA231), MCF-10A, ZR-75, T47D, T5, and MCF-7 cell lines, reverse-transcribed, and PCR amplified in the presence of P^{32} -dCTP using the three primers as described in the Materials and Methods section. PCR products were separated on acrylamide gel and visualized using a Molecular ImagerTM-FX. Bands migrating at 377 base pairs and 360 bp were sequenced and correspond to RNA containing intron-1 and fully spliced, respectively. **(C)** Signals from four experiments have been quantified as detailed in the Materials and Methods section. The average relative proportion of noncoding SRA1 RNA (intron-1 retained), expressed in an arbitrary unit, is graphed for each cell lines. Differences in relative intron-1 retained RNA expression between T5 (gray bar) and other cell lines (black bars) have been tested using the Student's *t*-test. Stars indicate a significant difference ($P < 0.05$). Standard deviations are shown.



serted transcript, this partial intron-1 inserted variant can also be detected using A–C primers. This simple observation suggests that SRA1 transcripts initiated upstream of the second ATG are alternatively spliced. This has been confirmed using other sets of primers (data not shown), and we have now submitted to Genbank a partial intron-1 SRA1 containing sequence starting at –250 (DQ286291). Interestingly, transcripts containing intron sequences (PCR product 4) or not (PCR product 5) and deleted in exon 3 were also identified. These transcripts are more highly expressed in T5 cells than in MDA-MB-468. They are noncoding, but are also not expected to act at the RNA level as they are missing a significant part of the core (202 nucleotides).

Altogether, these data confirmed the existence, within breast cancer cells, of alternatively spliced SRA1 transcripts, expressed at different levels between the two cell lines studied. The coexistence of transcripts fully spliced (not coding for SRAP) (PCR product 3) and of transcripts containing intron-1 sequences (not coding for the SRAP) (PCR products 1 and 2) is of particular interest. Indeed, it suggests that alternative splicing of this intron might be a possible mechanism used by breast cancer cells to regulate the balance between activator and repressor of estrogen receptor activity. To test for possible differences between cell lines in their relative amounts of coding (intron-1 normally spliced) and noncoding (total or partial intron-1 retained) mRNAs, we used a previously developed TP-PCR approach (Leygue *et al.*, 1996).

Differential relative splicing of SRA1 transcripts containing intron-1 sequence between breast cancer cell lines

As illustrated Figure 6A, the TP-PCR amplifies two PCR fragments corresponding to cDNAs sharing a common sequence (here exon 3) but different in their extremities (exon-1 versus intron-1 sequences). We have in the past shown that the proportion of obtained products was directly proportional to the ratio of the original cDNAs (Leygue *et al.*, 1996). The primer G spanned positions –42 to –24, and recognizes a fragment containing at least the second ATG, whereas primer H (+96 to +115) binds to both partial and full intron-1 retained RNA. Total RNA was extracted from a panel of breast cancer cell lines, reverse-transcribed and analyzed by TP-PCR as described in the Materials and Methods section. These cells were chosen to span a wide range of steroid receptors levels (e.g., MCF-7, ZR75, and T5 have a high level of estrogen receptor, T47D has high levels of progesterone receptor, and MDA-MB-231 is estrogen and progesterone receptor negative cells) and of invasion properties (MDA-MB-231 and MDA-MB-468 are highly invasive cells, whereas MCF-10A cells, coming from fibrocystic lesions, are often considered as “normal” (Miller *et al.*, 1993). PCR products, migrating around the expected sizes of 377 and 360 bases, were detectable in all cell lines tested. Sequencing of the products confirmed their identity as fragments corresponding to noncoding intron-1 and coding fully spliced SRA1 RNAs, respectively (data not shown). It should be stressed that a shift in size was observed

457 bp (Fig. 8A). The sequencing of these fragments revealed that they corresponded to a RNA containing a full SRA1 intron-1 and to a fully spliced RNA, respectively (Fig. 8B). Two other products migrating at 517 bp (PIR) and 443 bp (AD), respectively, are also observable. PIR corresponded to the partial (60 last base pairs) intron-1 previously observed. Interestingly, SRA1 AD product is fully spliced, but is missing 14 bases at the end of exon-1, suggesting the possible use of a cryptic splicing site (Fig. 8B). It should be stressed that this splicing event introduces a premature stop codon in the SRAP reading frame, making the corresponding SRA1 transcript unable to encode the SRAP.

Following quantification, a significant higher relative expression of the SRA1 intron-1 retained mini-gene transcript product was observed in MDA-MB-468, when compared to T5 cells (Fig. 8C). As a direct result, the relative proportion of fully spliced transcript was significantly lower in MDA-MB-468 cells. No difference in the relative expression of the other products, PIR and AD, was observed.

DISCUSSION

The main difference between the originally isolated non-coding SRA1 RNA (Lanz *et al.*, 1999) and the coding SRA1 isoforms we discovered, consisted in an extended exon 1 sequence (Emberley *et al.*, 2003). We therefore first investigated the possibility that these coding and noncoding RNAs might originate from transcriptions initiated at different positions within the SRA1 gene promoter. Predictive algorithms indeed suggested the presence, upstream of the junction between exon-1/intron-1, of two regions (A and B) likely to house a series of transcription initiation sites. The cloning by others of two transcripts, fully capped and corresponding to these two regions, corroborated this assumption. The observation that the transcript initiated in the region A was coding, whereas the one originating from the region B was noncoding, also suggested that the balance between coding and noncoding SRA1 RNA could be controlled by an initiation of transcription in one region versus another. The fact that these transcripts originally came from two different tissues (cerebellum and artery) also raised the possibility that the alternative use of a given site might be tissue specific.

Using 5'-RACE, we herein identified five putative transcription initiation sites, clustered in the region B, located around the second ATG codon, and leading to the production of fully spliced coding SRA1 RNAs (sites 1–4) or noncoding intron-1 retained transcript (site 5). Such “sliding” of the exact initiation site is a typical feature of TATA-less promoter (Smale, 1997). It is here worthy to note that, whereas we obtained coding SRA1 in 10 out of the 11 clones we sequenced, Lanz *et al.* (2000) reported that they did not obtain any sequences containing initiating ATG codons when performing 5'-RACE experiments. The simplest explanation for such a discrepancy possibly lays within the fact that we used breast cancer cells, whereas the tissue origin of cells used by these authors is not specified.

Even though we have not used capped RNAs as a template to perform the 5'-RACE experiments, we are confident that the sites identified are true initiation start sites. Indeed, the fact that we obtained exactly the same sites in different cell lines (sites

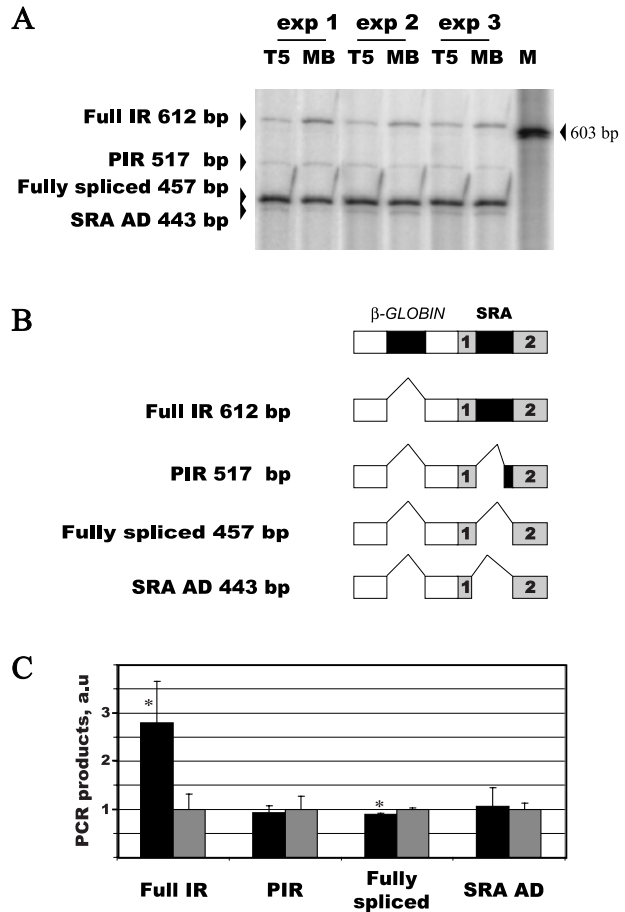


FIG. 8. Alternative splicing of minigene products in T5 and MDA-MB-468 cells. (A) Twenty-four hours following minigene transfection in T5 and MDA-MB-468 (MB) cells, total RNA was extracted, DNase treated, reverse-transcribed and PCR amplified as described in the Materials and Methods section, using primers P and Q (Table 1). PCR products were separated on acrylamide gel and visualized using a Molecular Imager™-FX. The results of three independent experiments (exp. 1–3) are presented. Four bands, migrating at 612, 517, 457, and 443 bp were cloned and sequenced. (B) These bands corresponded to full SRA1 intron-1 retention (full IR), partial SRA1 intron-1 retention (PIR), fully spliced minigene transcript and a new SRA1 alternative SRA1 donor site (SRA1 AD) transcript, respectively. M: molecular weight marker, Φ ×174 RF DNA/Hae III fragments. (C) For each band (full IR, PIR, fully spliced and SRA1 AD), signals have been quantified and expressed in arbitrary unit, as described in the Materials and Methods section. Differences in relative expression between T5 (gray bars) and MDA-MB-468 (black bars) have been tested using the Student's *t*-test. Stars indicate a significant difference ($P < 0.05$). Standard deviations are shown.

2–3–4, Fig. 3) reduces the possibility that we sequenced RNAs degraded exactly at the same position. Moreover, the site identified for the intron-1 retained RNA (site 5, position –12) is extremely close to the one reported for AK024640 (–14) in endothelial cells of human coronary artery. Altogether, these observations suggest that transcripts initiated in the region B in breast cancer cells can be either coding or noncoding.

The SRA1 RNAs we previously isolated from breast cancer cells (Genbank # AF293024-26) were amplified by RT-PCR using a primer at position -250 (Emberley *et al.*, 2003). Their transcription was therefore initiated upstream of region B we identified by 5'-RACE. Using the same upper primer, we were able to amplify SRA1 coding transcripts in all breast cancer cell lines we tested, including T5, MCF-7, and MDA-MB-468. This underlines the fact that transcripts likely initiated in the A region are also present in these cells (Emberley *et al.*, 2003). We believe that we did not detect such transcripts in our 5'-RACE as a direct result of a preferential amplification of shorter cDNAs in the PCR reaction. The observation that intron-1 retaining transcripts can be amplified with the -250 primer (DQ286291) also confirms that the transcript initiated in region A can be alternatively spliced, and overall noncoding.

Altogether, our data show that in breast cancer cells, SRA1 transcripts can be initiated within two distinct regions (A and B), and that they can be alternatively spliced, independently of the site used. These data therefore strongly suggest that the generation of coding and noncoding SRA1 RNAs in breast cancer cells does not result from the initiation of the SRA1 transcripts at a particular site, but rather results from alternative splicing events involving intron-1 sequences.

We showed here that, indeed, coexpression of coding fully spliced and noncoding retaining intron-1 SRA1 RNAs is a common feature of breast cancer cell lines. We also demonstrated that cell lines differ in their relative levels of both RNA species. The observation that an artificial minigene generated RNA, containing exon-1, -intron-1, and exon-2 sequences is differentially spliced in these cell lines, and that the relative trend observed with endogenous SRAs is also seen with the artificial minigene product is of particular interest. Indeed, this suggests that the primary sequence of the transcript in this region, that is, exon-1/intron-1 and exon-2, is sufficient for the cell type specific differential intron-1 retention to occur.

The use by cells of alternative splicing of an intron to control the balance between coding and noncoding RNAs is certainly not a unique feature of the SRA1 gene. Indeed, intron retention is now considered as a relatively common event involving almost 15% of human gene products (Galante *et al.*, 2004).

Alternative splicing has been shown to control major normal events such as sex determination or tuning of brain receptor sensitivity. Moreover, alteration of splicing has also been involved in many pathological situations including but not limited to skin diseases (Wessagowit *et al.*, 2005) neurodegeneration (Lee and Irizarry, 2003; Gallo *et al.*, 2005), and cancer (Hall and Russell, 2005; Kalnina *et al.*, 2005).

In some cases, alternative splicing leads to the production of new sequences (Stetefeld and Ruegg, 2005), which can encode for extended or truncated proteins with novel or altered functional properties (Stamm *et al.*, 2005). In other cases, alternative splicing results in the introduction of an early stop codon or a shift in the reading frame, leading to the generation of noncoding RNAs. In this latter situation, the noncoding RNA is often seen as a passive byproduct of an event, the major effect of which is usually thought to consist in the absence of protein. Since protein production is considered to be the functional outcome of most genes, the production of noncoding RNAs, as a result of differential splicing, is thought in this case to be a

mechanism for gene silencing. However, the bivalent nature of SRA1, that is, functional RNA as well as functional protein, sheds a new light on alternative splicing. Indeed, this mechanism can now also be seen as a way to control the balance between two functional units, RNA as well as its coding counterpart.

Regulation of alternative splicing is a complex phenomenon that involves a plethora of different factors and signal sequences (Staley and Guthrie, 1998; Shin and Manley, 2004). It is now admitted that the full understanding of the molecular mechanisms involved in alternative splicing events could lead to the development of new therapies, targeting the production of particular spliced forms (Garcia-Blanco, 2005; Hagiwara, 2005; Wilton and Fletcher, 2005).

In this context, we are currently investigating the exact mechanisms participating in the differential splicing of intron-1 in breast cancer cells. Indeed, we believe that the characterization of the mechanisms involved will lead to the development of new strategies aiming at modifying the balance between coding and noncoding SRA1 RNAs. Ultimately, this will allow us to control the equilibrium between activator and repressor of estrogen receptor activity in breast cancer.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Jiuyong Xie for his critical reading of the manuscript. This work was supported by grants from the Canadian Institute of Health Research (CIHR), the Canadian Breast Cancer Research Alliance (CBCRA), and the Cancer Care Manitoba Foundation (CCMF). F. Hube and S. Choniedass-Kothari are recipients of a postdoctoral Manitoba Health Research Council (MHRC) Fellowship and a Department of Defense (DOAD) student fellowship, respectively.

REFERENCES

- CHOONIEDASS-KOTHARI, S., EMBERLEY, E., HAMEDANI, M.K., TROUP, S., WANG, X., CZOSNEK, A., HUBE, F., MUTAWE, M., WATSON, P.H., and LEYGUE, E. (2004). The steroid receptor RNA activator is the first functional RNA encoding a protein. *FEBS Lett.* **566**, 43–47.
- CHOONIEDASS-KOTHARI, S., HAMEDANI, M.K., TROUP, S., HUBE, F., and LEYGUE, E. (2006). The steroid receptor RNA activator protein is expressed in breast tumor tissues. *Int. J. Cancer* **118**, 1054–1059.
- COUTTS, A.S., DAVIE, J.R., DOTZLAW, H., and MURPHY, L.C. (1996). Estrogen regulation of nuclear matrix-intermediate filament proteins in human breast cancer. *J. Cell. Biol.* **63**, 174–184.
- DEBLOIS, G., and GIGUERE, V. (2003). Ligand-independent coactivation of ERalpha AF-1 by steroid receptor RNA activator (SRA) via MAPK activation. *J. Steroid Biochem. Mol. Biol.* **85**, 123–131.
- DOMINSKI, Z., and KOLE, R. (1991). Selection of splice sites in pre-mRNAs with short internal exons. *Mol. Cell Biol.* **11**, 6075–6083.
- EMBERLEY, E., HUANG, G.J., HAMEDANI, M.K., CZOSNEK, A., ALI, D., GROLLA, A., LU, B., WATSON, P.H., MURPHY, L.C., and LEYGUE, E. (2003). Identification of new human coding steroid receptor RNA activator isoforms. *Biochem. Biophys. Res. Commun.* **301**, 509–515.
- EVANS, R.M. (1988). The steroid and thyroid hormone receptor superfamily. *Science* **240**, 889–895.

- GALANTE, P.A., SAKABE, N.J., KIRSCHBAUM-SLAGER, N., and DE SOUZA, S.J. (2004). Detection and evaluation of intron retention events in the human transcriptome. *RNA* **10**, 757–765.
- GALLO, J.M., JIN, P., THORNTON, C.A., LIN, H., ROBERTSON, J., D'SOUZA, I., and SCHLAEPFER, W.W. (2005). The role of RNA and RNA processing in neurodegeneration. *J. Neurosci.* **25**, 10372–10375.
- GAO, X., and NAWAZ, Z. (2002). Progesterone receptors—Animal models and cell signaling in breast cancer: Role of steroid receptor coactivators and corepressors of progesterone receptors in breast cancer. *Breast Cancer Res.* **4**, 182–186.
- GAO, X., LOGGIE, B.W., and NAWAZ, Z. (2002). The roles of sex steroid receptor coregulators in cancer. *Mol. Cancer* **1**, 7.
- GARCIA-BLANCO, M.A. (2005). Making antisense of splicing. *Curr. Opin. Mol. Ther.* **7**, 476–482.
- GREEN, S., WALTER, P., GREENE, G., KRUST, A., GOFFIN, C., JENSEN, E., SCRACE, G., WATERFIELD, M., and CHAMBON, P. (1986). Cloning of the human oestrogen receptor cDNA. *J. Steroid Biochem.* **24**, 77–83.
- HAGIWARA, M. (2005). Alternative splicing: A new drug target of the post-genome era. *Biochim. Biophys. Acta* **1754**, 324–331.
- HALL, P.A., and RUSSELL, S.H. (2005). New perspectives on neoplasia and the RNA world. *Hematol. Oncol.* **23**, 49–53.
- JENSEN, E.V., and JORDAN, V.C. (2003). The estrogen receptor: A model for molecular medicine. *Clin. Cancer Res.* **9**, 1980–1989.
- JORDAN, V.C., and MORROW, M. (1999). Tamoxifen, raloxifene, and the prevention of breast cancer. *Endocr. Rev.* **20**, 253–278.
- KALNINA, Z., ZAYAKIN, P., SILINA, K., and LINE, A. (2005). Alterations of pre-mRNA splicing in cancer. *Genes Chromosomes. Cancer* **42**, 342–357.
- LANZ, R.B., MCKENNA, N.J., ONATE, S.A., ALBRECHT, U., WONG, J., TSAI, S.Y., TSAI, M.J., and O'MALLEY, B.W. (1999). A steroid receptor coactivator, SRA, functions as an RNA and is present in an SRC-1 complex. *Cell* **97**, 17–27.
- LANZ, R.B., RAZANI, B., GOLDBERG, A.D., and O'MALLEY, B.W. (2002). Distinct RNA motifs are important for coactivation of steroid hormone receptors by steroid receptor RNA activator (SRA). *Proc. Natl. Acad. Sci. USA* **99**, 16081–16086.
- LANZ, R.B., CHUA, S.S., BARRON, N., SODER, B.M., DEMAYO, F., and O'MALLEY, B.W. (2003). Steroid receptor RNA activator stimulates proliferation as well as apoptosis in vivo. *Mol. Cell Biol.* **23**, 7163–7176.
- LEE, C.J., and IRIZARRY, K. (2003). Alternative splicing in the nervous system: An emerging source of diversity and regulation. *Biol. Psychiatry* **54**, 771–776.
- LEYGUE, E., MURPHY, L., KUTTENN, F., and WATSON, P. (1996). Triple primer polymerase chain reaction. A new way to quantify truncated mRNA expression. *Am. J. Pathol.* **148**, 1097–1103.
- LEYGUE, E., DOTZLAW, H., WATSON, P.H., and MURPHY, L.C. (1999). Expression of the steroid receptor RNA activator in human breast tumors. *Cancer Res.* **59**, 4190–4193.
- MCKENNA, N.J., and O'MALLEY, B.W. (2001). Nuclear receptors, coregulators, ligands, and selective receptor modulators: Making sense of the patchwork quilt. *Ann. N. Y. Acad. Sci.* **949**, 3–5.
- MCKENNA, N.J., LANZ, R.B., and O'MALLEY, B.W. (1999a). Nuclear receptor coregulators: Cellular and molecular biology. *Endocr. Rev.* **20**, 321–344.
- MCKENNA, N.J., XU, J., NAWAZ, Z., TSAI, S.Y., TSAI, M.J., and O'MALLEY, B.W. (1999b). Nuclear receptor coactivators: Multiple enzymes, multiple complexes, multiple functions. *J. Steroid Biochem. Mol. Biol.* **69**, 3–12.
- MILLER, F.R., SOULE, H.D., TAIT, L., PAULEY, R.J., WOLMAN, S.R., DAWSON, P.J., and HEPPNER, G.H. (1993). Xenograft model of progressive human proliferative breast disease. *J. Natl. Cancer Inst.* **85**, 1725–1732.
- MOSELMAN, S., POLMAN, J., and DIJKEMA, R. (1996). ER beta: Identification and characterization of a novel human estrogen receptor. *FEBS Lett.* **392**, 49–53.
- SHI, Y., DOWNES, M., XIE, W., KAO, H.Y., ORDENTLICH, P., TSAI, C.C., HON, M., and EVANS, R.M. (2001). Sharp, an inducible cofactor that integrates nuclear receptor repression and activation. *Genes Dev.* **15**, 1140–1151.
- SHIBATA, H., SPENCER, T.E., ONATE, S.A., JENSTER, G., TSAI, S.Y., TSAI, M.J., and O'MALLEY, B.W. (1997). Role of co-activators and co-repressors in the mechanism of steroid/thyroid receptor action. *Recent Prog. Horm. Res.* **52**, 141–164.
- SHIN, C., and MANLEY, J.L. (2004). Cell signaling and the control of pre-mRNA splicing. *Nat. Rev. Mol. Cell Biol.* **5**, 727–738.
- SMALE, S.T. (1997). Transcription initiation from TATA-less promoters within eukaryotic protein-coding genes. *Biochim. Biophys. Acta* **1351**, 73–88.
- STALEY, J.P., and GUTHRIE, C. (1998). Mechanical devices of the spliceosome: Motors, clocks, springs, and things. *Cell* **92**, 315–326.
- STAMM, S., BEN ARI, S., RAFALSKA, I., TANG, Y., ZHANG, Z., TOIBER, D., THANARAJ, T.A., and SOREQ, H. (2005). Function of alternative splicing. *Gene* **344**, 1–20.
- STETEFELD, J., and RUEGG, M.A. (2005). Structural and functional diversity generated by alternative mRNA splicing. *Trends Biochem. Sci.* **30**, 515–521.
- WATANABE, M., YANAGISAWA, J., KITIGAWA, H., TAKEYAMA, K., OGAWA, S., ARAO, Y., SUZAWA, M., KOBAYASHI, Y., YANO, T., YOSHIKAWA, H., *et al.* (2001). A subfamily of RNA-binding DEAD-box proteins acts as an estrogen receptor alpha coactivator through the N-terminal activation domain (AF-1) with an RNA coactivator, SRA. *EMBO J.* **20**, 1341–1352.
- WESSAGOWIT, V., NALLA, V.K., ROGAN, P.K., and MCGRATH, J.A. (2005). Normal and abnormal mechanisms of gene splicing and relevance to inherited skin diseases. *J. Dermatol. Sci.* **40**, 73–84.
- WILTON, S.D., and FLETCHER, S. (2005). RNA splicing manipulation: Strategies to modify gene expression for a variety of therapeutic outcomes. *Curr. Gene Ther.* **5**, 467–483.
- XIE, J., JAN, C., STOILOV, P., PARK, J., and BLACK, D.L. (2005). A consensus CaMK IV-responsive RNA sequence mediates regulation of alternative exons in neurons. *RNA* **11**, 1825–1834.

Address reprint requests to:

Etienne Leygue, Ph.D.

Department of Biochemistry and Medical Genetics

University of Manitoba

770 Bannatyne Avenue

Winnipeg, Manitoba R3E 0W3

Canada

E-mail: eleygue@cc.umanitoba.ca

Received March 29, 2006; received in revised form April 26, 2006; accepted May 4, 2006.