

Research Article

Motifs within the CA-repeat-rich region of Surfactant Protein B (SFTPB) intron 4 differentially affect mRNA splicing

Wenjun Yang^{1*}, Lan Ni^{1*}, Patricia Silveyra¹, Guirong Wang¹, Georgios T Noutsios¹, Anamika Singh¹, Susan L DiAngelo¹, Olabisi Sanusi¹, Manmeet Raval¹ and Joanna Floros^{1,2}

¹Center for Host Defense, Inflammation, and Lung Disease (CHILD) Research, Department of Pediatrics, Penn State Hershey College of Medicine, Pennsylvania State University, Hershey, Pennsylvania

²Department of Obstetrics and Gynecology, Penn State Hershey College of Medicine, Pennsylvania State University, Hershey, Pennsylvania.

* contributed equally

Received on September 28, 2012; Accepted on January 2, 2013; Published on February 20, 2013

Correspondence should be addressed to Joanna Floros; Email: jfloros@psu.edu

Current address: Ningxia Medical University, Ningxia, China, 750004 (WY), Wuhan University, Hubei, China, 430070 (LN), Temple University School of Medicine, Philadelphia, PA, 19140, USA (AS), Department of Surgery, SUNY Upstate Medical University, Syracuse, NY, 13210 USA (GW), Department of Physiology, Penn State University, College of Medicine, Hershey, PA 17033, USA (MR).

Abstract

The first half of the surfactant protein B (SP-B) gene intron 4 is a CA-repeat-rich region that contains 11 motifs. To study the role of this region on SP-B mRNA splicing, minigenes were generated by systematic removal of motifs from either the 5' or 3' end. These were transfected in CHO cells to study their splicing efficiency. The latter was determined as the ratio of completely to incompletely spliced SP-B RNA. Our results indicate that SP-B intron 4 motifs differentially affect splicing. Motifs 8 and 9 significantly enhanced and reduced splicing of intron 4, respectively. RNA mobility shift assays performed with

a Motif 8 sequence that contains a CAUC *cis*-element and cell extracts resulted in a RNA:protein shift that was lost upon mutation of the element. Furthermore, *in silico* analysis of mRNA secondary structure stability for minigenes with and without motif 8 indicated a correlation between mRNA stability and splicing ratio. We conclude that differential loss of specific intron 4 motifs results in one or more of the following: a) altered splicing, b) differences in RNA stability and c) changes in secondary structure. These, in turn, may affect SP-B content in lung health or disease.

Introduction

Messenger RNA splicing is essential for generating mature mRNAs and forming highly complex proteomes from a relatively small number of human genes (Hui *et al.* 2005, Hung *et al.* 2008, Lin *et al.* 2005). An accurate mRNA splicing process requires three essential core sequence elements: the 5' splice site (5'ss), 3' splice site (3'ss), and the branch point (BP) sequence. These sequences are recognized by components of the spliceosome during the processes of intron removal and exon joining (Pastor *et al.* 2009, Ward & Cooper 2010). However, these three core sequences are necessary but not sufficient for defining intron-exon junctions. There are additional sequences affect-

ing splicing efficiency, referred to as exonic or intronic splicing enhancers (ESE or ISE), and exonic or intronic silencers (ESS or ISS) (Pastor *et al.* 2009, Ward & Cooper 2010).

A growing number of diseases appear to rise due to altered splicing, the result of mutations within enhancer and/or silencer sequences. Such mutations may reduce the efficiency with which a constitutive exon is spliced or affects alternative splicing patterns, leading to missplicing and typical skipping, activation of cryptic splice sites, or intron retention (Ward & Cooper 2010). A number of diseases have been reported to result from intronic mutations. For example, a mutation in intron 20 of familial dysautonomia (FD) gene was associated with FD (Slaugenhaupt *et al.*

2001); two mutations in introns 4 and 14 of adenomatous polyposis (AP) *coli* gene with familial AP (Neklason *et al.* 2004, Tuohy *et al.* 2010); a single polymorphism in intron 30 of neurofibromatosis 1 gene with neurofibromatosis type 1 (Raponi *et al.* 2006); and a mutation in intron 9 of the Wilms' tumor 1 gene with Frasier's syndrome (Melo *et al.* 2002). In addition, work from us and others has shown size length variants of intron 4 of the human surfactant protein B gene (SP-B; Gene ID: 6439; Locus tag: HGNC:10801; MIM: 178640) to associate with respiratory distress syndrome (RDS), and other pulmonary diseases, either alone or in combination with the presence of specific SP-A variants (Floros *et al.* 1995, 2001, Hamvas *et al.* 2005, Kala *et al.* 1998, Lin *et al.* 2005, Rova *et al.* 2004, Seifart *et al.* 2002a).

SP-B is essential for maintaining normal surface tension at the air-liquid interface in the alveolus of the lung. The absence of SP-B is incompatible with life, and deficiency of SP-B protein compromises lung function (Gower & Noguee 2011, Nesslerin *et al.* 2005, Tokieda *et al.* 1997). SP-B protein also maintains alveolar epithelial integrity and inhibits endotoxin-induced lung inflammation (Gower & Noguee 2011, Wert *et al.* 2009). Studies in heterozygous SP-B (+/-) mice have demonstrated that a reduction in SP-B associates with decreased lung compliance, affects pulmonary mechanics, alters lung function, and increases susceptibility to lung injury caused by hyperoxia (Nesslerin *et al.* 2005, Tokieda *et al.* 1997, 1999). Both animal and clinical studies have shown associations of reduced SP-B levels with airway infection, indicating that infection by a variety of airway pathogens can induce changes in lung structure and function via reduction of SP-B expression (Beers *et al.* 1999, Kerr & Paton 1999).

The *SFTPB* (SP-B) gene maps on the short arm of chromosome 2, spans a 9.7kb region, and consists of 11 exons, of which exon 11 contains the 3' untranslated region. Ten exons are transcribed into one mRNA of about 2 kb of length that directs the synthesis of a 381 amino acid lung-specific, hydrophobic 42 kDa SP-B preproprotein. The mature SP-B protein is encoded by exons 6 and 7, yielding a 79 amino acid, 8.7 kDa protein (Guttentag *et al.* 1998, O'Reilly *et al.* 1989, Serrano *et al.* 2006, Weaver 1998, Weaver *et al.* 2002). A length polymorphism in intron 4 has been identified (Floros *et al.* 1995, Hamvas *et al.* 2005, Lin *et al.* 2005). The variable region is localized approximately in the first half of intron 4. This region is proximal to the 5'ss of intron 4 and is characterized by the presence of 11 stretches of CA-repeat rich motifs, in which more than 65% of the nucleotides are CAs. Each motif consists of two elements; a relatively con-

served sequence of approximately 20bp, followed by a variable number of CA repeats (Floros *et al.* 1995, Hamvas *et al.* 2005). At least 8 deletions and 11 insertions have been identified in this CA-repeat-rich region (Floros *et al.* 1995, Hamvas *et al.* 2005, Lin *et al.* 2005).

Although SP-B intron 4 length variants have been shown to associate with neonatal RDS (Floros & Fan 2001, Floros *et al.* 1995, 2001, Hallman *et al.* 2002, Hamvas *et al.* 2005), bronchopulmonary dysplasia (Rova *et al.* 2004), chronic obstructive pulmonary disease (Seifart *et al.* 2002a), and lung cancer (Ewis *et al.* 2006, Lin *et al.* 2005, Seifart *et al.* 2002b), the contribution of intron 4 size variation to the development of these conditions has not been yet elucidated. Motifs of CA-repeat-rich variants of SP-B intron 4 have been proposed to alter RNA splicing, a process necessary for normal gene expression (Lin *et al.* 2005, Rova *et al.* 2004). Study of several naturally occurring deletion mutants has indicated that SP-B intron 4 deletion variants affected SP-B mRNA splicing (Lin *et al.* 2005). Each motif in SP-B intron 4 contains a fairly conserved GC rich sequence and various lengths of CA repeats (3-17 CAs). These motifs may contain either important ISE or ISS sites for binding of splicing regulatory proteins or play key roles in maintaining RNA structural stability. In addition, several recent reports indicate that RNA secondary structure may play an important role in determining the binding of proteins to RNA sequences (Buratti & Baralle 2004).

In the present work, we investigated the hypothesis that deletion of one or more of the intron 4 motifs leads to the direct loss of *cis*-acting ISE or ISS and/or to a change of the secondary structure stability, resulting in altered SP-B ratio of completely to incompletely spliced RNA. Towards this, we studied the role of intron 4 motifs in a systematic way. First, SP-B intron 4 sequence (WT) was identified and used to generate SP-B minigene constructs (n=24) by removing motifs one by one from either the 5' or the 3' ends of the CA-repeat-rich region of intron 4. Next, the ratio of completely to incompletely spliced minigene RNAs was investigated by Northern blot and Real-time RT-PCR after transfection of cells. Motif 8 of intron 4 increased the splicing ratio, and interacted with nuclear *trans*-acting factors in a sequence-specific manner. Further *in silico* analysis of secondary structure stability of selected minigene RNAs (with and without motif 8) and comparison of secondary structures with their respective ratios indicated that both specific sequences and RNA stability may affect splicing of SP-B intron 4.

Materials and Methods

Oligonucleotides and constructs

All oligonucleotides used in the present study are listed in Table 1S.

WT intron 4 pGEM-T construct

First, because differences were reported within the motif-containing intron 4 sequence (Floros *et al.* 1995, Hamvas *et al.* 2005, Lin *et al.* 2005, Pilot-Matias *et al.* 1989), we sought to identify the most frequently found sequence, which we refer to as WT. A 547 bp fragment of the SP-B gene intron 4 region was amplified from genomic DNAs isolated from blood samples obtained from 21 individuals, including 11 patients with RDS

and 10 healthy control subjects that had been previously identified as having a complete intron 4 (without any naturally occurring deletions). Two independent PCRs were carried out for each sample, and the PCR products were cloned into the pGEM-T vector as described below. Reactions were carried out with 0.1 mM dNTPs, 50 ng of DNA, 40 ng of each primer, 1X buffer mix, 0.63 mM MgCl₂ and 1 U of high fidelity Taq polymerase (Roche, IN, USA). The PCR products (547 bp) included all 11 motifs and were generated by using primers, 2051 (starting at nt 2583 of human SP-B mRNA, GenBank: M24461.1, containing the *Nhe*I restriction site) and 2052 (starting at nt 3099, containing the *Afl*III restriction site) (Table 1S). The PCR cycles were: 95°C for 2 min, 35 cycles of 95°C for 30 s,

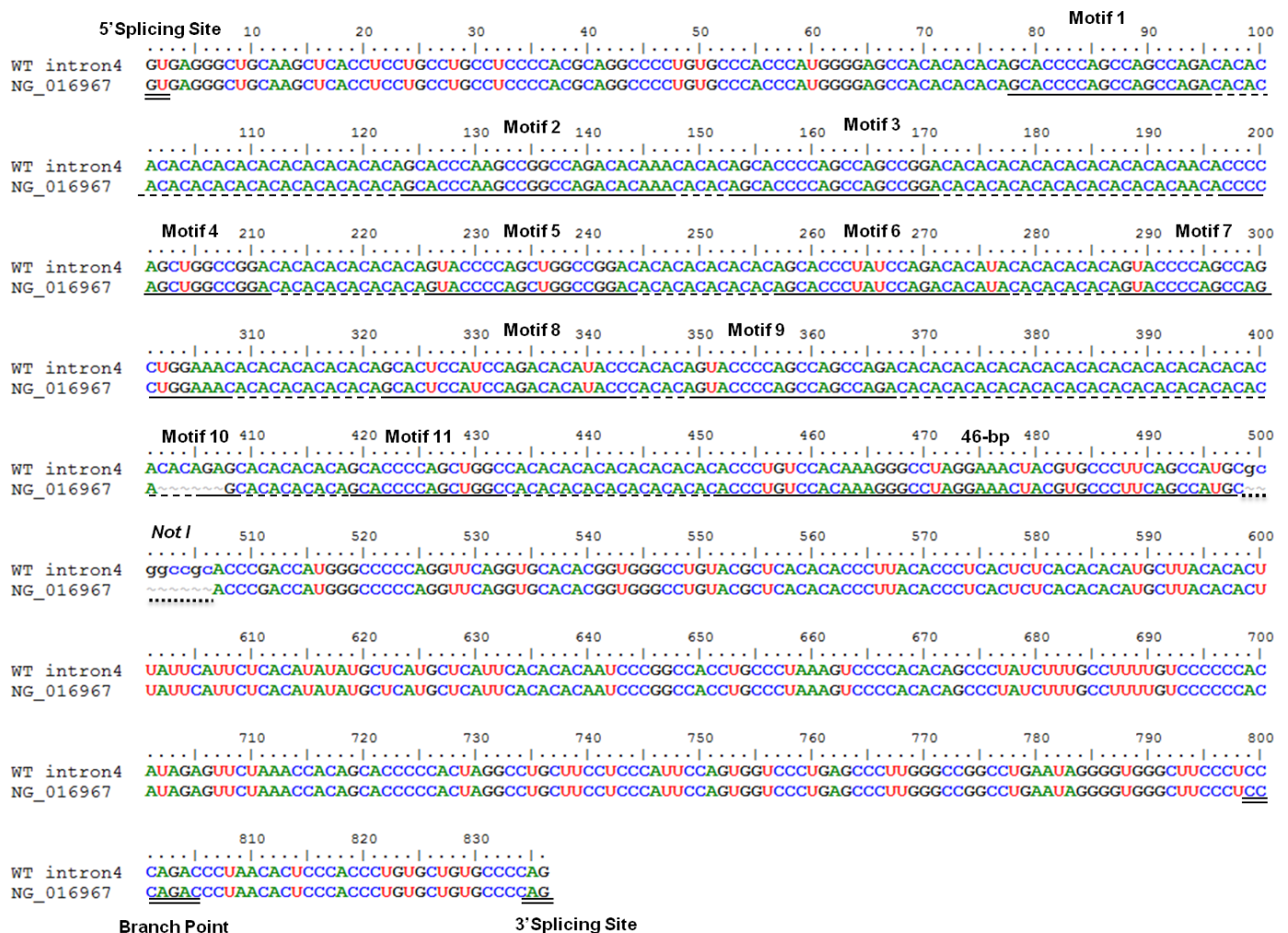


Figure 1. SP-B intron 4. Alignment of the human SP-B intron 4 sequence used in this study (top sequence) and the reference intron 4 sequence from GenBank (NG_016967). There are 836 nt in the SP-B intron 4 and the length variation polymorphism is localized approximately in the first half of the intron 4 sequence. This region is 78 nt away from the 5'ss in intron 4 and is characterized by the presence of 11 stretches of CA-repeat rich motifs. Each motif consists of two elements, a relatively conserved 18-20 bp sequence (solid underline), followed by a variable number of CA repeats (dashed underline). The 46-bp next to motif 11 is also underlined (solid underline). The 5' splicing site (5'ss), 3' splicing site (3'ss) and the branch point site are double underlined. The variant motifs are depicted in black as "motif 1" etc. on the top of their respective conservative sequence element. A *Not*I site (dotted line) and an *Eco*RI site (located between positions 78-79 of the intron 4 sequence, not shown) were inserted for cloning purposes.

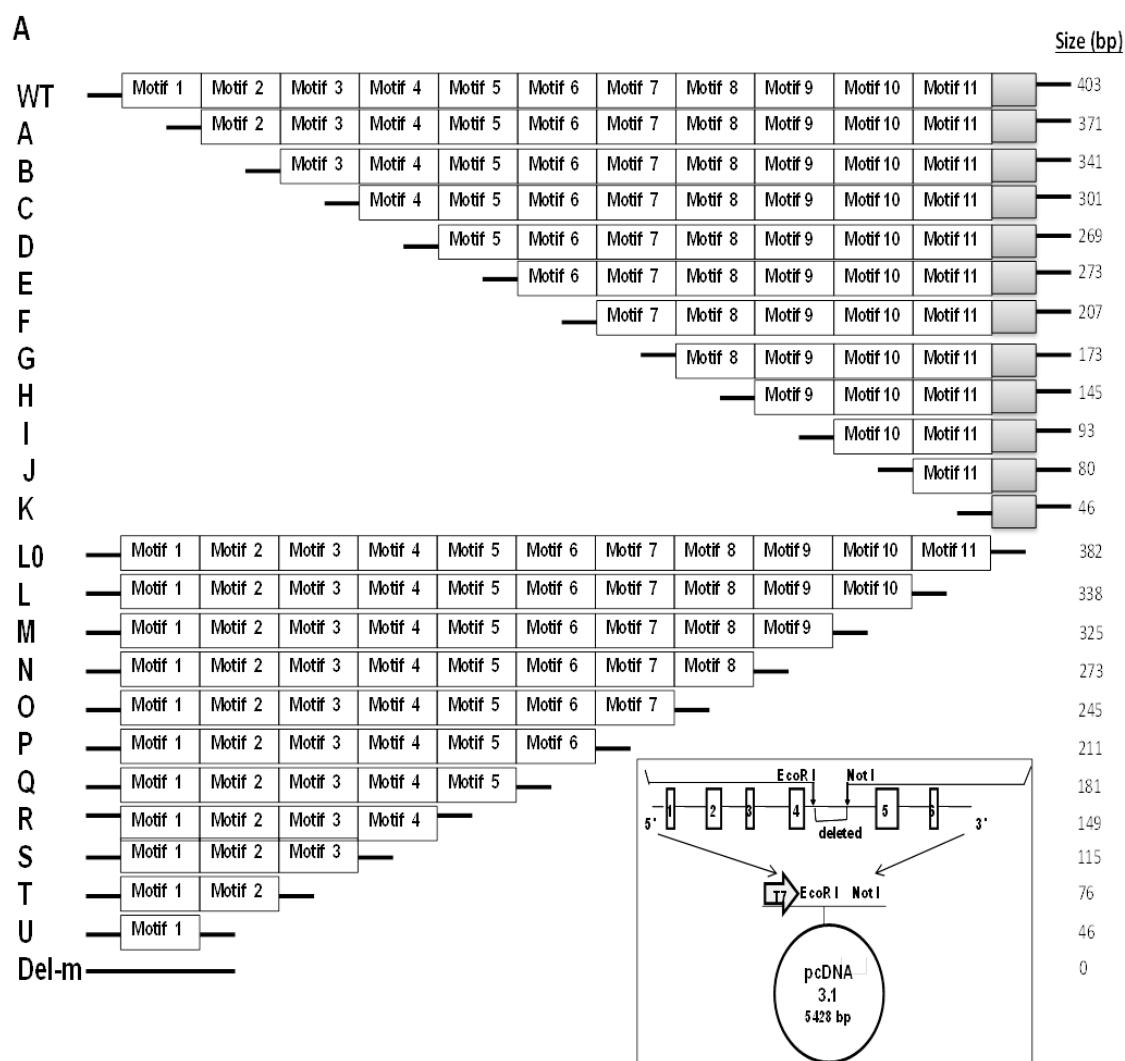


Figure 2. Description of the SP-B intron 4 and schematic representation of the intron 4 deletion constructs. Panel A depicts a schematic representation of the deletion constructs of intron 4 motifs. Two series of intron 4 deletion constructs were generated via systematic deletion of motifs either from the 5' end or the 3' end of the motif region. The WT contains all 11 motifs plus a 46 bp sequence (grey box) at the 3' end of motif 11. Compared to the WT, the sequence of motif 1 is deleted in construct A. In construct B, the sequence of motifs 1 and 2 are deleted, etc., as shown diagrammatically. The various intron 4 deletion fragments shown range in size from 46-417 nt. These were cloned in a pGEM-T vector, and then inserted between the *EcoRI* and *NotI* sites of pcDNA3.1 minigene Del-m construct, as shown in Panel A. The pcDNA3.1 minigene Del-m was obtained by modifying the previously generated pcDNA3.1 pBi4normal (Lin *et al.* 2005) as described in Materials and Methods. MCS upstream/downstream of the SP-B sequence were removed by site-directed mutagenesis and the *EcoRI* and *NotI* cloning sites were generated. *EcoRI* is located 78 nt from the end of exon 4 and *NotI* is located 502 nt upstream from the beginning of exon 5. Panel B depicts the variable region of the SP-B intron 4 WT sequence as identified in this study. This region has been characterized previously (Floros *et al.* 1995) and shown to consist of 11 motifs with each motif consisting of two elements, a relatively conserved sequence (bold) followed by a variable number of CA repeats (italics).

56°C for 60 s, 72°C for 1 min, and 72°C for 5 min.

The PCR product was separated on a 0.8% (w/v) agarose gel, purified and subjected to an overnight ligation at 16°C with the pGEM-T vector in a reaction containing 1X rapid ligation buffer, 50 ng of pGEM-T vector (Promega, WI, USA), 25 ng of purified PCR-product and 3 U of T4 DNA ligase. Transformation was carried out using JM109 *E. Coli* competent cells (Promega, WI, USA). Recombinant plasmid DNA was prepared using the Qiagen Miniprep kit (Qiagen, CA, USA) and digested with *AflIII* and *NheI*. A total of 57 clones were sequenced, and of these 39 (68%) had the same intron 4 sequence, referred here as intron 4 WT. These 39 clones were derived from 13 genomic samples (8 from RDS patients and 5 from controls). Seven of the remaining studied samples showed variation in the number of CA repeats in motifs 1 and 3, and one had a sequence identical to that of the reference sequence M24461. The WT SP-B intron 4 sequence was aligned and compared with the GenBank reference sequence (accession no. M24461) using the Clustal W multiple Alignment program from Bioedit version 7.0.5 (Figure 1).

pcDNA 3.1 SP-B minigene Del-m construct

A diagrammatic representation of the sequence of the 11 motifs and the deletion intron 4 constructs, along with the pcDNA 3.1 SP-B minigene Del-m, are shown in Figure 2. The pcDNA 3.1 SP-B minigene Del-m was generated by removing the 11 CA-repeat-rich motifs of SP-B intron 4 by modifying the parent plasmid pcDNA 3.1 pBi4normal (Lin *et al.* 2005) that contained the region starting at 8 bp upstream of the ATG start codon in exon 1 and ending at 433 bp downstream of exon 6. To remove the 11 motifs, the pBi4normal was modified as follows: First, MCS (multiple cloning sites) upstream/downstream of the SP-B sequence in pBi4normal were removed with primers 1816/1817 and 1818/1819 (Table 1S) by Site Directed Mutagenesis (Quick-changeII XL Kit, Agilent). Second, new restriction sites for *EcoRI* and *NotI* were created by Site Directed Mutagenesis (Quick-changeII XL Kit, Agilent) with primers 1811/1812. Third, the 11 motifs, including a 46-bp oligonucleotide in the 3' end of the eleventh motif, were sequentially removed. Of note, there is a 78 bp of intron 4 sequence between the 5' splicing site and motif 1 that contains a CA-repeat region between nucleotides 68-78 (Figure 1).

pcDNA 3.1 SP-B minigene constructs A-I & L0-T

All minigene constructs include the first 6 exons, 4 introns (1, 2, 3, 5) and part of intron 4, as shown in Figure 2A. A series of minigene constructs were made

by deletion of motifs from either the 5' or 3' end. All inserts, including the WT and the deletion fragments with different lengths of motifs of SP-B intron 4, were amplified using the pGEM-T WT intron 4 as a template; the primers used for the amplification are shown in Table 1S.

PCR was carried out in 1X High Fidelity Buffer, 2 mM MgSO₄, 0.2 mM of each dNTP, 100 ng of each primer, and 1 U of Platinum Taq High Fidelity (Invitrogen, CA, USA). To improve the specificity of amplification in order to generate different length variants, the following touchdown PCR program (Korbie & Mattick 2008) was used: 95°C for 2 min, 5 cycles of 95°C for 30 s, 70°C for 30 s and 68°C for 30 s, then 5 cycles of 95°C for 30 s, 68°C for 30 s and 68°C for 30 s, and then 5 cycles of 95°C for 30 s, 66°C for 30 s and 68°C for 30 s, followed by 25 cycles of 94°C for 30 s, 64°C for 30 s and 68°C for 30 s and a final extension step at 68°C for 5 min. All PCR fragments, ranging from 96 to 437 were separated on agarose gel, purified and cloned into a pGEM-T vector (Promega, WI, USA). The recombinant plasmid DNAs were digested, purified and sub-cloned into the *EcoRI* and *NotI* sites of pcDNA 3.1 minigene Del-m (Figure 2A, inset).

pcDNA 3.1 SP-B minigene K, U, motif 3, 4, 3+4, 8, 9, and 8+9 constructs

The constructs containing only motif 1 (minigene U), motif 3, motif 4, motifs 3+4, motif 8, motif 9, motifs 8+9 or the 46-bp sequence at the 3' end (minigene K) were generated by annealing two oligonucleotides and cloned into the *EcoRI* and *NotI* sites of pcDNA 3.1 minigene Del-m.

pGEM-11ZM8-D(CA) construct

To obtain motif 8 probes for RNA electromobility shift assays (REMSA), two oligos (2053/2054, Table 1S) were synthesized, annealed, and digested with *EcoRI* and *NotI*. The insert was cloned into the pGEM-11Z vector (Promega, Madison, WI, USA) to allow *in vitro* transcription under the T7 promoter. The CA-repeat-rich region was removed by site directed mutagenesis to obtain transcripts with the conserved sequence of motif 8 (i.e. without CA-repeats).

pGEM-11Z M8(G4)-D(CA) construct

A potential binding sequence element for SRp20 in motif 8 (CATC) was mutated, in the pGEM-11Z M8-D (CA) construct by site directed mutagenesis, using oligos 2061/2062 (Table 1S), and replaced by GGGG. The new construct was used as a template to generate negative control probes for REMSA.

Northern blot and RT-PCR related methods

Generation of probes

Two constructs were used to synthesize probes for Northern blot analysis: the pGEM-E3 that contained exon 3 sequences to detect total SP-B RNA and pGEM-I4 that contained intron 4 sequences to detect only incompletely spliced RNA or pre-mRNA. The probes exon 3 (probe-E3), consisting of 268 nucleotides of exon 3 (position 2060 of GenBank reference sequence M24461) and intron 4 (probe-I4), consisting of 275 nucleotides of intron 4 (position 3162 of reference sequence M24461) were generated by PCR using the WT minigene as template and primers 1836/1837 and 1858/1859, respectively. The PCR cycling conditions were 95°C for 2 min, 34 cycles of 95°C for 40 s, 58°C (probe-E3), or 61°C (probe-I4) for 30 s and 72°C for 30 s, followed by a final extension step at 72°C for 10 min. The PCR products were cloned into pGEM-T vector, resulting in pGEM-E3 and pGEM-I4. These were used to generate RNA probes (probe-E3, probe-I4) by *in vitro* transcription using a Northern blot starter kit (Roche, Indianapolis, IN, USA).

DNA Sequencing

All constructs were confirmed by sequence analysis using an Applied Biosystems ABI 377 DNA sequencer at the Molecular Genetics Core Facility, The Pennsylvania State University College of Medicine.

Cell Culture

The mammalian CHO (Chinese hamster ovary)-K1 cell line, purchased from the American Type Culture Collection (ATCC #CCL-61, VA, USA), was used for transfection. Cells were grown in GMEM (Glasgow Minimum Essential Medium; Invitrogen, Carlsbad, CA) supplemented with 10% (v/v) FBS (fetal bovine serum; Hyclone, UT, USA), 1X antimycotic/antibiotic solution (Sigma, MO, USA), 1x G+A (100× stock: 6 mg/ml L-glutamic acid and 6 mg/ml L-asparagine), 1X nucleoside mixture (Invitrogen, Carlsbad, CA) and 1X sodium pyruvate (Invitrogen, Carlsbad, CA), cultured at 37°C in a 5% (v/v) CO₂ atmosphere and subcultured weekly in 10-cm dishes to 50-80% confluence.

Transient transfection

CHO-K1 cells were used in a previous relevant work (Lin *et al.* 2005) and in this study. A preliminary experiment showed that no human SP-B mRNA homologue was detected by Northern blot and no PCR product was amplified by RT-PCR with human SP-B-specific primers from control CHO-K1 cells (data not shown). Cells were plated into 6-well culture plates until 50-80% confluence. A time course (5, 18, 24, 30,

36, 42 hr) with WT and minigenes L0 and N as well as a concentration course (0.5, 1, 2.5, 5.0, 10.0 µg) with minigene L0 and Del-m were carried out in CHO-K1 cells (Figure 1S) to identify optimal conditions for transfection of the constructs under study. The 30-36h time point and a concentration of 2.5 µg were chosen for further experimentation. The CHO-K1 cells were transfected with the Lipofectamine LTX Reagent kit (Invitrogen, Carlsbad, CA) following the manufacturer's protocol.

RNA Isolation

RNA was extracted using the RNA Bee solution (Tel-Test, FL, USA), precipitated in isopropyl alcohol, washed with 75% ethanol and resuspended in RNase-free water. Purity and concentration of extracted RNA was determined by Nanodrop 1000 spectrophotometer. Integrity of the RNA was also confirmed by gel electrophoresis by the presence of 28S and 18S ribosomal RNA bands.

Northern blot analysis

Samples of total RNA (5 µg from transfected cells), as well as RNA markers (Promega, WI, USA), were incubated with loading dye for 30 min at 65°C, and loaded on a 1% (w/v) agarose gel containing formaldehyde 37% (v/v). Samples were then separated by electrophoresis and transferred onto a positively charged nylon membrane (Roche, IN, USA). The blot was cross-linked using a UV Stratalink 2400 (Stratagene, CA, USA). The RNA marker (Promega, WI, USA) was used to estimate the size of SP-B transcript bands on the Northern blots. The gel and the membrane were photographed under UV light.

For the detection of completely and incompletely (containing intron 4 sequences) spliced RNA, DIG-labeled SP-B antisense RNA probe-E3 and probe-I4 were used, respectively. Hybridization was performed in DIG Easy Hyb Buffer (100 ng DIG-labeled RNA probe/ml) for 12 hours at 68°C. The blot was washed (DIG Wash and Block Buffer Set, Roche, IN, USA), bands were detected by exposure to film at RT, 1-5 min, and the band density was quantified by densitometry using a Molecular Dynamics model 100A Scanner (Amersham Pharmacia Biotech).

Real-time RT-PCR

All RNA samples were treated with DNA-free DNase Treatment & Removal Kit (Ambion, TX, USA). One step Real-time RT-PCR was performed using the Taqman One-Step RT-PCR Master Mix Reagents Kit (Applied Biosystems, NJ, USA). Completely spliced mRNA was detected with primers 1856/1857 and probe Vic1 [TaqMan VIC/minor groove binding

(MGB) probe] specific for the unique junction site of Exon 4 and Exon 5. For incompletely spliced RNA, we used primers 1854/1855 and probe Fam1 (TaqMan FAM/MGB probe) that specifically amplify the junction part of Intron 4 and Exon 5 (Table 1S). Reference samples for pre-mRNA and mRNA were plasmids pCDNA 3.1 SP-B WT minigene and pEE14-SP-B cDNA (Wang *et al.* 2003), respectively; these were serially diluted to generate standard curves. All reactions were carried out with 5 μ l of template RNA/DNA, 300 nM of each primer, 250 nM MGB TaqMan probe, 1 \times Master Mix (Applied Biosystems, NJ, USA), 1 \times MultiScribe and RNase inhibitor mix (Applied Biosystems, NJ, USA). The following Real-time RT-PCR thermal cycler conditions were used: 48°C for 30 min (cDNA synthesis) and 95°C for 10 min, followed by 42 cycles at 95°C for 15 s (denaturation) and 60°C for 1 min (annealing/extension). Reactions were performed on an ABI PRISM 7900 Sequence Detection System (Applied Biosystems, NJ, USA). The relative expression level for each construct was determined by the average value from three independent experiments. RNA target quantity was calculated by the Relative Standard Curve Methods (Applied Biosystems). Experiments were repeated 3 times (n=3).

REMSA- associated methods

SRp20 protein expression

In order to obtain nuclear extracts that contained human SRp20 protein for gel shift assays, we transiently expressed SRp20 in CHO-K1 cells. In brief, cells at 60% confluence were transfected with 6 μ g/dish (10cm) of a vector containing the human SRSF3 cDNA (*Homo sapiens* arginine-rich splicing factor 3 transcript variant 1) (Origene, Rockville MD, USA) that encodes SRp20 protein. Transfection was performed with the TurboFectin 8.0 (Origene, Rockville MD, USA) reagent diluted with Opti-MEM I (Invitrogen, Carlsbad, CA, USA), and transfected cells were incubated for 36 h at 37°C, as described above.

Nuclear Fractionation

Nuclear extracts from SRSF3 (SRp20) transfected and non-transfected CHO-K1 cells were obtained using the NE-PER Nuclear and Cytoplasmic Extraction kit (Thermo Scientific, Rockford IL, USA), after harvesting with PBS and cell scraping. Protein concentration was determined by the BCA Protein Assay (Thermo Scientific, Rockford IL, USA), following the manufacturer's instructions.

Western blot

Western blots were performed to confirm expression

of SRp20 in nuclear extracts. Four different antibodies (Abs) from two different vendors were tested. Ab 1 was a rabbit polyclonal Ab in which the immunogen was a synthetic peptide derived from an internal region of human SFRS3 (Abcam, Cambridge, MA, USA); Ab 2 was a mouse monoclonal Ab in which the immunogen was the amino acids 1-85 region of human SFRS3 (Abcam); Ab 3 was a rabbit polyclonal Ab in which the immunogen was the N terminal amino acids 35-84 region of human SFRS3 (Abcam); Ab 4 was a mouse monoclonal Ab in which the immunogen was the amino acids 1-85 region of human SFRS3 (Abnova, Walnut, CA, USA). Anti-SFRS3 primary Abs were used at 1:500 dilution, followed by incubation with goat anti-rabbit or anti-mouse IgG (H+L)-HRP conjugate (Biorad, Hercules CA, USA), 1:3000 dilution. HRP was detected by luminol-based chemiluminescence ECL Western Blotting Substrate (Thermo Scientific, Rockford IL, USA).

In vitro transcription

RNA probes for REMSA experiments were synthesized *in vitro* with the AmpliScribe T7-Flash Biotin-RNA Transcription Kit (Epicentre, Madison, WI, USA) using the plasmid M8-D(CA) and the mutated plasmid M8(G4)-D(CA). These were linearized by *NotI* and served as templates, following the manufacturer's protocol. In brief, a reaction containing 110 ng of linear plasmid DNA, 1X transcription buffer, 9 mM NTP / Biotin-UTP PreMix, 10 mM DTT, 0.5 μ l RiboGuard RNase Inhibitor, and 10 U of polymerase enzyme were incubated at 37°C for 1 h. One unit of DNase I (Epicentre, Madison, WI, USA) was then added and incubated at 37°C for another 15 min to remove the DNA template. Unlabeled probes were prepared likewise using 9 mM of non-biotinylated NTPs. RNA transcripts were purified using RNA-bee (TELTEST, Friendswood, TX, USA) in conjunction with Direct-zol RNA miniprep columns (Zymo Research, Irvine, CA, USA).

REMSA

Biotinylated RNA transcripts were heated at 80°C for 20 min and rapidly cooled on ice for 1 min prior to incubation with nuclear protein extracts. Each binding reaction contained 10 mM HEPES pH 7.3, 20 mM KCl, 1 mM MgCl₂, 1 mM DTT, 1% (v/v) glycerol, 40 U of RNaseOUT recombinant ribonuclease inhibitor (Invitrogen, Carlsbad, CA, USA), 2 μ g of tRNA as a non-specific competitor, 200 nM of *in vitro* transcribed biotinylated RNA, and 1 μ g of nuclear extract. Reaction mixtures were incubated at room temperature (RT) for 20 min, prior to loading on the gel. For competition shift assays, a 100- to 300-fold mass excess of

unlabeled *in vitro* M8-D (CA) and M8(G4)-D(CA) RNA transcripts were used to compete specific interactions between RNA and nuclear extracts. Unlabeled probes were incubated with protein extracts at RT for 10 min prior to the addition of biotinylated probe. The biotinylated M8 (G4)-D(CA) RNA probe was used as a negative control.

Samples were fractionated on native 6% polyacrylamide gels in 1X TBE and transferred to positively charged nylon membranes (Roche, Indianapolis, IN, USA) using a semidry Amersham Biosciences Nova blot apparatus in 1X TBE at 400mA for 30 min. Membranes were exposed to UV (254 nm) 120 mJ/cm² for 45 sec (Spectrolinker XL-1000 UV crosslinker, Spectronics Corporation, New Cassle, NY, USA) for crosslinking. Band shifts were detected using the Chemiluminescent Nucleic Acid Detection Module (Thermo Scientific, Rockford IL, USA).

For identification of SRp20:RNA M8-D(CA) interaction on REMSA, 1-2 µg of SRFS3 Ab were added to the binding reactions for supershift assays. Ab and nuclear extracts were incubated at RT for 20 min to form a protein-Ab complex which was then added to the RNA binding reaction. Four different Abs, as described above, from two different vendors were tested.

Statistical Analysis

All experiments were repeated at least three times as stated in the results section and triplicate assays were performed for each experimental sample. The data were analyzed by the SigmaStat software (version 3.5). Differences among groups were assessed by ANOVA or multiple-comparison ANOVA (Tukey's test). Results are expressed as means ±SE. Statistically significant differences were considered when $p < 0.05$.

In Silico Prediction Analysis

Branch point

To search for the predicted branch point of splicing sites of intron 4, we used an online program: the Human Splicing Finder website Version 2.4.1 (University of Montpellier, France, <http://www.umd.be/HSF/>) (Desmet *et al.* 2009), which was created to help the study of pre-mRNA splicing.

Prediction of RNA secondary structure

To study differences in secondary structure among SP-B intron 4 regions of all constructs, we used the RNAfold online program (<http://rna.tbi.univie.ac.at/>, Institute for Theoretical Chemistry, University of Vienna, Austria) (Mathews *et al.* 1999) to predict the secondary structures of intron 4 variants. This software fol-

lows the algorithm published by Zuker and Stiegler (Zuker & Stiegler 1981) to estimate the minimum total base pair distance of the RNA secondary structure and the minimum free energy at 37°C. The centroid secondary structure, with the best possible combination of paired bases, and the minimum free energy (dG) normalized for size by dividing dG by the total number of bases of each sequence under study, was obtained for the experimental intron 4 variant sequences.

Results

SP-B intron 4 motifs and minigene constructs

The sequence of intron 4 contains 11 motifs, each consisting of a relatively conserved sequence and a CA-repeat-rich region (Floros *et al.* 1995, Hamvas *et al.* 2005, Lin *et al.* 2005, Pilot-Matias *et al.* 1989). We sequenced 57 genomic clones from 21 individuals to identify the most frequently observed intron 4 sequence, which is referred to here as WT intron 4 sequence and was used in the present study. The WT sequence, when compared to the Genbank reference sequence (Figure 1), was nearly identical except that the Genbank sequence lacked 6 nucleotides in motif 10 (positions 402-408).

The Human consensus BP site, shown to be yUnAy, followed by poly pyrimidines tracts, was identified for intron 4 using the Human Splicing Finder program (Desmet *et al.* 2009). The underlined BP sequence (Figure 1) exhibited consensus values >70% (79.98%), and had a good position according to the AG dinucleotide exclusion zone (Gooding *et al.* 2006). The bold A is located 32 bp upstream of the 3' ss.

The intron 4 WT pGEM-T recombinant construct (see Materials and Methods) was used to generate a series of deletion constructs with motifs being removed one by one from either the 5' or the 3' end of the CA-repeat-rich region, as illustrated in Figure 2A, and cloned between the *EcoRI* and *NotI* sites of the Del-m minigene vector. The SP-B minigene Del-m (Figure 2A) that had all 11 motifs removed was generated as described in the Materials and Methods section. Twenty four SP-B minigene constructs containing a different number of motifs were studied. These were the WT SP-B minigene (normal-sized intron 4 with *EcoRI* and *NotI* sites) and the 23 SP-B minigene variant constructs (i.e. intron 4 deletions of varying size).

A schematic representation of the constructs and their missing motifs is shown in Figure 2A, and the sequence of each motif is shown in Figure 2B. The deletions were located within a 417 bp sequence of intron 4, found approximately in the first half of intron 4, nucleotides 79-496. This 417 bp region is a CA-repeat-rich region in which 65% of the total number of

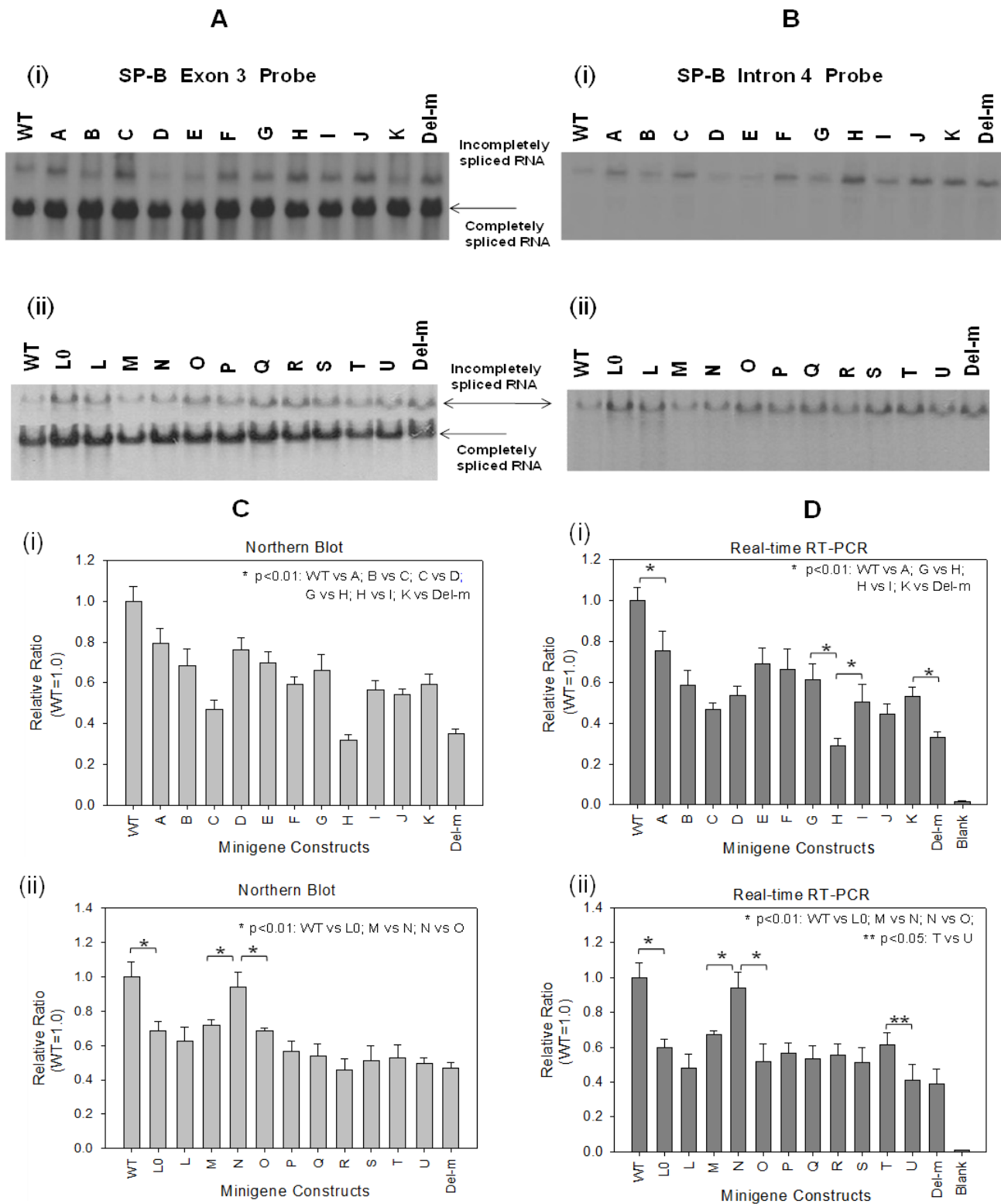


Figure 3. Northern blot and real-time RT-PCR analysis of SP-B minigene mRNAs. Panel A shows examples of Northern blot data with probe-E3, for constructs A-K (i) and L0-U (ii), respectively. Panel B shows examples of Northern blot data with probe-I4, for constructs A-K (i) and L0-U (ii), respectively. In addition to the deletion mutants, WT and Del-m were also included. Two bands were observed with probe-E3 (Panel A); a lower band with high intensity and an upper band with low intensity. The lower band represents normal mRNA and the upper band incompletely spliced RNA. With probe-I4, one band was observed (Panel B) representing incompletely spliced RNA. Panel C shows densitometric analyses of Northern Blot experiments for constructs A-K (i) and L0-U (ii), respectively. Panel D shows real-time RT-PCR results for constructs A-K (i) and L0-U (ii), respectively. The graphs show the ratios of normal mRNA/incompletely spliced RNA. The ratio of the WT minigene was set to 1.0 and all the relative ratios were evaluated based on the WT value (e.g. relative ratio of A in Northern Blot is 0.784), and the average \pm SE is shown. Both the amount of normal mRNA and of incompletely spliced RNA were determined as described in Materials and Methods. Experiments were repeated 3 times ($n=3$) and each sample was assayed in triplicate.

nucleotides are CAs. These CAs are clustered in 12 stretches of CA repeats. These included CA repeats within the 11 motifs and one within nucleotides 68–78, prior to the start of the 11 motifs. As shown in Figure 2B, motifs 1, 3, 9, and 11 had long CA repeats (10–17 CAs), and motifs 6, 7, and 8 had long conserved sequences (20–22 nt). In the other motifs (except motifs 10 and 11), the CA repeats were 18 nt long. Motif 10 had a truncated conserved sequence (3 nt) and motif 11 had a short sequence (14 nt).

Expression of SP-B minigenes in CHO-K1 cells

Northern blot analysis

Figure 3 depicts examples of Northern Blot analysis of all constructs. When the SP-B exon 3 probe (probe-E3) was used, all minigene constructs, including the WT, expressed a visible variable amount of incompletely spliced RNA in addition to normal mRNA bands [Figure 3A (i) and 3A (ii)]. When the probe containing only part of the second half of intron 4 sequence was used (probe-I4), only one band was observed, which had the same size as that of the incompletely spliced RNA [Figure 3B (i) and 3B (ii)]. To compare the efficiency of RNA splicing, we calculated the ratios of the amount of normal mRNA to incompletely spliced RNA (Figure 3C) for each minigene construct shown in Figure 2A. The ratio from the WT minigene was set to 1.0, and all the relative ratios were evaluated based on the WT value.

Although the total amount of DNA transfected and the RNA concentration loaded on the gel was the same as those of the WT, the relative ratios differed. Figures 3C (i) and 3C (ii) show that the relative ratio of minigene Del-m was the lowest (around 0.4), and all other relative ratios varied from 1.0 to 0.4, with the WT minigene expressing the highest relevant amount of completely spliced RNA. The ratios between WT and N had no significant difference [Figure 3C (ii)]. Several constructs exhibited levels similar to Del-m. These included minigenes C and H [Figure 3C (i)], and L0, L, M, O, P, Q, R, S, T, and U [Figure 3C (ii)]. Comparison of adjacent motifs showed that the ratios of minigene WT, B, G, K, or N, decreased sharply once motif 1, 3, 8, or the sequence (46 nt) at the 3' end of motif 11 was removed ($p < 0.01$), while the ratios of minigene C, H, or M, increased once motif 4 or 9 was removed ($p < 0.01$) [Figures 3C (i) and 3C (ii)].

Real-time RT-PCR analysis

To further assess expression and confirm Northern blot results, we used Real-time RT-PCR to quantify completely and incompletely spliced RNA [Figure 3D (i) & 3D (ii)]. The relative ratio of the WT minigene was set to 1.0, and with 2 exceptions, the RT-PCR ratios were similar to those from Northern blot [Figures 3C (i) and 3C (ii)]. Real-time RT-PCR showed that a) removal of motifs 3 & 4 did not change the ratio between minigene B/C and minigene C/D as shown by Northern blot analysis; and b) the ratio decreased

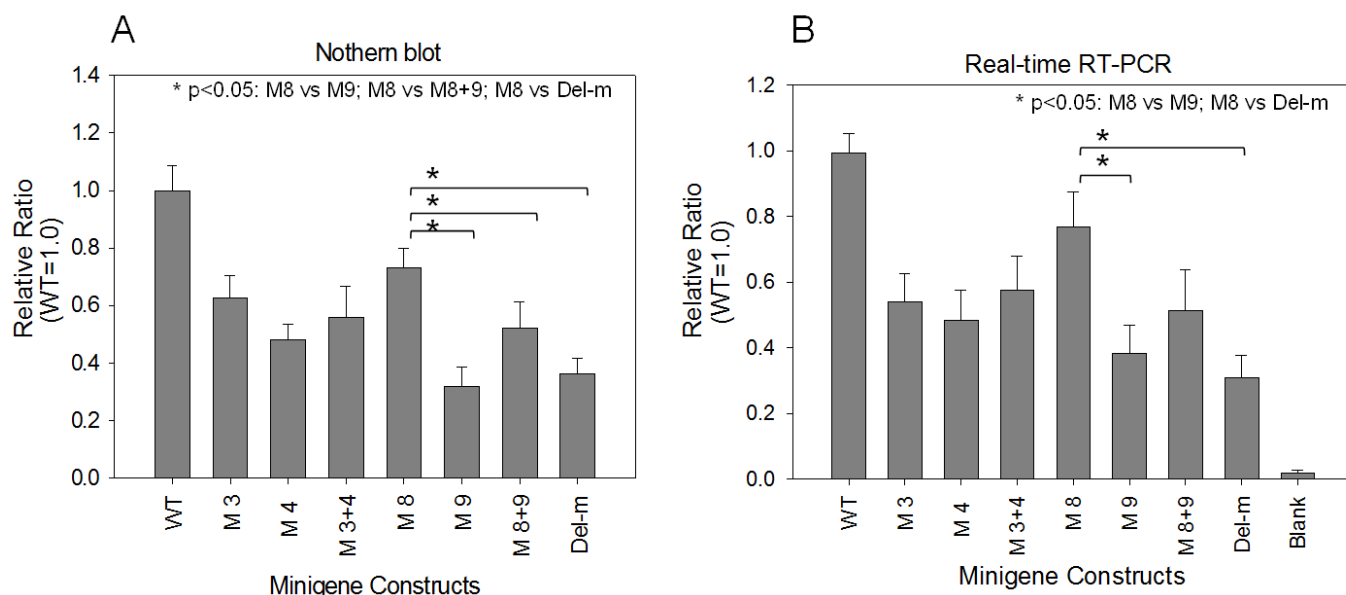


Figure 4. Northern blot and real-time RT-PCR analysis of selected motifs. Panels A and B show the results of Northern blot and real-time RT-PCR, respectively, for constructs containing selected motifs, M3, M4, M3+4, M8, M9, M8+9, including WT and Del-m. The ratios of normal mRNA/incompletely spliced RNA are shown. The amount of both normal mRNA and incompletely spliced RNA was determined as described in Materials and Methods. Experiments were repeated 3 times ($n=3$) and each sample was assayed in triplicate. Bars show mean \pm SE.

sharply from minigene T to U once motif 2 was removed ($p < 0.05$) while there was no difference in the ratios by Northern blot analysis [Figure 4C (ii)]. Removal of motif 8 from minigene G or N, and removal of the 46 nt sequence at the 3' end of motif 11, resulted in a significant decrease of the ratios, whereas removal of motif 9 from minigene H or M resulted in an increase of the ratio, as shown by Northern blot analysis.

Effect of selected motifs on splicing

Based on the experimental findings (Figure 3C, 3D), selected motifs were studied for their role on splicing, individually or in combination with their adjacent motif. Six constructs were generated as described in Materials and Methods, where each contained motifs 3, 4, 3+4, 8, 9, or 8+9. The impact of these motifs on splicing was studied by Northern blot analysis and Real-time RT-PCR. The rationale for studying motifs 8 and 9 was based on the observation that these consistently (by both Northern blot and Real-time RT-PCR) affected the ratio, as indicated by differences among constructs, G, H, and I, and of constructs M, N, and O. Motifs 3 and 4 on the other hand did not consistently

show significant differences between RT-PCR and Northern blot and were used here for comparison purposes.

Figures 4A and 4B depict the relative ratio of Northern blot analysis and of Real-time RT-PCR, respectively, of these minigene constructs. Consistent findings were observed by both methods (Figures 4A and 4B). Compared to the ratio of WT, all ratios, except for motif 8, exhibited significant differences ($p < 0.05$). Compared to the ratio of Del-m, the ratio of motif 8 increased ($p < 0.05$), and the ratio of motifs 9 and 8+9 did not change significantly, although significant changes ($p < 0.05$) were observed between motif 8 and motif 9. On the other hand, the ratio among motifs, 3, 4, and 3+4, did not show any significant changes by either method.

Motif 8 generated a sequence-specific electromobility shift with nuclear proteins; mutation of the CAUC→GGGG element abrogated the specific shift

Experimental evidence shown above identified motif 8 as an important regulatory sequence element. As proof

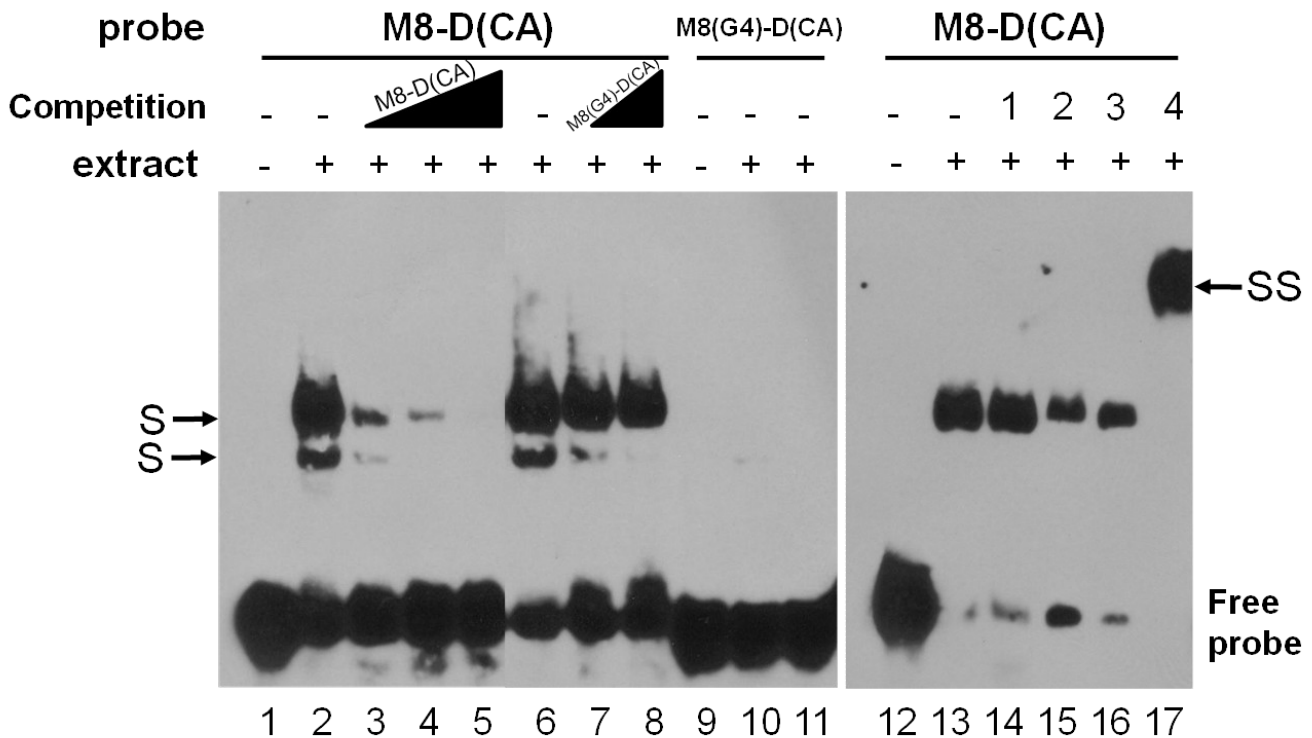


Figure 5. Motif 8 forms a sequence-specific shift with nuclear extracts. EMSA was performed to study specificity by incubating 200 nM of biotin-labeled RNA (obtained from construct M8-D(CA)) with 1 μ g of nuclear extracts obtained from CHO-K1 cells in which human SRP20 protein was overexpressed. Lanes 1 and 12 show free probe; lanes 2, 6, and 13 show specific shifts that were obtained by incubation of probe and nuclear extracts; lanes 3–5 show competition with 100, 200, and 300-fold molar excess of unlabeled M8-D(CA) RNA; lanes 7 and 8 show competition with 200 and 300-fold molar excess of unlabeled mutated M8(G4)-D(CA) RNA; lanes 10 and 11 show the lack of shifts when the mutated probe M8(G4)-D(CA) was incubated with nuclear extracts; lanes 14–16 show that inclusion of Abs 1, 2, and 3, respectively, does not abrogate the specific shift; lane 17 shows competition of the specific shift and an apparent supershift when M8-D(CA) RNA was incubated with nuclear extract and 1 μ g of Ab 4. Specific shifts (S), and super shift (SS) are labeled. Gel shift experiments were repeated 4 times.

of principle, we assessed the importance of motif 8, a) by carrying out shift assays to study whether sequence-specific RNA:protein complexes are formed, and b) by determining whether one of the binding proteins is SRp20. The rationale for the latter was based on the fact that motif 8 contains a sequence (CAUC) shown recently to be a potential binding site for SRp20 (SRFS3) (Ank \ddot{o} *et al.* 2012, Hargous *et al.* 2006). Although the RNA binding domain of SRp20 proteins is conserved among species for REMSA, we used CHO-K1 nuclear extracts (the cell line used for the transfection experiments described above) with and without transient expression of human SRp20 as described in Materials and Methods. Gel shift assays revealed the formation of specific RNA:protein complexes (Figure 5, lanes 2 and 6). The specificity of this interaction

was confirmed by a) competing with molar excess of unlabeled probe sequence (Figure 5, lanes 3-5), b) non-competition by a mutated probe (GGGG instead of CAUC) (Figure 5, lanes 7 and 8), and c) the absence of shift with the mutated probe (Figure 5, lanes 10 and 11). Nuclear extracts from CHO cells that were not transiently transfected with the human SRSF3 cDNA (SRp20) also formed specific RNA/protein complexes (not shown). However, we could not confirm experimentally whether the specific shift contains SRp20. Three different Abs targeting SRp20 (Abcam) did not disrupt the specific RNA:protein complex or form supershifts when these were included in the shift reaction assays (Figure 5 lanes 14-16). Although, a fourth SRp20 Ab (Abnova) both abrogated the formation of the complex and formed an apparent supershift (lane

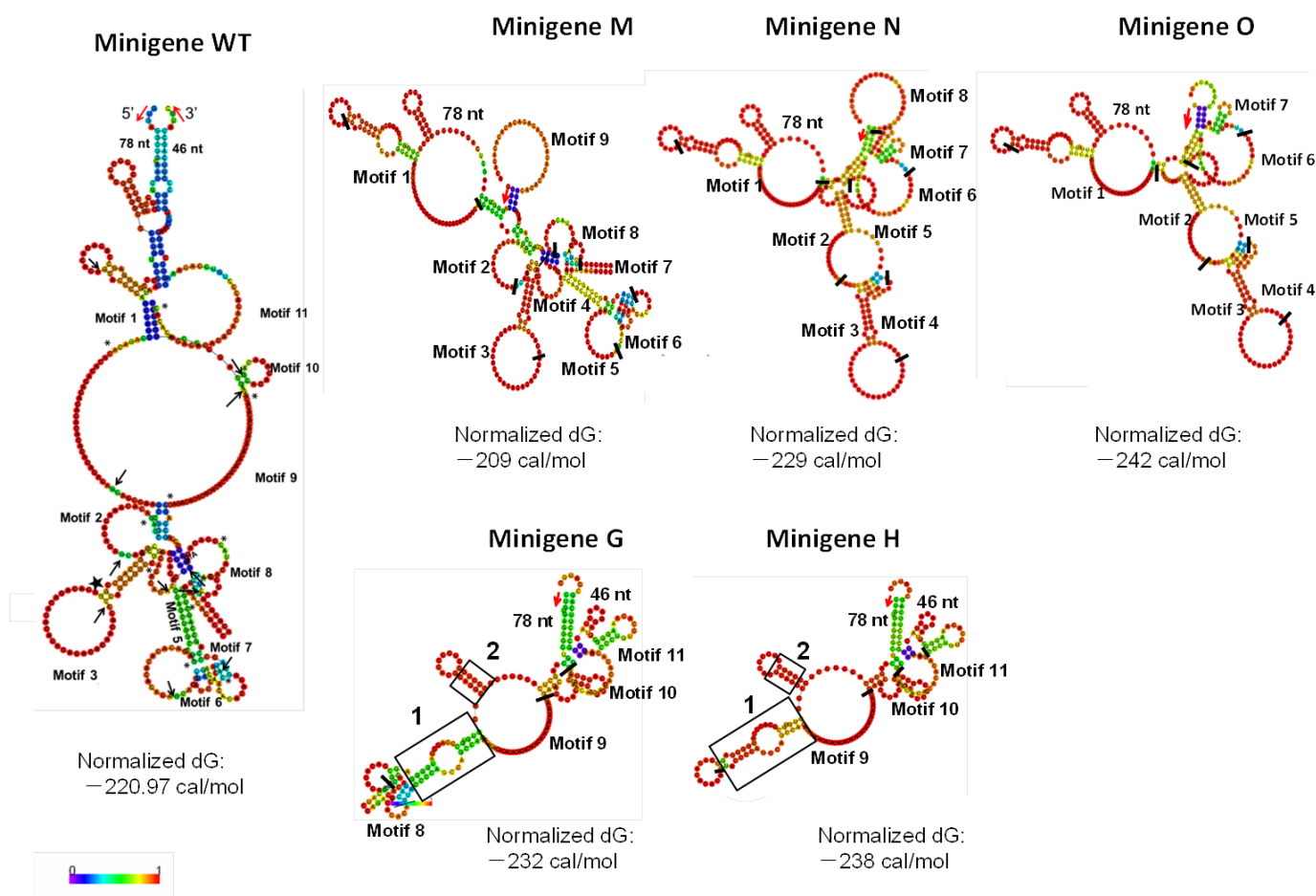


Figure 6. Predicted secondary structure for the WT intron 4 and of selected deletion intron 4 mutants. The RNA secondary structure of the WT minigene displayed the following characteristics: CA-repeat regions either individually, or with other CA rich regions, tended to form stable loops (red color). Among these, the CA repeats from motifs 1 & 9 formed the largest loop. On the contrary, the conservative sequences (GC rich) tended to form stems with varying stability. Among these, conserved sequences of motifs 3 and 4, as well as motifs 5 and 7, formed two long stems, respectively. The 5' and 3' ends are shown by red arrows (\rightarrow). The black arrow marks the beginning of each motif. The star (*) marks the beginning of each CA repeat within each motif. The normalized dG of WT is -220.97 cal/mol. The normalized dG for the secondary structure of each specific minigene is shown. Differences in the structural stability of stem 1 in minigenes G and H are marked by a box. The nucleotide region (78nt) prior to the start of intron 4 motifs and the nucleotide region (46nt) at the end of intron 4 motifs are noted along with the location of each motif. The color scale ranging from 0 (low stability) to 1 (high stability) is shown.

17), it is unclear whether the specificity of this Ab is for the bound proteins(s) or the probe, because a super-shift was observed when the probe and the Ab (in the absence of nuclear extract) were co-incubated (not shown). The specificity of the Abs was assessed by Western blot analysis. Abs 1, 3, and 4, but not Ab 2, identified by Western blot a specific nuclear protein with a size of that of the SRp20 protein (not shown). Also, mass spectroscopy analysis of the proteins in the specific shift did not identify SRp20, although other potential splicing factors were identified.

In silico analysis of secondary structures of constructs with and without motif 8

To further gain insight of the potential role of secondary structure stability in the observations made above, we analyzed the secondary structures of the WT minigene and of selected deletion mutants, using an online available program (RNAfold). dGs were normalized for length (i.e. by the number of nucleotides). The normalized dG for WT was -221 cal/mol and the normalized dGs of all other secondary structure constructs varied from -202 cal/mol (minigene B) to -258 cal/mol (minigene U) (Figure 6). Analysis of the WT minigene structure showed that the CA-repeat region, or CA cluster, tended to form a stable loop by itself, or with other CA rich regions. While the CA repeats of motif 1 and motif 9 formed the largest loop (Figure 6), the conserved sequences (GC rich) tended to form stems with varying stability. Among these, the conserved sequences of motifs 3 and 4, as well as motifs 5 and 7, formed, between themselves (3 with 4 and 5 with 7) two long stems with medium stability, and motifs 1 and 11 formed a stem with low stability. The intron 4 motif elements of the minigenes studied here also tended to exhibit the same structural characteristics as the WT, i.e. CA repeats formed loops and conserved sequences formed stems.

We next studied pairs of minigenes with and without motif 8. We found that: (i) comparison of secondary structure characteristics of minigenes, M, N, and O, revealed that the overall stability of N (dG -229 cal/mol) was similar to that of the WT (dG -221 cal/mol) whereas M (dG -209 cal/mol) and O (-242 cal/mol) were, respectively, less and more stable than N. Removal of motif 9 shown to have a negative impact on the ratio, resulted in a structure (minigene N) with stability similar to that of WT. However, removal of motif 8 (from minigene N) generated minigene O that had an even more stable structure (dG-242 cal/mol) than the WT. Thus, major changes in the overall structural stability in either direction (away from that observed in the WT) may negatively impact the ratio of completely to incompletely spliced RNA. (ii) Mini-

gene G that contained motif 8 and exhibited a higher ratio compared to minigene H [Figure 3C (i) and 3D (i)], revealed the following: although minigene H had a similar stability (dG -238 cal/mol) with minigene G (dG -232 cal/mol), a close inspection of their stems and loops showed that the “microstability” of a stem differed among these 2 minigenes, with minigene H showing higher microstability (color scale: 0: lower – 1: higher) in stem 1 (marked with a square) than G. Thus, in spite of a fairly similar overall dG, structural “microstability” of distinct elements within each RNA may affect differentially various downstream processes.

Discussion

A number of studies have shown associations of SP-B intron 4 variants and pulmonary disease in both infants and adults (Floros *et al.* 1995, Hamvas *et al.* 2005, Seifart *et al.* 2002b). Although this variation may alter mRNA splicing, RNA stability, or transcription factor binding sites necessary for normal gene expression, the details of these processes have not been defined (Lin *et al.* 2005, Rova *et al.* 2004). In this study, we investigated the hypothesis that one or more of the intron 4 motifs has a positive or a negative impact on SP-B RNA splicing. We found that: (i) virtually the entire SP-B intron 4 sequence was necessary for normal levels of RNA, although some motifs may be more important than others; (ii) motif 8 by itself, or with upstream motif sequences, exhibited a ratio of completely to incompletely spliced RNA comparable to that of the WT. However, motif 8, with downstream motif sequences that included motif 9, exhibited a significantly lower ratio than the WT and comparable to that when all intron 4 motifs were missing; (iii) motif 8 formed a sequence-specific shift, and mutation of a *cis*-element (CAUC) within motif 8 abrogated the specific RNA:protein interaction. Although the CAUC sequence had been shown previously to be a binding site for SRp20 (Ankő *et al.* 2012, Hargous *et al.* 2006), our present effort did not confirm this; (v) *in silico* analysis and comparison of secondary structure stability of intron 4 minigenes with or without motif 8 showed that secondary structure stability differences paralleled the splicing differences observed. Thus, both specific sequences and secondary structure stability appear to be important for correct splicing.

It has been proposed that pre-mRNA secondary structures can influence splicing activity (Buratti & Baralle 2004). Numerous studies have shown that RNA secondary structure masks or exposes splice regulatory elements, or modifies the distance between these elements (Hui 2009). For example, mutations in

the tau gene intronic region near the 5' splice site of exon 10 correlated closely with alterations in a characteristic stem-loop structure as determined by nuclear magnetic resonance spectroscopy (Buratti & Baralle 2004, Varani *et al.* 1999). Thus, the predictive ability in the search for novel RNA binding targets of proteins can be greatly enhanced if the secondary structure is taken into consideration.

Removal of motif 8 from minigene N had a negative effect on the ratio indicating that motif 8 positively affects splicing in the presence of upstream intron 4 motifs. Because motif 8 by itself also had a positive effect, we hypothesized that motif 8 possesses important *cis*-acting splice enhancers. Indeed, motif 8 was found to contain a unique "CAUC" sequence, a potential binding site for SRp20, leading us to further hypothesize that SRp20, or an unknown protein, binds the conservative sequence of motif 8. A sequence-specific RNA:protein interaction was consequently found between the conserved sequence of motif 8 and nuclear protein(s), supporting the hypothesis that the decrease in the ratio upon removal of motif 8 directly comes from the loss of a binding site for critical regulatory protein(s). Moreover, the thermal stability of minigene N (that contains motif 8) was similar to the WT (albeit slightly higher than WT) as was the ratio of completely to incompletely spliced RNA. Together, these indicate that the secondary structure, along with specific *cis* elements, may be important for splicing to occur properly.

However, when motif 8 was used in the presence of downstream motifs (especially motif 9) the ratio decreased, indicating that motif 9 had a dominant negative effect on splicing. This observation was further confirmed when motif 9 was used by itself, indicating that competing mechanisms (both positive and negative) may be at play for normal splicing. The underlying mechanisms may involve cooperative and/or antagonistic effects among regulatory proteins and binding for specific *cis*-elements. The final effect would probably rely on the group of factors whose function prevails within a given cellular context. Furthermore, comparison of the overall thermal stability of minigenes with (minigene G) and without (minigene H) motif 8 and downstream motifs, showed that removal of motif 8 resulted in slightly higher stability, which was considerably higher than the WT (WT dG: -220.97 cal/mol; minigene H dG: -238 cal/mol). A higher micro-stability of a stem loop structure that may contribute to the overall higher stability that was observed in minigene H in the absence of motif 8. We speculated that a certain structural fluidity is necessary and that a change in the overall stability in either direction (away from the WT) may have a negative impact

on splicing. It is also possible that deletion of certain motifs affects RNA stability itself. Future experiments are warranted to understand the underlying mechanisms of intron 4-mediated processes.

The available literature indicates that splicing is a complex process. The present findings indicate that SP-B splicing is indeed a complex and, most likely, a highly coordinated process that relies on the combination of several factors, including structural stability, and intronic sequence elements for proper control. Mutations that disrupt any of these critical features may alter production and/or function of the encoded proteins by affecting splicing and/or RNA stability. Several naturally occurring deletion and insertion variants have been reported for intron 4, and these may result in the retention of part of intron 4 (Floros *et al.* 1995, Hamvas *et al.* 2005, Lin *et al.* 2005), as shown in the present study with the mutants where intron 4 motifs were systematically removed. The incompletely spliced RNA may generate a new start or termination codon, or lead to codon frame shifts. In this scenario, because the normal mature SP-B protein is encoded by exons 6 and 7, the total amount of normal SP-B protein could decrease. Moreover, even if translation could be activated from an AUG within intron 4 in an IRES-dependent way, generation of abnormal transcripts could not be excluded.

In summary, the present findings show that: (i) motifs within the SP-B intron 4 can positively or negatively affect SP-B splicing and/or RNA stability; motifs 8 and 9 in particular may contain ISE and ISS, respectively; (ii) nuclear proteins interact in a sequence-specific manner with the CAUC element within motif 8 and this interaction may be important for the positive effect of motif 8 on splicing. Future experiments are required to identify the binding factors involved in this regulation and study how RNA:protein interactions can affect SP-B intron RNA splicing and/or stability. Another major future step is to develop functional experimental systems suitable for the interaction of *cis* elements and *trans* factors within SP-B intron 4.

Acknowledgments

The authors thank Joseph Bednarczyk for expert technical support, Gina Marcucci and Lynelle Murray for typing, and the Pennsylvania State University College of Medicine core facility for DNA sequencing, oligonucleotide synthesis and mass spectroscopy. This work was supported by the NIH HL-34788 grant. LN is the recipient of a fellowship from the Ministry of Education of China.

Conflicts of Interest

No conflicts of interest, financial or otherwise are declared by the author(s).

Author contributions

WY and LN equally contributed to this work by generating minigene constructs, performed Northern blot and real time PCR experiments, and contributed to the synthesis of the data and writing of the manuscript. PS contributed with experimental design, sequence alignments, interpretation and synthesis of data, and manuscript writing. GW contributed with experimental design, generation of the SP-B minigene Del-m construct, and interpretation of the initial data. GN performed shift assays and western blots for SRp20 and contributed to manuscript writing. AS sequenced genomic DNA samples, identified the WT intron 4 sequence and wrote the relevant methods. OS generated some of the intron 4 constructs. SD generated several intron 4 constructs and cloned the minigene constructs. MR provided input on RNA secondary structure predictions. JF provided oversight with all aspects of the project, including the overall experimental design, data interpretation, synthesis and writing of the manuscript.

References

Ankö ML, Müller-McNicoll M, Brandl H, Curk T, Gorup C, Henry I, Ule J & Neugebauer KM 2012 The RNA-binding landscapes of two SR proteins reveal unique functions and binding to diverse RNA classes. *Genome Biol* **13** R17.

Beers MF, Atochina EN, Preston AM & Beck JM 1999 Inhibition of lung surfactant protein B expression during *Pneumocystis carinii* pneumonia in mice. *J Lab Clin Med* **133** 423-433.

Buratti E & Baralle FE 2004 Influence of RNA secondary structure on the pre-mRNA splicing process. *Mol Cell Biol* **24** 10505-10514.

Desmet FO, Hamroun D, Lalande M, Collod-Bérout G, Claustres M & Bérout C 2009 Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res* **37** e67.

Ewis AA, Kondo K, Dang F, Nakahori Y, Shinohara Y, Ishikawa M & Baba Y 2006 Surfactant protein B gene variations and susceptibility to lung cancer in chromate workers. *Am J Ind Med* **49** 367-373.

Floros J & Fan R 2001 Surfactant protein A and B genetic variants and respiratory distress syndrome: allele interactions. *Biol Neonate* **80** 22-25.

Floros J, Fan R, Diangelo S, Guo X, Wert J & Luo J 2001 Surfactant protein (SP) B associations and inter-

actions with SP-A in white and black subjects with respiratory distress syndrome. *Pediatrics international : official journal of the Japan Pediatric Society* **43** 567-576.

Floros J, Veletza SV, Kotikalapudi P, Krizkova L, Karinch AM, Friedman C, Buchter S & Marks K 1995 Dinucleotide repeats in the human surfactant protein-B gene and respiratory-distress syndrome. *Biochem J* **305** 583-590.

Gooding C, Clark F, Wollerton MC, Grellscheid SN, Groom H & Smith CW 2006 A class of human exons with predicted distant branch points revealed by analysis of AG dinucleotide exclusion zones. *Genome Biol* **7** R1.

Gower WA & Noguee LM 2011 Surfactant dysfunction. *Paediatr Respir Rev* **12** 223-229.

Guttentag SH, Beers MF, Bieler BM & Ballard PL 1998 Surfactant protein B processing in human fetal lung. *Am J Physiol* **275** L559-566.

Hallman M, Haataja R & Marttila R 2002 Surfactant proteins and genetic predisposition to respiratory distress syndrome. *Semin Perinatol* **26** 450-460.

Hamvas A, Wegner DJ, Trusgnich MA, Madden K, Heins H, Liu Y, Rice T, An P, Watkins-Torry J & Cole FS 2005 Genetic variant characterization in intron 4 of the surfactant protein B gene. *Hum Mutat* **26** 494-495.

Hargous Y, Hautbergue GM, Tintaru AM, Skrisovska L, Golovanov AP, Stevenin J, Lian LY, Wilson SA & Allain FH 2006 Molecular basis of RNA recognition and TAP binding by the SR proteins SRp20 and 9G8. *EMBO J* **25** 5126-5137.

Hui J 2009 Regulation of mammalian pre-mRNA splicing. *Sci China C Life Sci* **52** 253-260.

Hui J, Hung LH, Heiner M, Schreiner S, Neumuller N, Reither G, Haas SA & Bindereif A 2005 Intronic CA-repeat and CA-rich elements: a new class of regulators of mammalian alternative splicing. *EMBO J* **24** 1988-1998.

Hung LH, Heiner M, Hui J, Schreiner S, Benes V & Bindereif A 2008 Diverse roles of hnRNP L in mammalian mRNA processing: a combined microarray and RNAi analysis. *RNA* **14** 284-296.

Kala P, Ten Have T, Nielsen H, Dunn M & Floros J 1998 Association of pulmonary surfactant protein A (SP-A) gene and respiratory distress syndrome: interaction with SP-B. *Pediatr Res* **43** 169-177.

Kerr MH & Paton JY 1999 Surfactant protein levels in severe respiratory syncytial virus infection. *Am J Respir Crit Care Med* **159** 1115-1118.

Korbie DJ & Mattick JS 2008 Touchdown PCR for increased specificity and sensitivity in PCR amplification. *Nat Protoc* **3** 1452-1456.

Lin Z, Thomas NJ, Wang Y, Guo X, Seifart C, Shaikoor H & Floros J 2005 Deletions within a CA-repeat-

- rich region of intron 4 of the human SP-B gene affect mRNA splicing. *Biochem J* **389** 403-412.
- Mathews DH, Sabina J, Zuker M & Turner DH 1999 Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol* **288** 911-940.
- Melo KF, Martin RM, Costa EM, Carvalho FM, Jorge AA, Arnhold IJ & Mendonca BB 2002 An unusual phenotype of Frasier syndrome due to IVS9 +4C>T mutation in the WT1 gene: predominantly male ambiguous genitalia and absence of gonadal dysgenesis. *J Clin Endocrinol Metab* **87** 2500-2505.
- Neklason DW, Solomon CH, Dalton AL, Kuwada SK & Burt RW 2004 Intron 4 mutation in APC gene results in splice defect and attenuated FAP phenotype. *Fam Cancer* **3** 35-40.
- Nesslein LL, Melton KR, Ikegami M, Na CL, Wert SE, Rice WR, Whitsett JA & Weaver TE 2005 Partial SP-B deficiency perturbs lung function and causes air space abnormalities. *Am J Physiol Lung Cell Mol Physiol* **288** L1154-1161.
- O'Reilly MA, Weaver TE, Pilot-Matias TJ, Sarin VK, Gazdar AF & Whitsett JA 1989 In vitro translation, post-translational processing and secretion of pulmonary surfactant protein B precursors. *Biochim Biophys Acta* **1011** 140-148.
- Pastor T, Talotti G, Lewandowska MA & Pagani F 2009 An Alu-derived intronic splicing enhancer facilitates intronic processing and modulates aberrant splicing in ATM. *Nucleic Acids Res* **37** 7258-7267.
- Pilot-Matias TJ, Kister SE, Fox JL, Kropp K, Glasser SW & Whitsett JA 1989 Structure and organization of the gene encoding human pulmonary surfactant proteolipid SP-B. *DNA* **8** 75-86.
- Raponi M, Upadhyaya M & Baralle D 2006 Functional splicing assay shows a pathogenic intronic mutation in neurofibromatosis type 1 (NF1) due to intronic sequence exonization. *Hum Mutat* **27** 294-295.
- Rova M, Haataja R, Marttila R, Ollikainen V, Tamela O & Hallman M 2004 Data mining and multi-parameter analysis of lung surfactant protein genes in bronchopulmonary dysplasia. *Hum Mol Genet* **13** 1095-1104.
- Seifart C, Plagens A, Brödje D, Müller B, von Wichert P & Floros J 2002a Surfactant protein B intron 4 variation in German patients with COPD and acute respiratory failure. *Dis Markers* **18** 129-136.
- Seifart C, Seifart U, Plagens A, Wolf M & von Wichert P 2002b Surfactant protein B gene variations enhance susceptibility to squamous cell carcinoma of the lung in German patients. *Br J Cancer* **87** 212-217.
- Serrano AG, Ryan M, Weaver TE & Pérez-Gil J 2006 Critical structure-function determinants within the N-terminal region of pulmonary surfactant protein SP-B. *Biophys J* **90** 238-249.
- Slaughaupt SA, Blumenfeld A, Gill SP, Leyne M, Mull J, Cuajungco MP, Liebert CB, Chadwick B, Idelson M, Reznik L, Robbins C, Makalowska I, Brownstein M, Krappmann D, Scheidereit C, Maayan C, Axelrod FB & Gusella JF 2001 Tissue-specific expression of a splicing mutation in the IKBKAP gene causes familial dysautonomia. *Am J Hum Genet* **68** 598-605.
- Tokieda K, Iwamoto HS, Bachurski C, Wert SE, Hull WM, Ikeda K & Whitsett JA 1999 Surfactant protein-B-deficient mice are susceptible to hyperoxic lung injury. *Am J Respir Cell Mol Biol* **21** 463-472.
- Tokieda K, Whitsett JA, Clark JC, Weaver TE, Ikeda K, McConnell KB, Jobe AH, Ikegami M & Iwamoto HS 1997 Pulmonary dysfunction in neonatal SP-B-deficient mice. *Am J Physiol* **273** L875-882.
- Tuohy TM, Done MW, Lewandowski MS, Shires PM, Saraiya DS, Huang SC, Neklason DW & Burt RW 2010 Large intron 14 rearrangement in APC results in splice defect and attenuated FAP. *Hum Genet* **127** 359-369.
- Varani L, Hasegawa M, Spillantini MG, Smith MJ, Murrell JR, Ghetti B, Klug A, Goedert M & Varani G 1999 Structure of tau exon 10 splicing regulatory element RNA and destabilization by mutations of frontotemporal dementia and parkinsonism linked to chromosome 17. *Proc Natl Acad Sci U S A* **96** 8229-8234.
- Wang G, Christensen ND, Wigdahl B, Guttentag SH & Floros J 2003 Differences in N-linked glycosylation between human surfactant protein-B variants of the C or T allele at the single-nucleotide polymorphism at position 1580: implications for disease. *Biochem J* **369** 179-184.
- Ward AJ & Cooper TA 2010 The pathobiology of splicing. *J Pathol* **220** 152-163.
- Weaver TE 1998 Synthesis, processing and secretion of surfactant proteins B and C. *Biochim Biophys Acta* **1408** 173-179.
- Weaver TE, Na CL & Stahlman M 2002 Biogenesis of lamellar bodies, lysosome-related organelles involved in storage and secretion of pulmonary surfactant. *Semin Cell Dev Biol* **13** 263-270.
- Wert SE, Whitsett JA & Noguee LM 2009 Genetic disorders of surfactant dysfunction. *Pediatr Dev Pathol* **12** 253-274.
- Zuker M & Stiegler P 1981 Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res* **9** 133-148.