



HAL
open science

The CD44 alternative v9 exon contains a splicing enhancer responsive to the SR proteins 9G8, ASF/SF2, and SRp20

Delphine Galiana-Arnoux, Fabrice Lejeune, Marie-Claude Gesnel, James Stevenin, Richard Breathnach, Fabienne del Gatto-Konczak

► **To cite this version:**

Delphine Galiana-Arnoux, Fabrice Lejeune, Marie-Claude Gesnel, James Stevenin, Richard Breathnach, et al.. The CD44 alternative v9 exon contains a splicing enhancer responsive to the SR proteins 9G8, ASF/SF2, and SRp20. *Journal of Biological Chemistry*, 2003, 278 (35), pp.32943-32953. 10.1074/jbc.M301090200 . hal-04146336

HAL Id: hal-04146336

<https://hal.science/hal-04146336>

Submitted on 29 Jun 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The CD44 Alternative v9 Exon Contains a Splicing Enhancer Responsive to the SR Proteins 9G8, ASF/SF2, and SRp20*

Received for publication, January 31, 2003, and in revised form, June 23, 2003
Published, JBC Papers in Press, June 24, 2003, DOI 10.1074/jbc.M301090200

Delphine Galiana-Arnoux[‡], Fabrice Lejeune^{§¶}, Marie-Claude Gesnel[‡], James Stevenin[§],
Richard Breathnach^{‡||}, and Fabienne Del Gatto-Konczak[‡]

From the [‡]INSERM U463, Institut de Biologie-CHR, 9 Quai Moncoussu, 44093 Nantes Cedex 1 and [§]Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS/INSERM/IULP, 67404 Illkirch Cedex, France

The *CD44* gene alternative exons v8, v9, and v10 are frequently spliced as a block by epithelial cells. By transfecting minigenes containing only one of these alternative exons, we show that splicing of each of them is under cell type-specific control. By using minigenes carrying short block mutations within exons v8 and v9, we detected a candidate exon splicing enhancer in each of these exons. These candidates activated splicing *in vitro* of a heterologous transcript and are thus true exon splicing enhancers. We analyzed further a v9 exon splicing enhancer covering ~30 nucleotides. This enhancer can be UV cross-linked to SR proteins of 35 and 20 kDa in HeLa nuclear extract. By using individual recombinant SR proteins for UV cross-linking in S100 extract, these proteins were identified as 9G8, ASF/SF2, and SRp20. S100 complementation studies using recombinant 9G8, ASF/SF2, and SRp20 showed that all three proteins can activate splicing *in vitro* of a heterologous exon containing the v9 enhancer; the strongest activation was obtained with 9G8. Progressive truncation of the 30-nucleotide enhancer leads to a progressive decrease in splicing activation. We propose that 9G8, ASF/SF2, SRp20, and possibly other non-SR proteins cooperate *in vivo* to activate v9 exon splicing.

Alternative splicing of pre-mRNA transcripts allows a gene to code for more than one protein (1, 2). Very often similar but distinct versions of the same protein are made from a gene in this way. When alternative splicing is controlled in a cell type-specific manner, cells benefit from the particular version of a protein best suited to their needs. The biological importance of this is underlined by numerous observations linking impaired control of alternative splicing to human diseases, including cancer (3). Understanding splicing control will thus shed light on important aspects of normal and pathological cell life.

In mammalian systems, alternative splicing control is often very complex and is characterized by multiple combinatorial control (4). For example, splicing of the neuron-specific exon N1 of the mouse *c-src* gene is activated by a purine-rich exonic

sequence and an intronic splicing enhancer; the latter contains binding sites for hnRNP¹ H, hnRNP F, and KH-type splicing regulatory protein (5). Polypyrimidine tract binding protein (PTB) binds to intron sites flanking the N1 exon to repress splicing, and one of these lies within the enhancer (6). A neurally enriched homologue of PTB binds more stably to this site than does PTB but is a weaker repressor of splicing *in vitro*. A neurally enriched homologue of PTB also greatly enhances the binding of hnRNP H and KH-type splicing-regulatory protein to the enhancer (5). The N1 exon is thus under control of both activators and repressors, and their relative influence determines its splicing efficiency (7).

We and others have investigated another example of multiple combinatorial control: splicing of the *FGFR-2* gene K-SAM alternative exon, an exon spliced by epithelial cells. Splicing requires the action of three activating sequences in the downstream intron (8), one of which binds the activator TIA-1 (9). Splicing is repressed by an exon-splicing silencer that binds hnRNP A1 (10) and by PTB-binding sites in the flanking introns (11). The other alternative exon (BEK) is repressed by a sequence in the upstream intron that overlaps with one of the sequences needed for efficient K-SAM exon splicing (12). The BEK exon also contains binding sites for PTB that contribute to repression of its splicing (13).

We wished to analyze splicing of other epithelial cell-specific exons. Variable exons v8, v9, and v10 of the *CD44* gene are spliced by epithelial cells and are attractive models for a number of reasons. *CD44* is a transmembrane receptor expressed by numerous cell types, which participates in cell-cell or cell-extracellular matrix interactions (14–17). It has been implicated in a number of normal and pathological processes depending on such interactions. Many different isoforms of *CD44* exist, as a result of extensive alternative splicing of the *CD44* pre-mRNA (18). The human *CD44* gene contains 20 exons. Exons 1–5 are spliced constitutively. Between exons 5 and 16 lie 10 exons (v1–v10) that can be alternatively spliced in a very large number of different combinations. A report (19) that metastasizing rat pancreatic adenocarcinoma and mammary carcinoma cell lines express specific *CD44* isoforms (CD44v4–v7 and CD44v6,v7), whose forced expression suffices to confer metastatic properties upon a related non-metastasizing cell line, provoked many studies of *CD44* transcripts and proteins in a wide variety of human tumor samples (20). Although no simple, universal rule has emerged, expression of certain *CD44* isoforms appears to be inversely correlated with survival probability in some specific cancers (17). Much of this work has focused on exon v6, but some studies have also addressed

* This work was supported in part by INSERM, the CNRS, the Hôpitaux Universitaires de Nantes et de Strasbourg, the Association pour la Recherche sur le Cancer, and the Ligue Nationale Contre le Cancer, Comité Départemental de Loire-Atlantique. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¶ Supported by a fellowship from the Ligue Nationale Contre le Cancer.

|| To whom correspondence should be addressed: INSERM U463, Institut de Biologie-CHR, 9 Quai Moncoussu, 44093 Nantes Cedex 1, France. Tel.: 33 2 40 08 47 50; Fax: 33 2 40 35 66 97; E-mail: breathna@nantes.inserm.fr.

¹ The abbreviations used are: hnRNP, heterogeneous nuclear ribonucleoprotein; ESE, exon splicing enhancers; RT, reverse transcriptase; dsx, doublesex; 5'ss, 5' splice site; SCR, scrambled.

exons v8–v10. Thus it has been suggested, for example, that CD44v8–v10 expression is a marker of prognostic significance in gallbladder cancer (21).

The link between CD44 and cancer makes the CD44 pre-mRNA an attractive model for studying the possible modification of alternative splicing control in cancer. Previous studies have concentrated on elements controlling splicing of CD44 exons v4 (22–24) and v5 (24–28). As stated above, we were interested in exons v8, v9, and v10, splicing of which allows production of the CD44E form, preferentially expressed on epithelial cells. We have begun an analysis of sequence elements important for splicing of these exons, and we show here that exons v8 and v9 contain exon splicing enhancers (ESEs). We identify sequences within a v9 ESE, which bind to SR proteins 9G8, ASF/SF2, and SRp20, and we show that these proteins can activate splicing *in vitro* of a heterologous exon containing the v9 ESE. Our results are thus in favor of involvement of 9G8, ASF/SF2, and SRp20 in v9 exon splicing *in vivo*.

EXPERIMENTAL PROCEDURES

Plasmids—All plasmids used were constructed by standard subcloning procedures (details are available on request). The CD44 minigene has been described elsewhere (9). To make the v8 minigene, exons v9 and v10 were removed from the CD44 minigene by deleting fragments containing them (a 319-bp fragment was deleted for v9, and a 307-bp fragment was deleted for v10). To make the v9 minigene, a 856-bp deletion between *ClaI* and *BlnI* sites in the CD44 minigene was used to remove the v8 exon, and the v10 exon was deleted on a 307-bp fragment. The v10 minigene was made by deleting a 2988-bp fragment containing the v8 and v9 exons from the CD44 minigene. To make v8 and v9 minigenes containing blocks of mutated exon sequences, fragments of the corresponding minigenes were transferred to pBluescript SK(+) (an 856-bp *ClaI*-*BlnI* fragment for v8 and a 2132-bp *BlnI*-*HpaI* fragment for v9), mutated, and replaced in the parent minigene. Mutagenesis was carried out using 5'-phosphorylated divergent oligonucleotide primers for PCR. PCR products were circularized by ligation prior to introduction into *Escherichia coli*. Mutations were verified by sequencing. Oligonucleotides used are as follows: for v8Mb9, 5'-AATGTTCTGAAGGATCAATATTAGT-3' and 5'-CGCGAAGTCATAGTACAACGCTTCA-GCC-3'; for v8Mb8, 5'-CGGAGTGAGGAGTCCATATCTGAAGGAATC-3' and 5'-ATATCGCCTTCAAGCTACTGCAATCC-3'; for v8Mb7, 5'-TATCGTTCGTTGTACTACTGACTGGAGTCC-3' and 5'-TCCGACGCAATCCAAACACAGGTTTGGTGG-3'; for v8Mb6, 5'-ATCCGGTGCAGT-AGGCTGAAGCGTTATAC-3' and 5'-CCGCAAGGTTTGGTGGAAAGAT-TTGGAC-3'; for v8Mb5, 5'-TCAGGAGTGTGGATTGCGAGTAGG-C-3' and 5'-TATCCAAGATTTGGACAGGACAGGACC-3'; for v8Mb4, 5'-TCGAGCCACCAACCTGTGTTGG-3' and 5'-TATCACAGGACAGGACCTCTTC-3'; for v8Mb3, 5'-ATCCGCCAAATCTCCACCAACCTG-3' and 5'-ATCCATCTCTTTCAATGACAACGCG-3'; for v8Mb2, 5'-TCGCCTCTGCTGTCCAAATCTTC-3' and 5'-TATCATGACACGCGTAAGAATAACG-3'; for v8Mb1, 5'-ATCGTGAAGAGGTCCTGTCTGTCTGCC-3' and ATCCGCTAAGAATAACGAGTCTCAG-3'; for v9Mb8, 5'-TATCGCTCTGTTTCAATGAGGAATGAATCC-3' and 5'-TCGCCTCTCAGAGTCTCTACATCAGATG-3'; for v9Mb7, 5'-TATCGTTATTACTCTGCTGTTTCAATGAGG-3' and 5'-TCCGACGCA-CATCACATGAAGGCTTGAAGAAG-3'; for v9Mb6, 5'-ATCATCAGAGGACTCTGAGAATTACTCTG-3' and 5'-ATCCGAGGCTTGAAGAAGATAAAGACC-3'; for v9Mb5, 5'-TCAGAGTATGTGATGTGATGTAGAGAAGCTCTGAG-3' and 5'-TATCCAAGATAAAGACCATCAAC-AACTTC-3'; for v9Mb4, 5'-TCGAGCTTCCAAGCCTTCAATGTGATG-3' and 5'-TATCACCATCCAACAACCTTCTACTC-3'; for v9Mb3, 5'-ATCCGCTTTATCTTCTCCAAGCCTTC-3' and 5'-ATCCATCTACTCTACTC-TGACATCAAGCAG-3'; for v9Mb2, 5'-TCGCATGTTGGATGGTCTTT-ATCTTCTTC-3' and 5'-TATCCTGACATCAAGCAGTAAGG-3'; and for v9Mb1, 5'-TATCGAGTAGAAGTGTGGATGGTCTTTATC-3' and 5'-TCAATGCAGTAAGGATTATAAAACCTAGTTGGC-3'.

pdsx(RI/*FspI*)T7 (29) was a gift from T. Maniatis. pdsx(AAG)₆, pdsx-SCR, pdsx-v8d9, pdsx-v8d4, pdsx-v8b4, pdsx-v9d8, and pdsx-v9b6-5-4 were made by introducing complementary oligonucleotides corresponding to the sequences shown in Fig. 4A into the *Bst*BI site of pdsx(RI/*FspI*)T7. These constructs were linearized by *Hind*III before *in vitro* transcription using T7 RNA polymerase. For UV cross-linking experiments, the same sequences as those introduced into pdsx(RI/*FspI*)T7 (plus v9b5-4, v9b4a, see sequences Fig. 5A) were introduced into pBluescript SK(+). Constructs were linearized with *Eco*RI before *in vitro*

transcription. Plasmid Sp1 ex2 inv was as described previously (30). The v9 exon sequences to be tested (see Fig. 5A for individual sequences) were inserted between the *KpnI* and *Bam*HI sites of the inverted exon of this plasmid to make Sp1 ex2 inv v9b6-5-4, v9b5-4, and v9b4a. The positive control sequences for 9G8, ASF/SF2, and SRp20 are as listed in the legend to Fig. 6; they were also inserted between the *KpnI* and *Bam*HI of Sp1 ex2 inv to make derivatives responding specifically to each SR protein. These constructs were linearized by *Hind*III before *in vitro* transcription.

Transfections and RT-PCR—RNA was prepared from SVK14 cells (31) and 293-EBNA cells (Invitrogen), which had either been transfected or not, and used for RT-PCR with appropriate primers as described elsewhere (10). When applicable, cycloheximide (50 μ g/ml) was added to cells 48 h after transfection, and cells were harvested 6 h later. 24–28 cycles of amplification (1 min at 94 °C, 1 min at 58 °C, and 2.5 min at 72 °C) were used in order to remain in the range of exponential amplification. PCR products were migrated on 2% agarose gels, transferred to nylon filters (Hybond N+, Amersham Biosciences), and hybridized with appropriate ³²P-labeled probes. Filters were subjected to autoradiography or used for PhosphorImager analysis (PhosphorImager:445SI, Amersham Biosciences), with quantification using ImageQuant software. For analysis of endogenous CD44 mRNAs, primers used were 5'-CTCCACCTGAAGAAGATTGTACATC-3' and 5'-TCAGATCCATGAGTGGTATGGGACC-3' for amplification, and hybridization probes were either a mixture of 5' end-labeled oligonucleotides (5'-CAATGCCTTTGATGGACC-3' and 5'-ACTGGGGTGAATGTGTC-3') from constitutive exons 4 and 16, respectively, or 5' end-labeled oligonucleotide probes specific for the v8 exon (5'-GGATTTGCAGTAG-GCTGAAGCG-3'), the v9 exon (5'-CATGTGATGTAGAGAAGCTCTG-3'), or the v10 exon (5'-CTTCGTGTGGGTAATGAGAG-3'). For analysis of minigene mRNAs primers used were P1 (5'-ATCCAGTGGAT-CAAGCAC-3') and P2 (5'-GGCAACCTAGAAGGCACAG-3'), and probes used were either linked C1 and C2 sequences of the *FGFR-2* gene or the CD44 exon-specific probes listed above. For transfections, SVK14 cells and 293-EBNA cells were transfected with 2 μ g of minigene and 18 μ g of pCI-MS2 (10) by the calcium phosphate coprecipitation technique as described elsewhere (10). 48 h after transfection, RNA was harvested and analyzed by RT-PCR.

Extract Preparation and *In Vitro* Splicing Assays—HeLa cell nuclear extract, the 15–45% ammonium sulfate nuclear fraction, a total SR protein preparation, and recombinant 9G8, ASF/SF2, SC35, and SRp20 were prepared as described previously (30, 32). Capped ³²P-labeled pre-mRNA substrates were made by run-off *in vitro* transcription with T7 RNA polymerase for dsx constructs and with SP6 RNA polymerase for Sp1 ex2 inv constructs. For standard *in vitro* splicing assays, 132,000 dpm of dsx or Sp1 ex2 inv transcripts (40 fmol) were incubated in a final volume of 25 μ l, with 10 μ l of nuclear extract for dsx constructs, or with a mixture of 4 μ l of nuclear extract and 4 μ l of S100 cytoplasmic extract for Sp1 ex2 inv transcripts, in 11.2 mM Hepes, pH 7.9, 0.76 mM ATP, 24.8 mM creatine phosphate, 3.2 mM MgCl₂, 2.6% polyvinyl alcohol, 0.28 mM dithiothreitol, 11.2% glycerol, 56 mM KCl, 0.11 mM EDTA, 40 units of RNasin (Promega) at 30 °C for 2 h. Splicing products were resolved on denaturing 10 (dsx-based transcripts) or 6% (Sp1 ex2 inv-based transcripts) polyacrylamide gels, followed by autoradiography. For *in vitro* splicing of Sp1 ex2 inv-based transcripts in S100 complementation assays, transcripts were incubated in a final volume of 25 μ l with 9 μ l of S100 cytoplasmic extract plus 1 μ l of a 15–45% ammonium sulfate nuclear fraction. Under these conditions only very weak splicing is observed even with good splicing substrates, and this weak splicing can most probably be attributed to residual amounts of SR proteins in the 15–45% ammonium sulfate fraction. Depending on the experiment, various amounts of recombinant SR proteins were added to assays as indicated in the figure legends. Incubations were in 11.2 mM Hepes, pH 7.9, 1.52 mM ATP, 24.8 mM creatine phosphate, 3.2 mM MgCl₂, 2.6% polyvinyl alcohol, 0.28 mM dithiothreitol, 6.4% glycerol, 56 mM KCl, 0.11 mM EDTA, 40 units of RNasin at 30 °C for 90 min, and splicing products were resolved on denaturing 6% polyacrylamide gels, before analysis by autoradiography.

UV Cross-linking Assays—High specific activity ³²P-labeled RNA probes were made by run-off *in vitro* transcription with T7 RNA polymerase from the pBluescript SK(+)-based plasmids (³²P-labeled ATP was used to make the probes). Probes (250,000 dpm) were incubated in a final volume of 15 μ l with 3 μ l of nuclear extract, or 3 μ l of S100 cytoplasmic extract, or 3 μ l of S100 cytoplasmic extract plus, depending on the experiment, 200 ng of total SR proteins, or 200 ng of individual recombinant SR proteins. Incubations were in 10 mM Hepes, pH 7.9, 0.75 mM ATP, 25 mM creatine phosphate, 3.2 mM MgCl₂, 11.3% glycerol, 60 mM KCl, 33 ng/ μ l tRNA, and 55 units of RNasin at 30 °C for 20 min.

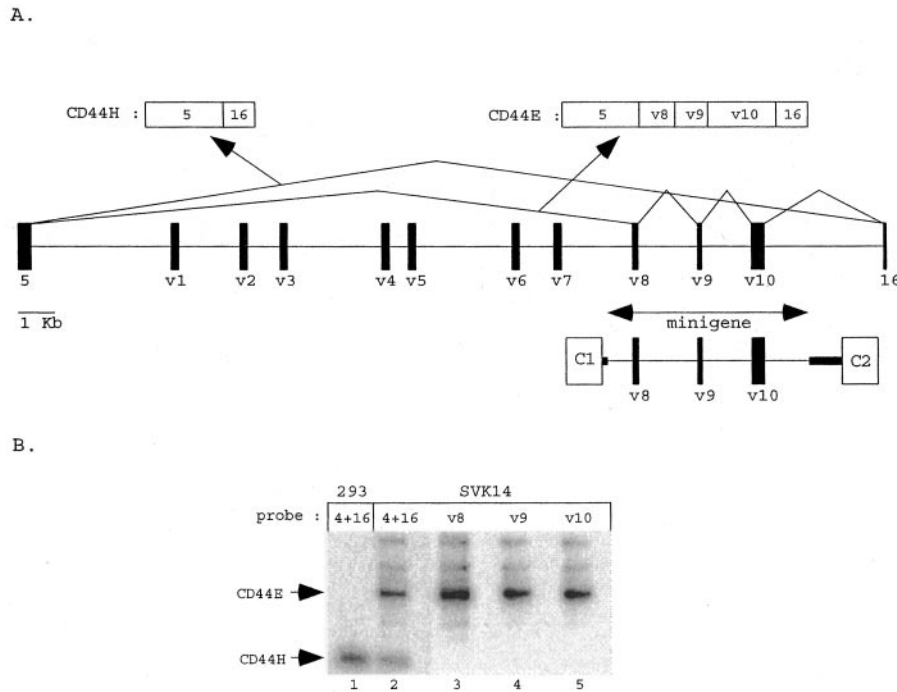


FIG. 1. Splicing of CD44 variable exons. *A*, a portion of the CD44 pre-mRNA is shown, with exons marked as *black boxes* and introns as *lines*. The alternative splicing events yielding CD44H (splicing of exon 5 to exon 16, skipping of all v exons) and CD44E (splicing of exons v8, v9, and v10 between exons 5 and 16) are shown *above* the pre-mRNA. The CD44 sequences present in minigene pre-mRNA are identified by an *arrow below* the CD44 pre-mRNA. Also shown is a portion of minigene pre-mRNA, with the FGFR-2 constitutive exons C1 and C2 represented as *boxes* and FGFR-2 intron sequences as *thick lines*. *B*, RT-PCR analysis of CD44 pre-mRNA splicing in 293 cells and in SVK14 cells. RNA from these cells was subjected to RT-PCR using primers from exons 4 and 16, before Southern blotting analysis. For *lanes 1* and *2*, the probe used was a mixture of oligonucleotides from exons 4 and 16. For *lanes 3–5*, oligonucleotide probes specific for exons v8, v9, and v10, respectively, were used. The expected migration positions of CD44E and CD44H products are shown.

Incubated samples were subjected to UV cross-linking, and then the samples were treated with 250 ng of RNase A and 100 units of RNase T1 for 1 h at 37 °C. Samples were diluted with a 2× SDS protein loading buffer, heated for 5 min at 100 °C, and resolved by SDS-PAGE on 12% polyacrylamide gels. Dried gels were then autoradiographed.

RESULTS

A Minigene for Studying Exon v8–v10 Splicing—A schematic map of part of the *CD44* gene with its 10 variable exons (v1–v10) lying between constitutive exons 5 and 16 is shown in Fig. 1A, together with representations of the CD44E mRNA (containing exons v8–v10) and the CD44H mRNA (containing none of the variable exons). We analyzed the CD44 splicing pattern in human 293 cells and in the human keratinocyte cell line SVK14, using RT-PCR with primers in constitutively spliced exons 4 and 16. The major RT-PCR product from 293 cells corresponded in size to that expected from CD44H mRNA (Fig. 1B, *lane 1*) and did not hybridize to oligonucleotide probes specific for the v8, v9, and v10 exons (data not shown). Whereas SVK14 cells express some CD44H mRNA also, the major RT-PCR product obtained from these cells corresponded in size to that expected from CD44E mRNA (Fig. 1B, *lane 2*) and hybridized to oligonucleotide probes specific for the v8, v9, and v10 exons (Fig. 1B, *lanes 3–5*, respectively). Some RT-PCR products larger than CD44E were observed in SVK14 cells; they contain exons v8–v10 (Fig. 1B, *lanes 3–5*) and presumably a selection of additional v exons chosen from v1–v7.

We constructed a minigene (v8v9v10) harboring human *CD44* gene sequences including variable exons v8, v9, and v10 (see Fig. 1A), with 0.74 kb of the 18.4-kb intron separating exons 5 and v8, and 1.5 kb of the 3.7-kb intron separating exons v10 and v16. In our minigene these CD44 sequences lie between constitutively spliced *FGFR-2* gene exons C1 and C2. A Rous sarcoma virus long terminal repeat promoter and the

polyadenylation signal for a bovine growth hormone gene are provided for driving expression (Fig. 2A). This minigene was transfected into SVK14 and 293 cells, and RNA was harvested and analyzed by RT-PCR using minigene-specific primers P1 and P2. In SVK14 cells, the major RT-PCR product obtained (Fig. 2B, *lane 1*) contained exons v8, v9, and v10 (this product hybridized to oligonucleotide probes specific for these exons, data not shown).

In 293 cells, the RT-PCR products were more complex (Fig. 2B, *lane 5*). The major product (C1C2) corresponded to skipping of all three variable exons, whereas other products corresponded to splicing of different combinations of the variable exons as marked (products were identified by hybridization to oligonucleotide probes specific for the v8, v9, and v10 exons, data not shown). Note that in 293 cells, an alternative 5' splice site (5'ss 2, Fig. 2A) can be used for exon v9 on minigene pre-mRNA. When 5'ss 2 is used, additional sequences (represented by an *asterisk* in Fig. 2, A and B, and in following figures) will be present in mRNAs and RT-PCR products. These additional sequences contain in-frame stop codons. 5'ss 2 is not used for splicing of the endogenous CD44 pre-mRNA in 293 cells (Fig. 1B, *lane 1*). However, in the mouse (33), splicing to exon v9 of an alternative exon lying in the intron separating exons v9 and v10 leads to formation of soluble CD44 (the alternative exon contains in-frame stop codons). It is thus possible that 5'ss 2 is used by some human cell types for soluble CD44 production.

Compared with endogenous CD44 pre-mRNA, minigene pre-mRNA allows more efficient splicing of exons v8, v9, and v10 in both SVK14 cells and in 293 cells (compare Fig. 1B, *lane 2*, to Fig. 2B, *lane 1*, and Fig. 1B, *lane 1*, to Fig. 2B, *lane 5*). This is probably because, as shown in Fig. 1A, the 3' splice sites of these exons are much closer (at least 17.6 kb closer) to the 5'

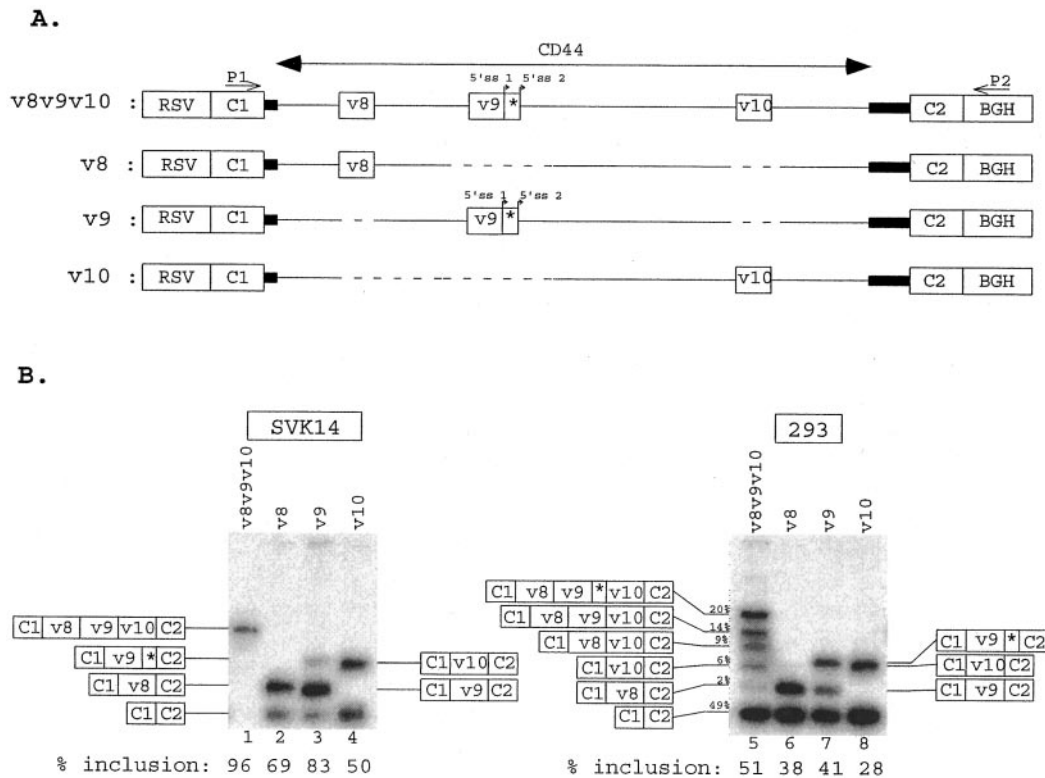


FIG. 2. Splicing of CD44 minigene pre-mRNAs. *A*, schematic representations of minigenes v8v9v10, v8, v9, and v10. Drawings are not to scale. The extent of CD44 gene sequence in minigenes is shown by an arrow. The Rous sarcoma virus long terminal repeat promoter and the bovine growth hormone polyadenylation sequence are marked RSV and BGH, respectively. FGFR-2 exons C1 and C2 are identified, and FGFR-2 gene intron sequences are represented by thick lines. CD44 exons v8, v9, and v10 are shown, and thin lines represent CD44 intron sequences. Two alternative 5' splice sites are marked for the v9 exon (5'ss 1, the normal v9 5' splice site, and 5'ss 2). The box with an asterisk represents sequences that are included in mRNAs if the v9 exon is spliced using 5'ss 2. Broken lines in v8, v9, and v10 minigene representations indicate deleted sequences. The positions of primers P1 and P2 used for RT-PCR analysis are marked. *B*, RT-PCR analysis of minigene pre-mRNA splicing. RNA from 293 or SVK14 cells transfected with minigenes as shown was subjected to RT-PCR using primers P1 and P2 shown in *A*, before Southern blot analysis with a probe comprising linked C1 and C2 exon sequences. Filters were subjected to autoradiography or used for PhosphorImager analysis to determine the relative abundance of RT-PCR products. Individual products are identified by their structures. The box with an asterisk represents sequences included as a result of use of the v9 exon's alternative 5'ss (5'ss 2, see *A*). The % inclusion was calculated by dividing the signal furnished by products containing one or more variable exon (for the v9 exon, products derived from use of both 5'ss 1 and 5'ss 2 were taken into consideration) by the total signal and multiplying by 100%. For lane 5, individual products were identified by hybridization to oligonucleotide probes in a separate set of experiments not shown here. The relative contributions of the products to the total signal is shown to the immediate left of lane 5. Transfections were carried out by series, each series involving transfection of the four minigenes in SVK14 and 293 cells in parallel. Three series were analyzed. Typical results are shown here.

splice site of the upstream constitutive exon C1 in the minigene than they are to the 5' splice site of the upstream constitutive exon 5 in the endogenous gene. Reducing intron size has been shown previously to increase v exon splicing (34). Some cell type-specific splicing control can nevertheless be observed with minigene RNA, as inclusion of variable exons is ~2-fold more efficient in SVK14 cells than in 293 cells (compare Fig. 2B, lanes 1 and 5).

Splicing of Each v Exon Is Controlled—Exons v8, v9, and v10 are very often spliced as a block, and it has been proposed previously that splicing of one of the exons favors splicing of the others (34). As noted above, one possible explanation could be intron size (34), which has been shown to have an influence on v exon splicing. For example, if exon v8 is defined for splicing (as in the exon definition model), the intron separating the 3' splice site of exon v9 from the nearest upstream active 5' splice site (the 5'ss of the v8 exon) will be only 2 kb, as opposed to 20 kb (the distance to the 5' splice site of constitutive exon 5) if exon v8 is not defined. Definition of exon v8 could thus enhance splicing of exon v9. So one possible model for cell type-specific splicing requires splicing control to act directly on just one v exon, splicing of which then induces splicing of the other v exons, without there being a need for any direct control acting on these latter exons. In an alternative model, splicing of each

v exon is under direct control. To distinguish between these two possibilities, we tested minigenes containing only one variable exon (Fig. 2A, minigenes v8, v9, and v10). For each of these minigenes, variable exon inclusion was about 2-fold more efficient in SVK14 cells than in 293 cells (Fig. 2B, compare lanes 2 and 6 for v8, lanes 3 and 7 for v9, and lanes 4 and 8 for v10). This result is similar to that obtained with the minigene containing all three variable exons where, as mentioned above, inclusion of the variable exons is ~2-fold more efficient in SVK14 cells than in 293 cells. Each exon is thus under direct control. However, exons v8, v9, and v10 are spliced less efficiently in SVK14 cells when alone than when together (Fig. 2B, compare lanes 2–4 to lane 1). This suggests that in addition splicing of one of these exons can favor splicing of the others.

v8 and v9 Exon Sequences That Affect Splicing in Vivo—To begin our analysis of sequences controlling splicing of the CD44 epithelial exons, we decided to search for control elements within exons v8 and v9. The sequences of these exons are shown in Fig. 3A and were arbitrarily divided into short blocks as shown (blocks b1–b9 for exon v8 and b1–b8 for exon v9). We made versions of minigenes v8 and v9 (Fig. 2A) in which blocks (Fig. 3A) were replaced by a random sequence of the same size. For example, in minigene v8Mb9, block 9 of exon v8 has been replaced by a random sequence. These minigenes were trans-

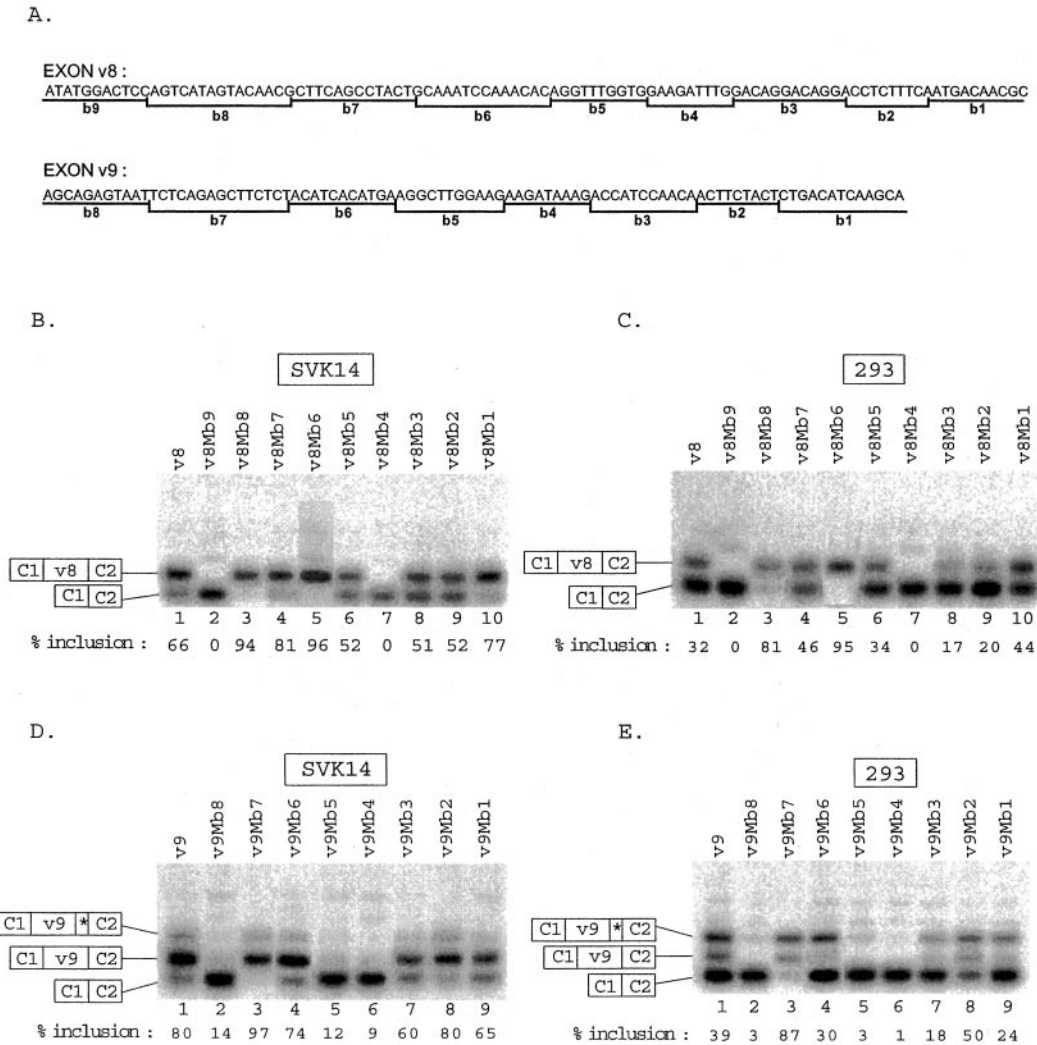


FIG. 3. Exon sequences involved in controlling splicing *in vivo*. A, sequences of v8 and v9 exons are shown, with identification of portions subjected to block mutation (marked as *bn*, where *n* is a number). B–E, RT-PCR analysis of RNA from SVK14 or 293 cells transfected with indicated minigenes. In minigenes v8Mb*n*, sequence *bn* of the v8 exon of minigene v8 has been replaced by a random sequence of the same length. Nomenclature follows the same principle for v9 minigenes. Analysis was as described in the legend to Fig. 2B.

ected into SVK14 or 293 cells, and RNA was harvested and analyzed by RT-PCR using minigene-specific primers. The results are shown in Fig. 3, B–E.

Whereas most of the block mutations have some effect on the relative amounts of RT-PCR products representing v exon splicing or v exon skipping, certain block mutations have particularly marked effects in both cell lines. For the v8 minigene, mutation of b8 and b6 increases the amount of RT-PCR product containing the v8 exon (Fig. 3, B and C, lanes 3 and 5). For the v9 minigene, mutation of b7 increases the amount of RT-PCR product containing the v9 exon (Fig. 3, D and E, lanes 3). In contrast, mutation of v8b9 or v8b4 reduces the amount of RT-PCR product containing the v8 exon to undetectable levels (see Fig. 3, B and C, lanes 2 and 7, respectively). For the v9 minigene, mutation of b8, b5, or b4 markedly reduces the amount of RT-PCR product containing the v9 exon (see Fig. 3, D and E, lanes 2, 5, and 6, respectively).

These results could indicate a role for the above blocks in controlling exon splicing, acting either as exon splicing silencers, whose mutation would increase exon splicing, or ESEs, whose mutation would decrease exon splicing. However, the nucleotide replacements we effected might actually have created motifs with an effect on splicing for some blocks. Alternatively, the replacements may have had an effect on the RT-PCR

technique used for analysis, possibly by rendering fragments containing the replacements difficult to amplify, or they may have rendered RNAs containing these replacements unstable. (This latter possibility seems unlikely, as addition of cycloheximide to transfected cells before harvesting RNA for analysis had no effect on the results obtained, data not shown.) We decided to concentrate our attention on the candidate ESEs v8b9, v8b4, v9b8, v9b5, and v9b4. If these blocks are ESEs, they should be able to activate splicing in a heterologous context when inserted in a poor pre-mRNA substrate. We set out to test whether this was the case.

Some v8 and v9 Exon Sequences Activate Splicing in Vitro—A truncated dsx pre-mRNA containing exons 3 and 4 with the intron separating them (Fig. 4A) has been used previously to analyze ESEs (29). If some of the v8 or v9 blocks discussed above are ESEs, inserting them behind the dsx exon 4 should enhance splicing of the pre-mRNA. However, it seemed possible that any ESE activity of the isolated blocks might be weak, as none of the blocks alone can activate v exon splicing *in vivo*, even partially, but require the assistance of at least one other block (Fig. 3). For this reason we chose to begin our analysis using dimers of the individual monomer blocks v8b9, v8b4, and v9d8, separated by the linker UACGCGUC (Fig. 4A; for example, v8b9 was used to make v8d9, d for

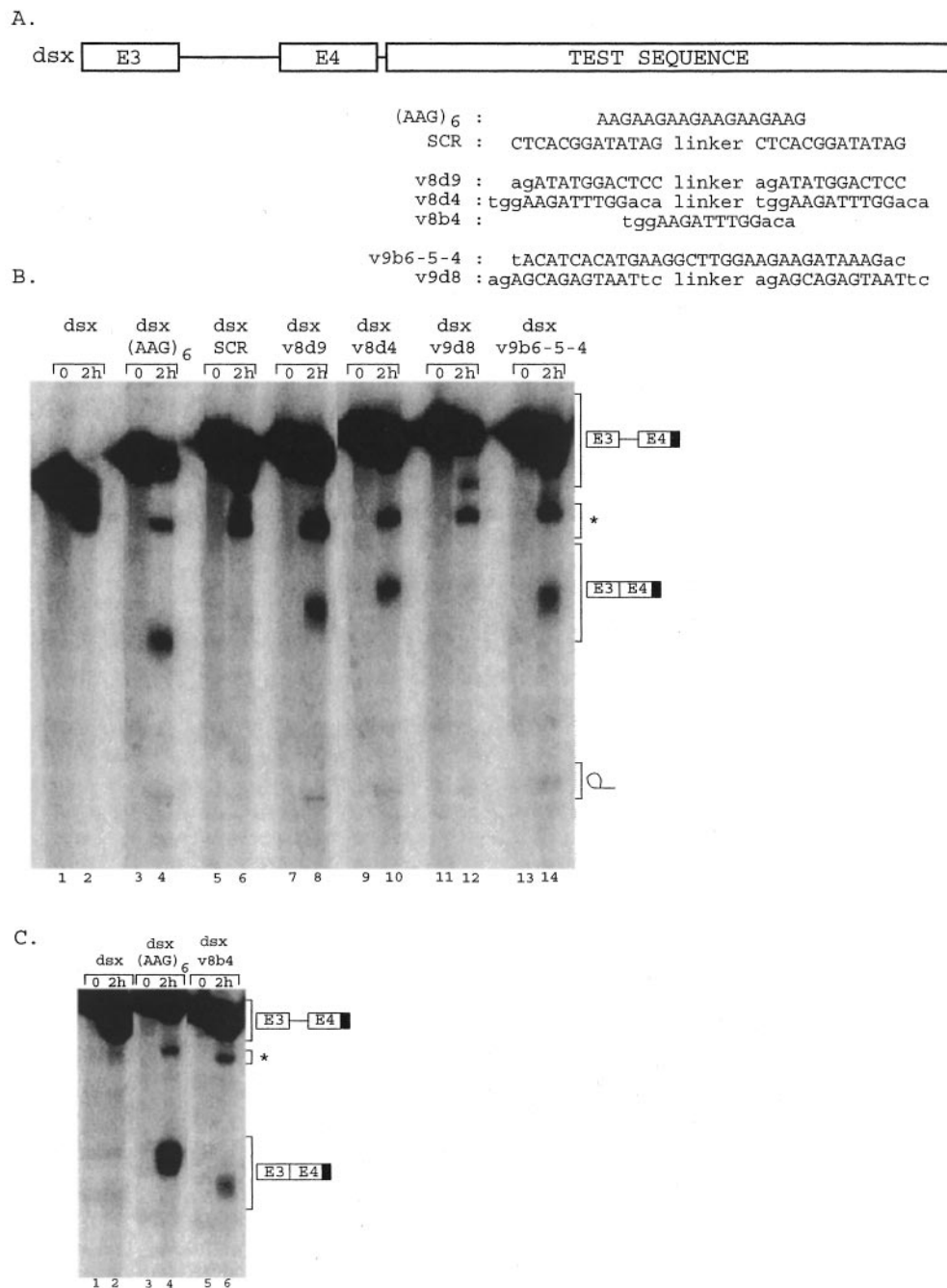


FIG. 4. **Some v8 and v9 exon sequences activate splicing *in vitro*.** A, schematic representation of doublesex plasmids. The drawing is not to scale. Exons E3 and E4 are shown as boxes; the intron separating them is represented by a line. Sequences to be tested for ESE activity were placed behind exon 4 and are represented by the test sequence box. The actual sequences used are shown below the test sequence box. Where dimers were used, the sequences corresponding to the v8 or v9 sequences chosen for block mutation (Fig. 3A) are shown in capital letters, and the flanking natural nucleotides are shown in lowercase letters. Monomers are separated by a linker (sequence UACGCGUC). SCR is a scrambled version of v8d9. B and C, various pre-mRNAs made from doublesex-based plasmids containing test sequences described in A (dsx is the parental plasmid with no test sequence) by *in vitro* transcription were incubated in HeLa cell nuclear extract for the indicated times (0 or 2 h) before denaturing gel analysis. The migration positions of pre-mRNAs, spliced products, and excised intron lariats are indicated. The asterisk identifies a pre-mRNA degradation product. The migration positions of pre-mRNAs and spliced products vary somewhat depending on the length of the test sequence (represented by a black box after exon 4 in the drawings).

dimer). In certain cases, additional natural flanking residues (lowercase letters in Fig. 4A) were added to block sequences, as we were concerned that the totality of an activating motif might not be represented precisely in just one block. However, we tested as a monomer a contiguous block of v9 exon sequence (v9b6-5-4), as this block already contains two of the blocks (b4 and b5) identified in Fig. 3 as needed for v9 splicing. We chose to add the b6 motif as it clearly reduces use of the normal v9 exon 5' ss in 293 cells (compare Fig. 3E, lanes 1 and 4, band

C1-v9-C2), whereas mutation of b6 has little effect on v9 exon splicing in SVK14 cells (Fig. 3D, lane 4), and so b6 may contribute to ESE function exerted by b4 and b5.

As a positive control, we used a 6-fold repetition of an AAG motif (AAG)₆ known to activate splicing in this system (29). As a negative control, we used a scrambled version of v8d9 (SCR). These pre-mRNAs were tested in an *in vitro* splicing assay using HeLa cell nuclear extract. As expected, little or no splicing was observed with either the parental dsx pre-mRNA or the

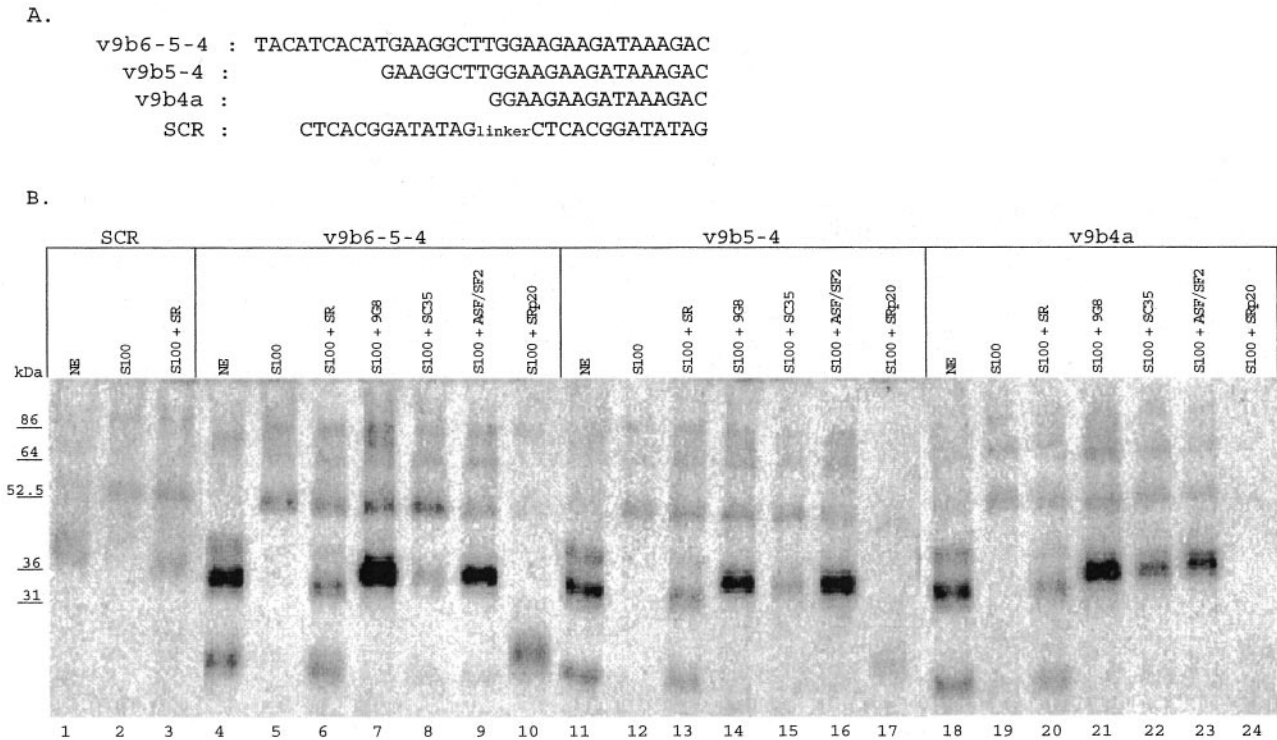


FIG. 5. **A v9 ESE cross-links to SR proteins.** A, sequences of RNA probes used for cross-linking. B, RNA molecules containing the sequences indicated were incubated either in HeLa cell nuclear extract (NE), S100 extract (S100), or S100 extract to which a preparation of total SR proteins (SR) or individual SR proteins as shown had been added, before UV cross-linking and SDS-PAGE. Note that the recombinant SRp20 has an N-terminal tag and so is slightly larger than the SRp20 in nuclear extract. This explains why the SRp20 band observed in nuclear extract samples migrates slightly faster than the SRp20 band observed in S100 samples with added recombinant protein.

pre-mRNA containing the scrambled sequence (Fig. 4B, compare lanes 1 and 2 and compare lanes 5 and 6). Note that the band indicated by an asterisk in Fig. 4B is not a splicing product but is probably the result of pre-mRNA degradation taking place in the nuclear extract. In contrast, spliced product (E3E4) and excised intron lariat were detected in splicing assays using the pre-mRNA containing the (AAG)₆ motif (Fig. 4B, compare lanes 3 and 4).

Both the two v8 exon block dimers tested (v8d9 and v8d4) activated splicing *in vitro*, giving results qualitatively similar to those obtained with the (AAG)₆-positive control (Fig. 4B, compare lanes 7 and 8, and 9 and 10, to lanes 3 and 4, respectively). We then tested the corresponding monomers v8b9 and v8b4. v8b4 activated splicing *in vitro* when placed in the dsx substrate (Fig. 4C, lane 6). However, the dsx substrate containing v8b9 was very unstable in nuclear extract for reasons not clear to us, and so we were unable to determine whether v8b9 can activate splicing in this system.

The v9 exon sequence v9d8 did not activate splicing *in vitro* (Fig. 4B, lanes 11 and 12) and so may not be a true ESE. However, the contiguous block of v9 exon sequence v9b6-5-4 did activate splicing (Fig. 4B, compare lanes 13 and 14).

v9b6-5-4 Binds to SR Proteins 9G8 and SRp20—v8d9, v8d4, v8b4, and v9b6-5-4 act as functional ESEs *in vitro*. We wished to identify the proteins that activate splicing through them. These sequences, together with the scrambled (SCR) negative control, were incubated in HeLa cell nuclear extract before UV cross-linking and gel analysis. The contiguous block of v9 exon sequence covering blocks 4–6 (v9b6-5-4) cross-linked to proteins of ~35 and 20 kDa and more weakly to a ~40-kDa protein (Fig. 5B, lane 4). In comparison, these proteins cross-linked only very weakly to the SCR sequence (Fig. 5B, lane 1). We also tested two shorter versions of v9b6-5-4 (Fig. 5A): v9b5-4 (corresponding to blocks 4 and 5 of Fig. 3 with some additional nucleotides), and v9b4a (corresponding to block 4 of

Fig. 3 with some additional nucleotides). Both these RNAs also cross-linked to 40-, 35-, and 20-kDa proteins, although cross-linking to the 35-kDa protein was somewhat reduced relative to v9b6-5-4 (see Fig. 5B, lanes 11 and 18 for v9b5-4 and v9b4a, respectively).

Given their size, it seemed worthwhile testing if the cross-linked proteins were SR proteins. S100 extract contained very low levels of SR proteins as compared with nuclear extract. v9b6-5-4, v9b5-4, and v9b4a, together with the SCR control, were incubated in parallel in S100 extract, and in S100 extract to which a preparation of total SR proteins had been added. As shown in Fig. 5B, no significant cross-linking to the 35- and 20-kDa proteins was observed in S100 extract for any of the RNAs (lanes 2, 5, 12, and 19). However, for v9b6-5-4, v9b5-4, and v9b4a, cross-linking to 35- and 20-kDa proteins was readily observed in S100 with added total SR proteins, and this cross-linking was significantly stronger than that observed with the SCR sequence (compare lanes 6, 13, and 20 to lane 3, Fig. 5B). These results show that the 35- and 20-kDa cross-linking proteins are SR proteins. We were not able to identify the 40-kDa cross-linking protein as an SR protein in these experiments (SRp40 was a possible candidate), as there was no clear increase in cross-linking to a 40-kDa protein when cross-linkings carried out in S100 or S100 plus SR proteins were compared for either v9b6-5-4, v9b5-4, or v9b4a (Fig. 5B, lanes 6, 13, and 20).

Several 35-kDa SR proteins are known, including 9G8, ASF/SF2, and SC35, whereas the 20-kDa protein could correspond to SRp20. To identify which SR proteins bind to the v9 exon sequences, cross-linking experiments were repeated in S100 extract using individual recombinant SR proteins instead of a total preparation. v9b6-5-4 cross-linked to 9G8, ASF/SF2, and SRp20 but only very weakly to SC35 (Fig. 5B, lanes 7, 9, 10, and 8, respectively). Similar results were obtained with v9b5-4, although cross-linking to both 9G8 and SRp20 was weaker

than for v9b6-5-4 (Fig. 5B, lanes 14–17). One likely explanation for this is that v9b6-5-4 (but not v9b5-4) contains the sequence UACAUCACAU, which conforms to one of the consensus motifs (AC(A/U)(A/U)C) proposed for SRp20-binding sites (30). In addition, there is some evidence that 9G8 and SRp20 can bind to the same RNA sequences (30).

v9b4a cross-linked to 9G8 and to ASF/SF2, but v9b4a cross-linking to SRp20 was even weaker than for v9b5-4 (Fig. 5B, lanes 21, 23, and 24, respectively). However, both v9b5-4 and v9b4a did cross-link to a 20-kDa protein in nuclear extract when total SR proteins were tested (lanes 11 and 18). By immunoprecipitating the nuclear extract cross-linkings with an anti-SRp20 antibody, we demonstrated that this 20-kDa protein was indeed SRp20 (data not shown). These observations suggest that v9b5-4 and v9b4a can bind SRp20 but that this binding depends on binding of other proteins, probably SR proteins. This is in contrast to v9b6-5-4, which probably binds SRp20 independently via its UACAUCACAU motif, see above.

For analysis of the v8 exon sequences, we used the same approach as that detailed above for the v9 exon sequences. v8d4 and v8d9, together with the scrambled (SCR) negative control, were incubated in HeLa cell nuclear extract before UV cross-linking and gel analysis. We detected a 45-kDa protein that cross-linked strongly to v8d4 but not to the SCR control, but comparing cross-linking patterns in S100 and S100 plus total SR proteins showed that this protein was not an SR protein (data not shown), and we have not yet identified it. We were unable to detect any protein showing specific cross-linking to v8b9 (data not shown).

9G8, ASF/SF2, and SRp20 Activate Splicing in Vitro of a Heterologous Exon Carrying v9b6-5-4—9G8, ASF/SF2, and SRp20 bind to v9 exon sequences, but will this binding activate splicing? This can be tested by comparing splicing of a pre-mRNA substrate in S100 extract with or without added SR proteins. For this approach, we decided to use the pre-mRNA substrate Sp1 ex2 inv (Fig. 6A), based on the adenoviral E1A pre-mRNA. We preferred this substrate to the dsx substrate for the splicing activation tests for two reasons. First, this substrate has been used previously for studies of splicing activation by SR proteins, including SRp20, in one of our laboratories (30). Second, v9 sequences activated splicing more efficiently in the Sp1 ex2 inv substrate as compared with the dsx substrate, although the relative activation ability of individual blocks was the same for both substrates (data not shown).

In the absence of any exogenous activating sequence, the Sp1 ex2 inv pre-mRNA substrate is not spliced to any significant degree in the presence of nuclear extract (Fig. 6B, lane 2). Splicing of Sp1 ex2 inv substrates was activated by v9b6-5-4, v9b5-4, and v9b4a in the presence of nuclear extract (compare Fig. 6C, lanes 1, 6, and 11 to Fig. 6B, lane 2). For the substrate containing v9b5-4, splicing in nuclear extract was as efficient as for that containing v9b6-5-4 (compare lanes 1 and 6, Fig. 6C). This observation is consistent with results obtained *in vivo*, as mutation of the extra sequences (b6) present in v9b6-5-4 had only a minor effect on v9 exon splicing *in vivo* (Fig. 3, D and E, lanes 4). Relative to results obtained with substrates with v9b6-5-4 and v9b5-4, splicing observed in nuclear extract of the substrate with v9b4a (Fig. 6C, lane 11) dropped ~2-fold. This is again consistent with results obtained *in vivo*, as v9b4a lacks the b5 sequences required for v9 exon splicing *in vivo* (Fig. 3, D and E, lanes 5).

In the absence of any exogenous activating sequence, the Sp1 ex2 inv pre-mRNA substrate is not spliced to any significant degree in the presence of S100 extract, even if SR proteins are

added (Fig. 6B, lanes 3–6). However, if known binding sites for individual SR proteins are inserted in the downstream exon, splicing now takes place in S100 extract with the corresponding SR protein added (see Fig. 6B, lanes 7–12). To allow meaningful comparisons of the abilities of the different SR proteins to use the v9 sequences as a splicing enhancer, we first determined, for each SR protein, the molar amount of protein required to obtain about 30% splicing activation using a substrate containing an already characterized enhancer for it (Fig. 6B, lanes 8, 10, and 12; see also the figure legend). These amounts of the various proteins were then used to activate splicing of Sp1 ex2 inv substrates with added v9 sequences, as shown in Fig. 6C.

9G8, ASF/SF2, and SRp20 activated splicing of substrates with v9b6-5-4 or v9b5-4, whereas only 9G8 still activated significantly substrates with v9b4a. Of the proteins tested, 9G8 gave the highest splicing activation in S100 extract for all three substrates tested (Fig. 6C, lanes 4, 9, and 14). Splicing of the v9b5-4 substrate could be activated in S100 by 9G8 to an extent similar to that seen with the v9b6-5-4 substrate (compare lanes 4 and 9), whereas activation by ASF/SF2 and SRp20 was less efficient for the v9b5-4 substrate than for the v9b6-5-4 substrate (compare lanes 3 and 8 and lanes 5 and 10). Given that splicing of the two substrates is equally efficient in the presence of nuclear extract (compare lanes 1 and 6), these comparisons show that 9G8 probably makes the major contribution to activation in the nuclear extract. With the v9b4a substrate, splicing activation dropped both with nuclear extract (lane 11) and with 9G8 in S100 extract (lane 14).

When we compared the cross-linking and *in vitro* activation data, we did not always observe a perfect correlation. For example, ASF/SF2 cross-links equally well to v9b6-5-4, v9b5-4, and v9b4a, yet *in vitro* activation by ASF/SF2 drops progressively from v9b6-5-4, to v9b5-4, to v9b4a. Perhaps activation here requires not only binding but also some more specific sequence interaction, which is progressively lost when passing from v9b6-5-4 to v9b5-4 to v9b4a. For SRp20, the situation appears simpler, as there is a progressive loss of cross-linking to SRp20 alone and of SRp20-induced splicing activation, when passing from v9b6-5-4 to v9b5-4 to v9b4a.

DISCUSSION

Alternative exons v8, v9, and v10 are included as a block in CD44 mRNA in a variety of epithelial cells but are all excluded in many other cell types. We have studied splicing of these exons using a minigene-based approach in two cell lines, which either splice or skip the three exons as a block. With minigenes containing only v8, or only v9, or only v10, the same extent of cell type-specific splicing of the exon is observed as with a minigene containing all three alternative exons. Thus cell type-specific control of splicing is exerted on each of the three alternative exons.

In an attempt to identify internal exon sequences involved in v8 and v9 splicing, we tested the effects of short block mutations on splicing of these exons on minigene-derived pre-mRNAs *in vivo*. In this way we identified for both exons sequences whose mutation either increased or decreased the abundance of spliced transcripts containing the exons. In particular, mutation of one v8 exon sequence (v8b4, UGG-AAGAUUUGGACA) and one v9 exon sequence (v9b5-4, GAA-GGCUUGGAAGAAGAUAAAGAC) blocks splicing of the corresponding exons *in vivo* almost completely. Both these sequences can activate splicing of a heterologous exon *in vitro* in nuclear extract, and so they are true ESEs. In current models for ESE function, the enhancer binds proteins which in turn aid binding of splicing factors to the 5' or 3' splice sites of the exon (2, 35–37), although enhancer-bound proteins may also be able to activate a later step of splicing (38). The 3' splice site-

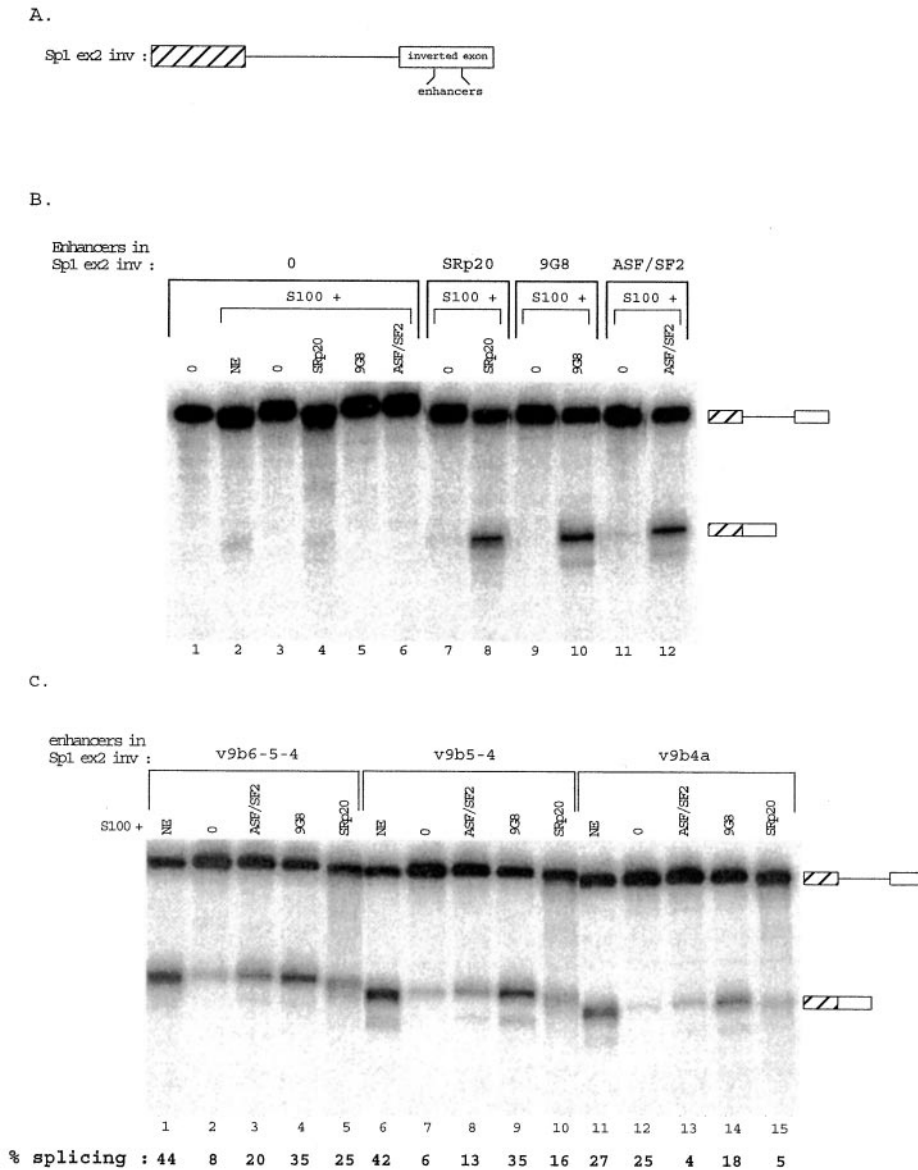


FIG. 6. A v9 ESE renders splicing of a heterologous exon sensitive to SRp20, 9G8, and ASF/SF2. *A*, schematic representations of Sp1 ex2 inv-based plasmids containing two exons (boxes) separated by an intron (line). Sequences to be tested were inserted within the inverted exon as shown. *B*, determination of the molar amounts of individual SR species required to activate splicing to 30–33% efficiency through the use of specific SR enhancers. The 20 nucleotide sequences used for each SR species are as follows: CUUCGAUCAACGCCACGCCA for SRp20, GACAACGACGACGACUAGAA for 9G8, and GCACAGGACGAAGCUGCACC for ASF/SF2. The Sp1 ex2 inv substrate without enhancer (lanes 1–6) or in the presence of S100, complemented or not by individual SR proteins (amounts used were 20, 18, and 6 μ M for SRp20, 9G8, and ASF/SF2, respectively) as indicated above the lanes. *C*, abilities of individual SR proteins to activate splicing through the v9 ESE. Sp1 ex2 inv substrates containing various v9 sequences as indicated above the lanes (and which correspond to sequences shown in Fig. 5A) were spliced in the presence of a mixture of nuclear extract and S100 (lanes 1, 6, and 11) or in the presence of S100 complemented or not by individual SR proteins (20 μ M SRp20, 18 μ M 9G8, or 6 μ M ASF/SF2), as indicated above the lanes. The detection of a weak splicing in the absence of added SR proteins in lanes 2, 7, and 12 may be due to the presence of residual amounts of SR proteins in the S100 and/or the 15–45% extracts (see “Experimental Procedures”). For *B* and *C*, pre-mRNAs and spliced products are identified by drawings.

associated polypyrimidine sequences of the v8 and v9 exons are very short, making U2AF recruitment one attractive candidate target for v9b5-4-bound activators.

We identified the SR proteins 9G8, ASF/SF2, and SRp20 as proteins that bind to v9b5-4, and can activate splicing of an exon carrying the v9b5-4 motif *in vitro*. In addition, an ~40-kDa protein, which we have not identified but which does not appear to be an SR protein, cross-linked to v9b5-4 in nuclear extract. Of the three SR proteins, 9G8 appears to be the major activator of splicing *in vitro*, activating splicing via v9b5-4 2- or 3-fold more efficiently than ASF/SF2 or SRp20. It is interesting in this respect to note that part of v9b5-4, which we termed

v9b4a (GGAAGAAGAUAAAGAC) is quite similar to part of a natural ESE from exon 4 of the 9G8 gene (GGAAGGCGAUA), which interacts preferentially with 9G8 (40). Our results strongly suggest that 9G8, and to a lesser extent ASF/SF2 and SRp20, possibly together with non-SR proteins, participate in activating a v9 ESE to induce splicing of the v9 exon *in vivo*.

The list of cellular exons known to contain SR protein-dependent ESEs is quite short (4), and our data add an important new member, the CD44 v9 exon, to this list. Indeed, if the notion that SR proteins can activate exon splicing by binding to ESEs is widely accepted, much of the experimental evidence supporting this notion stems from analysis of sequences se-

lected *in vitro* from pools of random sequences, based on their ability to bind SR proteins or to activate splicing *in vitro* (see Ref. 39 for a review). We have identified RNA sequences on a genuine exon that bind SR proteins for splicing activation. How do these sequences compare with the *in vitro* selected sequences?

RNA sequences selected in a SELEX approach by 9G8 (30) conform to consensus AGAC(G/U)ACGA(C/U) and ACGAGA-GA(C/U). v9b6-5-4 (UACAUCACAUGAAGGCUUGGAAGAA-GAUAAAGAC), v9b5-4, and v9b4a all cross-link to 9G8, and all contain several repeats of the motif GAN. They are thus reasonably similar to the first SELEX consensus. 9G8 cross-links more strongly to v9b6-5-4 than to v9b5-4 or v9b4a, and this may be because of its cross-linking to the UACAUCACAUG motif present in v9b6-5-4 (although binding of 9G8 to this motif does not appear to participate in splicing activation). This motif is similar to one of the consensus motifs (AC(A/U)(A/U)C) proposed for SRp20-binding sites (30), and there is some evidence that 9G8 and SRp20 can bind to the same RNA sequences, even if their respective high affinity targets are significantly different (30). The RNA binding domains of 9G8 and SRp20 are similar, although 9G8 differs from SRp20 in that it contains a zinc knuckle (41), and it has been shown that a 9G8 variant mutated in its zinc knuckle is a functional homologue of SRp20 (30). As a further example, RBP1, the *Drosophila* homologue of SRp20, binds with other proteins to a doublesex enhancer in *Drosophila* extracts, whereas if binding is carried out in HeLa cell nuclear extract, it is 9G8 that binds instead (42).

One SELEX sequence for ASF/SF2 is RGAAGAAC, and the motif GGAAGAAG is present in v9b6-5-4, v9b5-4, and v9b4a (all of which cross-link to ASF/SF2). As was observed for 9G8, cross-linking to SRp20 is significantly stronger to v9b6-5-4 than to v9b5-4 or v9b4a when SRp20 is tested alone. This is probably because, as discussed above, the v9b6 sequence (UACAUCACAUGAA) conforms to one of the consensus motifs (AC(A/U)(A/U)C) proposed for SRp20-binding sites (30). There is no significant similarity between the purine-rich v9b5-4 and v9b4a and consensus motifs proposed for SRp20-binding sites, which are generally pyrimidine-rich (29). However, SRp20 cross-links weakly to v9b5-4 and activates splicing *in vitro* via this motif, although less effectively than 9G8. SRp20 cross-links significantly more efficiently to v9b5-4 and v9b4a in nuclear extract or when part of a preparation of total SR proteins than when tested alone, suggesting that other proteins, present in the SR protein preparation, improve SRp20 interaction with v9b5-4 and v9b4a. It is thus possible that the ability of SRp20 to activate v9 exon splicing via v9b5-4 may be underestimated by the *in vitro* analysis testing SRp20 alone, as it may be more effective in the presence of other proteins.

The v9 exon is spliced in a cell type-specific manner. Could 9G8, SRp20, or ASF/SF2 binding to the v9 ESE be implicated? A difference in 9G8, SRp20, or ASF/SF2 protein levels cannot explain the difference in v9 exon splicing we have observed between the two cell lines used here (Western blot analysis detects similar levels of the three proteins in the two lines).² In agreement with this, overexpression of SRp20 *in vivo* does not increase v9 exon splicing in 293 cells; we have not been able to analyze the effect of 9G8 or ASF/SF2 overexpression, as overexpression of either protein activates very efficiently a cryptic splice site within FGFR-2 intron sequences of our minigene.² However, protein levels do not necessarily reflect activity levels with proteins such as SR proteins, which are known to undergo phosphorylation (43). It remains possible that the activity lev-

els of 9G8, ASF/SF2, or SRp20 differ between the cell lines. This could contribute to the observed difference in v9 exon splicing. It is equally possible that various negative influences, including large intron sizes, and various positive influences, including 9G8/SRp20/ASF/SF2 bound to the v9 ESE, are used to poise the v9 exon on the brink of splicing. Cell type-specific input from yet to be identified sequences would then tip the balance in favor of splicing or splicing repression.

Differences in the relative levels of individual SR proteins and their inhibitors have been observed during tumorigenesis, and attempts have been made to correlate these changes with changes in CD44 pre-mRNA splicing (44, 45). To the best of our knowledge, our results with the v9 exon are the first indication that individual SR proteins are involved in splicing regulation of any CD44 v exon (previous work has identified proteins activating splicing of exons v4 and v5, but the proteins in question are not members of the SR protein family (22, 24, 27, 28)). It is thus unfortunate that v9 splicing was not analyzed directly in any of the studies. However, it is intriguing that SRp20 is one of the SR proteins induced upon tumorigenesis in a mouse mammary gland model, with a progressive increase in expression as normal cells progress to pre-neoplasias and then to tumors (48). Inclusion of CD44 exon v8 was only detected in RNA from tumor cells. Exons v8, v9, and v10 are frequently spliced as a block (34), and so it is conceivable that v9 exon splicing is also elevated in the tumor cells, possibly as a result of high level expression of SRp20.

Acknowledgments—We thank Renata Gattoni for helpful discussions and for a critical reading of the manuscript. We thank the staff of the IGBMC facilities for their assistance.

REFERENCES

- Graveley, B. R. (2001) *Trends Genet.* **17**, 100–107
- Maniatis, T., and Tasic, B. (2002) *Nature* **418**, 236–243
- Caceres, J. F., and Kornbliht, A. R. (2002) *Trends Genet.* **18**, 186–193
- Smith, C. W., and Valcarcel, J. (2000) *Trends Biochem. Sci.* **25**, 381–388
- Markovtsov, V., Nikolic, J. M., Goldman, J. A., Turck, C. W., Chou, M. Y., and Black, D. L. (2000) *Mol. Cell. Biol.* **20**, 7463–7479
- Chou, M. Y., Underwood, J. G., Nikolic, J., Luu, M. H., and Black, D. L. (2000) *Mol. Cell* **5**, 949–957
- Modafferi, E. F., and Black, D. L. (1999) *RNA (New York)* **5**, 687–706
- Del Gatto, F., Plet, A., Gesnel, M. C., Fort, C., and Breathnach, R. (1997) *Mol. Cell. Biol.* **17**, 5106–5116
- Del Gatto-Konczak, F., Bourgeois, C. F., Le Guiner, C., Kister, L., Gesnel, M. C., Stevenin, J., and Breathnach, R. (2000) *Mol. Cell. Biol.* **20**, 6287–6299
- Del Gatto-Konczak, F., Olive, M., Gesnel, M. C., and Breathnach, R. (1999) *Mol. Cell. Biol.* **19**, 251–260
- Carstens, R. P., Wagner, E. J., and Garcia-Blanco, M. A. (2000) *Mol. Cell. Biol.* **20**, 7388–7400
- Carstens, R. P., McKeehan, W. L., and Garcia-Blanco, M. A. (1998) *Mol. Cell. Biol.* **18**, 2205–2217
- Le Guiner, C., Plet, A., Galiana, D., Gesnel, M. C., Del Gatto-Konczak, F., and Breathnach, R. (2001) *J. Biol. Chem.* **276**, 43677–43687
- Ponta, H., Wainwright, D., and Herrlich, P. (1998) *Int. J. Biochem. Cell Biol.* **30**, 299–305
- Sneath, R. J., and Mangham, D. C. (1998) *Mol. Pathol.* **51**, 191–200
- Goodison, S., Urquidí, V., and Tarin, D. (1999) *Mol. Pathol.* **52**, 189–196
- Herrlich, P., Morrison, H., Sleeman, J., Orian-Rousseau, V., König, H., Weg-Remers, S., and Ponta, H. (2000) *Ann. N. Y. Acad. Sci.* **910**, 106–118
- Screation, G. R., Bell, M. V., Jackson, D. G., Cornelis, F. B., Gerth, U., and Bell, J. I. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89**, 12160–12164
- Gunther, U., Hofmann, M., Rudy, W., Reber, S., Zoller, M., Haussmann, I., Matzku, S., Wenzel, A., Ponta, H., and Herrlich, P. (1991) *Cell* **65**, 13–24
- Ponta, H., Sleeman, J., Dall, P., Moll, J., Sherman, L., and Herrlich, P. (1994) *Invasion Metastasis* **14**, 82–86
- Yamaguchi, A., Zhang, M., Goi, T., Fujita, T., Niimoto, S., Katayama, K., and Hirose, K. (2000) *Oncol. Rep.* **7**, 541–544
- Stickeler, E., Fraser, S. D., Honig, A., Chen, A. L., Berget, S. M., and Cooper, T. A. (2001) *EMBO J.* **20**, 3821–3830
- Weg-Remers, S., Ponta, H., Herrlich, P., and König, H. (2001) *EMBO J.* **20**, 4194–4203
- Honig, A., Auboeuf, D., Parker, M. M., O'Malley, B. W., and Berget, S. M. (2002) *Mol. Cell. Biol.* **16**, 5698–5707
- König, H., Ponta, H., and Herrlich, P. (1998) *EMBO J.* **17**, 2904–2913
- Matter, N., Marx, M., Weg-Remers, S., Ponta, H., Herrlich, P., and König, H. (2000) *J. Biol. Chem.* **275**, 35353–35360
- Stoss, O., Olbrich, M., Hartmann, A. M., König, H., Memmott, J., Andreadis, A., and Stamm, S. (2001) *J. Biol. Chem.* **276**, 8665–8673
- Matter, N., Herrlich, P., and König, H. (2002) *Nature* **420**, 691–695
- Schaal, T. D., and Maniatis, T. (1999) *Mol. Cell. Biol.* **19**, 1705–1719

² D. Galiana-Arnoux, R. Breathnach, and F. Del Gatto-Konczak, unpublished results.

30. Cavaloc, Y., Bourgeois, C. F., Kister, L., and Stevenin, J. (1999) *RNA (New York)* **5**, 468–483
31. Taylor-Papadimitriou, J., Purkis, P., Lane, E. B., McKay, I. A., and Chang, S. E. (1982) *Cell Differ.* **11**, 169–180
32. Gallego, M. E., Gattoni, R., Stevenin, J., Marie, J., and Expert-Bezancon, A. (1997) *EMBO J.* **16**, 1772–1784
33. Yu, Q., and Toole, B. P. (1996) *J. Biol. Chem.* **271**, 20603–20607
34. Bell, M. V., Cowper, A. E., Lefranc, M. P., Bell, J. L., and Screaton, G. R. (1998) *Mol. Cell. Biol.* **18**, 5930–5941
35. Valcarcel, J., and Green, M. R. (1996) *Trends Biochem. Sci.* **21**, 296–301
36. Blencowe, B. J. (2000) *Trends Biochem. Sci.* **25**, 106–110
37. Graveley, B. R. (2000) *RNA (New York)* **6**, 1197–1211
38. Chew, S. L., Liu, H. X., Mayeda, A., and Krainer, A. R. (1999) *Proc. Natl. Acad. Sci. U. S. A.* **96**, 10655–10660
39. Cartegni, L., Chew, S. L., and Krainer, A. R. (2002) *Nat. Rev. Genet.* **3**, 285–298
40. Lejeune, F., Cavaloc, Y., and Stevenin, J. (2001) *J. Biol. Chem.* **276**, 7850–7858
41. Cavaloc, Y., Popielarz, M., Fuchs, J. P., Gattoni, R., and Stevenin, J. (1994) *EMBO J.* **13**, 2639–2649
42. Lynch, K. W., and Maniatis, T. (1996) *Genes Dev.* **10**, 2089–2101
43. Stojdl, D. F., and Bell, J. C. (1999) *Biochem. Cell Biol.* **77**, 293–298
44. Ghigna, C., Moroni, M., Porta, C., Riva, S., and Biamonti, G. (1998) *Cancer Res.* **58**, 5818–5824
45. Stickeler, E., Kittrell, F., Medina, D., and Berget, S. M. (1999) *Oncogene* **18**, 3574–3582