 Human Molecular Genetics, 2000, Vol. 9, No. 2 **259–265**

# An exonic enhancer is required for inclusion of an essential exon in the SMA-determining gene *SMN*

**Christian L. Lorson[1] and Elliot J. Androphy[1,2,+]**

[1]Department of Dermatology, New England Medical Center and Tufts University School of Medicine, Boston, MA 02111, USA and [2]Department of Molecular Biology and Microbiology, Tufts University School of Medicine, Boston, MA 02111, USA

Received 15 September 1999; Revised and Accepted 5 November 1999

**The *survival motor neuron* genes, *SMN1* and *SMN2*, encode identical proteins; however, only homozygous loss of *SMN1* correlates with the development of spinal muscular atrophy (SMA). We have previously shown that a single non-polymorphic nucleotide difference in *SMN* exon 7 dramatically affects *SMN* mRNA processing. *SMN1* primarily produces a full-length RNA whereas *SMN2* expresses dramatically reduced full-length RNA and abundant levels of an aberrantly spliced transcript lacking exon 7. The importance of proper exon 7 processing has been underscored by the identification of several mutations within splice sites adjacent to exon 7. Here we show that an AG-rich exonic splice enhancer (ESE) in the center of *SMN* exon 7 is required for inclusion of exon 7. This region functioned as an ESE in a heterologous context, supporting efficient *in vitro* splicing of the *Drosophila double-sex* gene. Finally, the protein encoded by the exon-skipping event, Δ7, was less stable than full-length SMN, providing additional evidence of why *SMN2* fails to compensate for the loss of *SMN1* and leads to the development of SMA.**

## INTRODUCTION

Spinal muscular atrophy (SMA) is a common autosomal recessive disease characterized by degeneration of α-motor neurons. Two nearly identical *survival of motor neuron* (*SMN*) genes are typically present on chromosome 5q13: *SMN1* and *SMN2* (1,2). Only homozygous loss of *SMN1* correlates with SMA development (1). These two *SMN* genes are >99% identical at the nucleotide level and encode SMN (1,3), a 294 amino acid RNA-binding protein (4). SMN has been shown to play a role in the biogenesis of small ribonuclear proteins (snRNPs) through association with SMN-interacting protein 1 (SIP-1) and Sm proteins (5,6), and enhances *in vitro* splicing, potentially as a regenerative factor for the general splicing machinery (7).

We have shown that development of SMA correlates with aberrant splicing of *SMN* exon 7 (8). *SMN1* mRNA predominantly expresses a full-length transcript whereas *SMN2* produces low levels of full-length RNA and an aberrantly spliced product lacking exon 7, termed Δ7 (1,9). SMNΔ7 protein, predicted to be slightly smaller than full-length SMN, has not been detected by western analysis, and has therefore been presumed to be less stable. We have reported that a single nucleotide difference between *SMN1* and *SMN2* within exon 7 (exon 7+6, where +6 refers to the sixth nucleotide in *SMN* exon 7) is responsible for their splicing patterns and the subsequent development of SMA (8). Proper mRNA processing of this exon is critical since several mutations in the splice sites adjacent to exon 7 have been isolated from SMA patients and SMA hybrid genes exhibit high levels of aberrant mRNA processing in this region (1,10). Little is known concerning the processing of *SMN* exon 7 or the functional significance of the resultant Δ7 protein in the development of SMA. The exon 7+6 nucleotide was predicted to have disrupted an exonic splicing enhancer (ESE); however, this activity has yet to be demonstrated. Elucidating the processing mechanism(s) regulating *SMN* full-length expression relative to the exon-skipped mRNA is an important step in furthering the molecular pathogenesis of SMA.

Pre-mRNA processing is a critical step in eukaryotic gene expression (11,12). The accurate identification and excision of intron sequences and the subsequent ligation of 5′ and 3′ splice sites is carried out by a large complex consisting of U1, U2, U4/U6 and U5 snRNPs and a group of RNA-binding factors termed SR proteins (13,14). SR proteins typically contain two critical regions: an arginine/serine-rich (RS) domain (13,14) and at least one RNA recognition motif (RRM), which function to mediate protein–protein interactions and RNA binding, respectively (15,16). Processing of constitutive and regulated pre-mRNAs requires SR proteins (17–19). ESEs are found in constitutive and regulated exons and serve as binding sites for SR proteins. Two general classes of binding motif have been identified: a purine-rich enhancer consisting of GAR repeats (20–25) (where R represents either G or A) and CA-rich motifs, termed ACEs (22,26). Through a series of protein–protein interactions, SR factors bound to ESEs are generally believed to function by recruiting the splicing machinery and promoting the cooperative assembly of competent splicing complexes (17,22–24,27,28). Another function of ESEs is to compensate for suboptimal 3′ splice sites, thereby promoting inclusion of an otherwise poorly defined exon (29,30). Potentially, one or more ESEs are operative in the constitutive inclu-

[+]To whom correspondence should be addressed. Tel: +1 617 636 1493; Fax: +1 617 636 6190; Email: eandroph@opal.tufts.edu

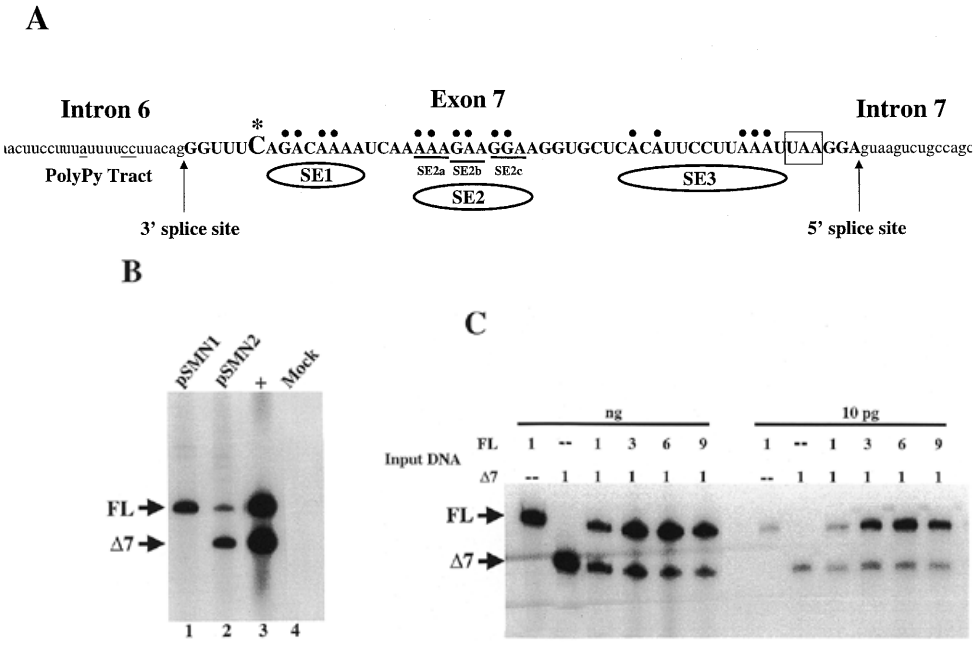


**Figure 1.** (**A**) RNA sequence of the *SMN* exon 7 region indicating exon 7 sequences (bold upper case) and intron sequences (lower case). Non-polymorphic *SMN1* and *SMN2* exon 7+6 position (large bold and asterisk), splice signals, SMN stop codon (boxed) and mutations used in Figures 1–3 are indicated. Base pairs indicated with dots and underlining were mutated to U residues in the identified regions [SE1, SE2(a–c), SE3]. (**B**) [$^{32}$P]dATP-labeled RT–PCR. Thirty-six hours post-$CaPO_4$ transfection with *pSMN1* and *pSMN2*, or mock transfection of U2OS cells, total RNA was isolated and used to synthesize oligo-dT anchored cDNAs that were amplified by 15 cycles [94°C for 1 min; (94°C for 30 s; 56°C for 30 s; 72°C for 1.5 min) × 15; 72°C for 10 min] using pCI Fwd#2 and pCI Rev primers. Cloned full-length and Δ7 cDNA served as a PCR positive control ('+' lane). Full-length (FL) and the exon 7 skipped (Δ7) products are indicated by arrows. (**C**) [$^{32}$P]dATP-labeled competitive PCR [see (B)] with cloned FL and Δ7 cDNA using the indicated nanogram (ng) or picogram (pg) amounts. Quantitations of full-length:Δ7 ratios (predicted/actual) with standard deviations are as follows: ng, 1:1/1:1 ± 0.01; 3:1/3.2:1 ± 0.3; 6:1/6.8:1 ± 0.4; 9:1/9.7:1 ± 0.4; pg, 1:1/.9:1 ± 0.1; 3:1/2.7:1 ± 0.2; 6:1/6.1:1 ± 0.2; 9:1/8.1 ± 0.8.

sion of *SMN* exon 7. In this report, we sought to identify and characterize the *cis*-acting elements responsible for *SMN* exon 7 processing.

## RESULTS

Two nearly identical *SMN* genes are typically present in normal unaffected individuals; however, only the presence of *SMN1* and the production of high levels of full-length SMN protein protect against the development of SMA. A single nucleotide difference between *SMN1* and *SMN2* affects the generation of mRNA species (3,8). The *SMN1*-derived nucleotide (exon 7+6/C) results in the production of full-length *SMN*, and the *SMN2*-derived nucleotide (a C→T transition) (Fig. 1A) results in an aberrantly spliced product lacking exon 7, called Δ7. The following experiments were designed to identify and characterize the regulatory mRNA-processing elements surrounding *SMN* exon 7. A semi-quantitative radiolabeled polymerase chain reaction (PCR)-based assay was developed to analyze the effects of the *SMN* exon 7 mutagenesis. *pSMN1* and *pSMN2* plasmids contain genomic sequences from *SMN* exons 6–8 and have previously been shown to recapitulate the splicing patterns of the respective endogenous *SMN* genes (8). Following transient transfection, total RNA was isolated and used to generate cDNA. PCR amplification utilized a primer set specific to the plasmid segment flanking the *SMN* sequences; therefore, only plasmid-derived *SMN* transcripts were detected. This was necessary since all cell lines contain endogenous *SMN* transcripts. Following 15 rounds of amplification, detectable *pSMN1* expression was exclusively detected as a full-length cDNA product that included exons 6, 7 and 8, whereas *pSMN2* expression was predominantly the exon-skipped product, Δ7 (Fig. 1B). These results have been repeated numerous times and the quantitative relative ratios of full-length:Δ7 for *pSMN1* and *pSMN2* are routinely >99 ± 1 and 16 ± 3%, respectively. Control reactions that reflected the typical full-length:Δ7 ratios were performed with these templates in each of the following experiments to maintain consistency throughout (data not shown).

It was critical to demonstrate that the PCR results represented linear amplification and that accurate amplification of both mRNA species in a single reaction occurred. To this end, full-length and Δ7 cDNAs derived from the *pSMN2* mini-gene were cloned into the same vector used for genomic expression. Therefore, the amplification products from the following control PCR reactions were identically sized as the spliced products from *pSMN1* and *pSMN2*. The same primer set was then used in a series of competitive PCR reactions utilizing known concentrations of cDNA-containing plasmids. Using 15 PCR cycles, representative amplification occurred over two orders of magnitude in template concentration (Fig. 1C). Similar results were obtained with up to 20 cycles (data not shown). Therefore, this assay could be used for the analysis of *SMN* exon 7 mRNA processing requirements.

To characterize *SMN* exon 7 mRNA processing, we sought to determine whether the critical *SMN* exon 7+6 position absolutely required the *SMN1*-derived C nucleotide. Previously, we reported that the C→T (U in RNA sequence) transition within

*SMN* exon 7 resulted in the abundant accumulation of the exon-skipped mRNA species, Δ7. Although referred to throughout this proposal as a T nucleotide, it is implied that T refers to the DNA sequence and will transcribe into a U residue in RNA sequence. Site-specific mutagenesis introduced either an A or G substitution for this C nucleotide. In the nomenclature for the following mutations, the parent plasmid is indicated and the mutation follows the dash. Introduction of an A nucleotide resulted in a low level of exon skipping, whereas the G substitution had an even lower, though detectable, level of exon-skipped product (Fig. 2A). Interestingly, ESEs are typically sensitive to U substitutions in the enhancer regions, whereas C and G residues typically have a neutral effect (24,25,31), suggesting that *SMN* exon 7 may contain one or more ESEs.

ESEs are *cis*-elements that serve as binding substrates for splicing factors that are recruited to exons, and in turn regulate mRNA processing in addition to the typical 5′ and 3′ splice signals. The T at exon 7+6 in *SMN2* might disrupt such an ESE. We therefore determined whether *SMN* exon 7 contained splice enhancers. Three potential enhancer regions were identified by sequence (20–25). Putative splice enhancer region 1 (SE1) is a 5′ CA-rich region, SE2 is a centrally located AG-rich element and SE3 contains CA sequences in the 3′ end of the exon (Fig. 1A). SE2 would be predicted to represent the strongest putative enhancer element, since uninterrupted AG-rich stretches have previously been shown to function as potent ESEs (22–25). The three potential splicing enhancer motifs were disrupted by introducing mutations into an otherwise wild-type *pSMN1* background. In this strategy, each mutated nucleotide would result in a U residue in the RNA sequence. The mutations are indicated above each nucleotide within the putative elements (Fig. 1A). Importantly, none of these mutations introduced a premature termination codon (PTC), since PTCs can affect exon retention (32). In contrast to the parent construct *pSMN1*, which expressed undetectable levels of the exon-skipped product, the introduction of SE1 or SE3 mutations resulted in the accumulation of full-length and Δ7 species to similar levels (Fig. 2B, lanes 3 and 5). The SE2 mutation dramatically altered splicing patterns, yielding abundant Δ7 and undetectable levels of full-length transcripts (Fig. 2B, lane 4). The combination of either SE1 or SE3 mutations in conjunction with SE2 resulted in expression levels similar to those observed with SE2 alone (Fig. 2B, lanes 6 and 8), i.e. exclusively Δ7 transcripts. Introduction of both SE1 and SE3 mutations into *pSMN1* further increased the amount of Δ7 relative to full-length compared with either mutation individually; nonetheless, a low level of full-length product was still readily detectable (Fig. 2B, lane 7). Taken together, these results suggest that the AG-rich SE2 region functions as an ESE within a constitutively spliced exon.

One function of ESEs is to strengthen suboptimal splice signals flanking the exon (11,30). Although the 3′ and 5′ splice sites flanking *SMN* exon 7 are near consensus, the polypyrimidine tract (PPT) could be suboptimal since a non-consensus A residue interrupts the pyrimidine stretch (29,30). In the following mutations, the relative strength of the *SMN* exon 7 PPT was increased by replacing the exon 7 PPT with 12 consecutive U residues (Fig. 1A). The uninterrupted stretch of 12 U residues in the RNA abrogated the typical exon-skipped pattern of *pSMN2*-derived transcripts and resulted in high

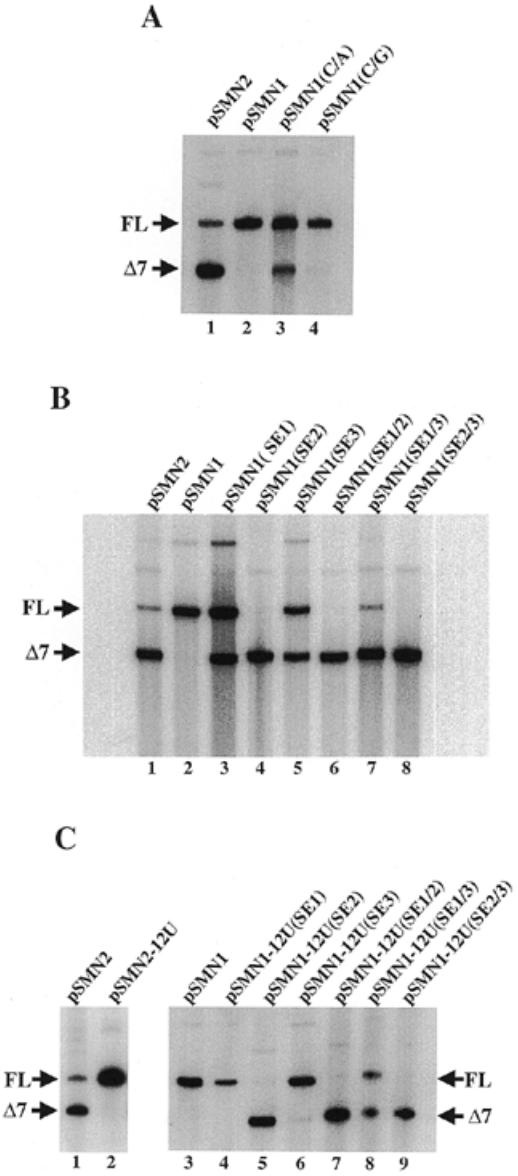


**Figure 2.** (**A**) [$^{32}$P]dATP-labeled RT–PCR of total RNA isolated from U2OS cells 36 h post-transfection with 10 µg of *pSMN2*, *pSMN1*, *pSMN1*(C/A) or *pSMN1*(C/G); exon 7+6 nucleotide was a U, C, A or G, respectively. (**B**) *pSMN1* exon 7 mutations SE1, SE2 or SE3, and double mutations SE1/2, 1/3 or 2/3. (**C**) Single and double exon 7 ESE mutations with 12U PPT mutation. *pSMN1-12U* creates 12 consecutive U residues in the upstream PPT.

levels of full-length transcript and no Δ7 (Fig. 2C, lanes 1 and 2). The enhanced PPT (12U PPT) also restored full-length expression to *pSMN1* constructs with either SE1 or SE3 mutations (Fig. 2C, lanes 4 and 6). However, no increase in full-length mRNA was detected when the exon 7 RNA contained the SE2 mutation (Fig. 2C, lane 5). Furthermore, exon 7 inclusion was not increased by the 12U PPT within the context of SE1/2 or SE2/3 double mutations, whereas the SE1/3 double mutation was partially compensated by the increased strength of the PPT (Fig. 2C, lanes 7–9).

These studies highlight the importance of the SE2 element in *SMN* exon 7 processing. To characterize this region further, a series of smaller sequential substitution mutations were intro-

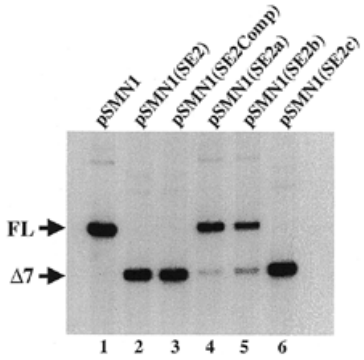


**Figure 3.** Analysis of the *SMN* exon 7 SE2 AG-rich element. [$^{32}$P]dATP RT–PCR analysis (Fig. 1B) of transiently expressed transcripts from *pSMN1*, -SE2, -SE2Comp, -SE2a, -SE2b and -SE2c. 'Comp' introduced the complementary strand into the coding strand and SE2a–c introduced UUU substitutions into sequential segments of SE2.

duced into the AG-rich region. These mutations replaced three consecutive bases from the AG-rich region with three U residues, dividing SE2 into three subdomains: SE2a, SE2b and SE2c (Fig. 1A). Disruption of the first two SE2 regions, SE2a and SE2b, with UUU substitutions slightly increased the relative amount of the exon-skipped product. Mutation of SE2c resulted in a concomitant decrease of full-length and an increase of Δ7 mRNA (Fig. 3, lanes 4–6). Finally, the entire region spanning SE2 was replaced with its complementary sequence. This severe mutation also resulted in high levels of Δ7 expression (Fig. 3, lane 3). Taken together, these results demonstrate that the AG-rich SE2 ESE is necessary for constitutive inclusion of *SMN* exon 7. This ESE is also present in *SMN2*; however, it fails to overcome the T at position +6 that causes exon 7 skipping.

The AG-rich element in *SMN* exon 7 functions in this experimental context to facilitate high levels of constitutive full-length expression, but it is generally acknowledged that a true ESE should function in a heterologous context. To demonstrate conclusively that the AG-rich region of *SMN* exon 7 functioned as an ESE, we tested the ability of the SE2 region to support efficient splicing of the *Drosophila double-sex* (*dsx*) gene (17,24). Using the *dsx* cassette, which has been used previously to characterize a number of ESEs (17,25), the complete sequence of *SMN1* or *SMN2* exon 7 was cloned downstream of *dsx* exon 4 (Fig. 4A). Additionally, three or five wild-type copies of the AG-rich region identified by the SE2 mutation were similarly positioned downstream of *dsx*. All four *in vitro* splicing templates contained fusion exons of essentially the same size, and therefore produced similarly sized spliced products. Relatively low levels of properly spliced product were detected with either the *dsx-SMN1* or *-SMN2* templates (Fig. 4B, lanes 1 and 2). Insertion of the three copies of the AG-rich element had a modest effect on splicing efficiency; however, five copies of wild-type SE2 sequences supported high-efficiency *in vitro* splicing (Fig. 4B, lanes 3 and 4). Similar results were obtained following a 30 or 60 min reaction time (data not shown). The positive control contains multiple strong ESEs from the bovine papillomavirus (BPV) that result in similar sized products (Fig. 4B, lane 5) (20–25). The AG-rich region within *SMN* exon 7 therefore functions as

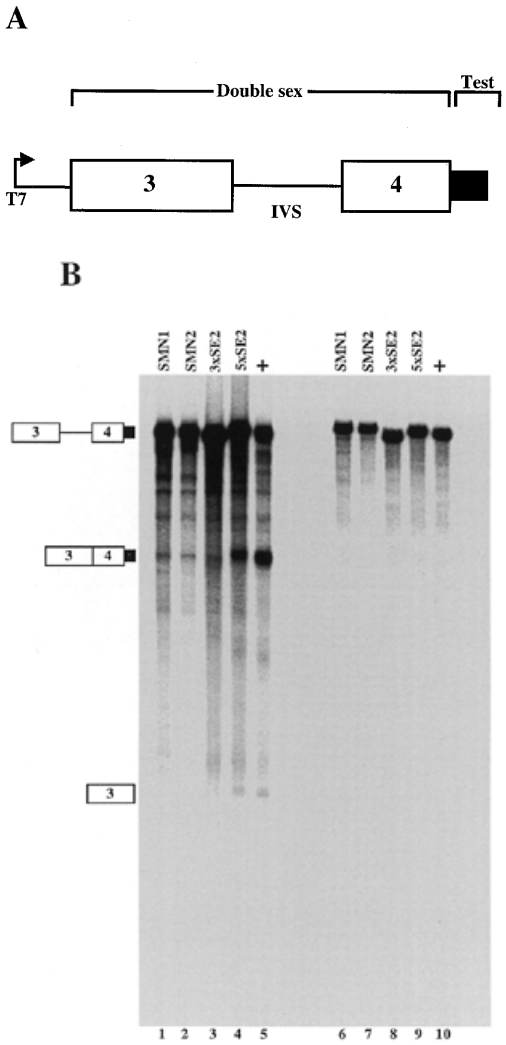


**Figure 4.** *SMN* SE2 functions as an efficient ESE *in vitro*. (**A**) Expanded view of the *in vitro* splicing cassette. *Drosophila double-sex* (*dsx*) exons (boxes) and intervening sequences (IVS) are bracketed downstream of a T7 promoter. Test sequences (black box) were positioned downstream of exon 4. Exon 3, IVS and exon 4 are 168, 114 and 44 bp, respectively. (**B**) *In vitro* splicing of *dsx*-fusion templates with *SMN1* or *SMN2* exon 7 and three or five wild-type copies of the *SMN* exon 7 SE2 region (lanes 1–4, respectively) and unspliced templates (left and right sides of panel, respectively). Positions of unspliced, spliced and exon 3 alone are indicated schematically (left).

an ESE within its natural context *in vivo* and in a heterologous *in vitro* splicing system.

Our data show that the aberrant splicing of *SMN2* results in expression of SMNΔ7, a protein that cannot functionally compensate for the loss of full-length protein in SMA. Having identified a critical genetic element that regulates the expression of full-length versus Δ7, we addressed the significance of the aberrant splice in the development of SMA and SMN protein biochemistry. Whereas the Δ7 protein self-associates less efficiently and fails to enhance *in vitro* splicing, there has been speculation that Δ7 is less stable than full-length protein, and therefore is unable to provide protection from development of disease. To compare the intracellular stability of the SMN proteins, N-terminal hemagglutinin (HA) epitope-tagged HA:SMN or HA:Δ7 SMN expression constructs were tran-

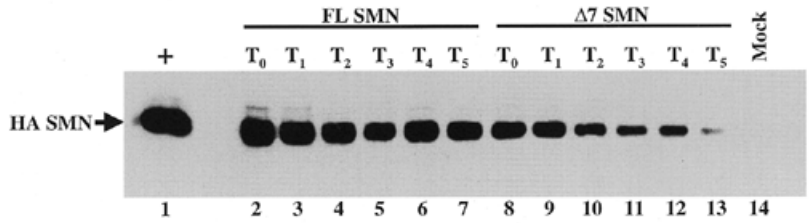


**Figure 5.** Relative protein stability of full-length (FL) and Δ7 SMN. HA-tagged FL SMN and Δ7 SMN expression vectors were transiently expressed in U2OS cells for 24 h post-transfection, washed and re-fed in 10 mM cycloheximide. Cells were harvested at 0, 45, 90, 120, 150 and 180 min after cycloheximide addition. Equivalent amounts (~100 μg) of total cellular extract were analyzed by western blot with anti-HA monoclonal antibody (12CA5) and a horseradish peroxidase-conjugated anti-mouse monoclonal antibody. Extract from mock-transfected cells was a negative control.

siently expressed in U2OS cells, a human osteosarcoma cell line. Expression was under the control of the constitutively expressing cytomegolovirus immediate-early promoter. Twenty-four hours post-transfection, cells were washed extensively and treated with 10 mM cycloheximide to inhibit protein synthesis. Cells were harvested at time zero and five subsequent time points ($T_0 = 0$; $T_1 = 45$; $T_2 = 90$; $T_3 = 120$; $T_4 = 150$; $T_5 = 180$ min). Equivalent amounts of cellular extract were analyzed by western blot using αHA monoclonal antibody. Full-length HA:SMN remained largely present even at the final time point, 180 min (Fig. 5, lanes 2–7), whereas Δ7 degradation was readily apparent (Fig. 5, lanes 8–13). Although full-length HA:SMN was more abundant at the first time point compared with HA:SMNΔ7, the difference in stability can still be observed since at the last time point only very low levels of SMNΔ7 were present. The membrane was stained to confirm that loading and transfer were equivalent (data not shown). Taken together, these results identify a critical genetic element required for the constitutive inclusion of *SMN* exon 7. The subsequent protein product of the exon-skipping event, which is the principal product of the *SMN2* genes, was less stable and therefore provides further evidence of why SMNΔ7 protein fails to protect against the development of SMA.

## DISCUSSION

### Identification and characterization of a purine-rich enhancer within the constitutively spliced *SMN* exon 7

One of the most critical aspects of SMA pathogenesis is understanding why *SMN2* cannot protect against SMA in the absence of the nearly identical *SMN1* gene. SMA patients are clinically categorized based on age of onset and severity of symptoms: type I (severe) to type III (mild) (33). Although not protective against the development of SMA, *SMN2* acts as a disease modifier and decreases SMA severity in a dose-dependent manner. There is an excellent correlation between higher number of *SMN2* copies and milder forms of SMA (34). Several groups have demonstrated that the steady-state levels of full-length, wild-type SMN protein correlated with disease development (35,36). Thus, the low levels of *SMN2*-derived full-length SMN protein are likely to provide partial protection from disease. Furthermore, the primary product of *SMN2*, SMNΔ7 protein, is defective for self-association (37), stimulation of *in vitro* splicing (7) and, as we show here, is much less stable.

Our study was designed to identify and characterize the mRNA processing elements regulating *SMN* exon 7 splicing events. An AG-rich enhancer in the center of exon 7 was necessary for constitutive inclusion of *SMN* exon 7 *in vivo* and was sufficient for stimulating efficient *in vitro* splicing. The mutational analysis of SE1 and SE3 identified additional requirements for maximal levels of exon inclusion and suggested that exon 7 contains multiple layers of regulation. Interestingly, the effects of the SE1 and SE3 mutations were abrogated by increasing the relative strength of the upstream PPT. In contrast, the SE2 region was essential and disruption of this region could not be overcome by increasing the PPT strength. Notably, the AG-rich motif within SE2 was the closest of the three elements to a consensus ESE motif (13,29,38). These results are consistent with the exon 7 AG-rich element functioning as an ESE. ESEs have been identified within constitutively and alternatively regulated exons, and one mechanistic role of ESEs is to compensate for suboptimal flanking splice signals. RNA stability is not likely to be contributing to these results since several mutations that reduce full-length expression were retained in the final transcript by increasing the strength of the PPT. In further support of SE2 acting as a bona fide ESE, SE2 functioned in a heterologous context by stimulating splicing of the *dsx* splicing cassette. Interestingly, the AG-rich motif is extremely well conserved across divergent species in which SMN has been identified.

Proper splicing of *SMN* exon 7 is likely to involve multiple regulatory elements. It is possible that the critical C/T nucleotide difference in exon 7 between *SMN1* and *SMN2* does not disrupt the AG-rich SE2. The small effect of the C→A and C→G substitutions (Fig. 2A) supports the notion that an enhancer has been disrupted. Previously characterized ESEs are particularly sensitive to U residues, less so to As, and Gs can be neutral (24,25,31). This pattern of nucleotide preference was closely adhered to at the exon 7+6 position. Relatively little is known, however, about the role of the 5′-most exonic sequences regarding splice site selection, and the C/T transition may represent a disruption at a different step of the splicing reaction. A single nucleotide substitution flanking the +6 position, however, did not affect exon 7 processing (data not shown). Since little sequence homology exists between heterologous 5′ exon sequences, the mechanism underlying why the +6 T causes exon skipping is currently being investigated.

The *in vitro* splicing results confirm that the SE2 region acts as an ESE. However, in the context of a single copy of full-length exon 7, *SMN1* or *SMN2 in vitro* splicing was inefficient

and the difference between *SMN1* and *SMN2* observed *in vivo* was not apparent. Heterologous ESEs have been identified using similar approaches, and it is not uncommon to require multiple copies of ESEs to mediate efficient splicing *in vitro*, likely due to the absence of additional elements that serve to complement the enhancer present in its natural context. Several possibilities exist as to why a single *SMN1* or *SMN2* chimeric exon 7 template did not recapitulate their respective *in vivo* splicing patterns. The distance from the 3′ splice site may influence the activity. The *SMN* sequences are fused downstream of the *dsx* exon 4; therefore, ~50 bp are present between the natural 5′ end of exon 7 and the 3′ splice site of the *dsx* gene. Since ESEs typically function in a distance-dependent manner (22,24,31), this additional distance may disrupt the regulatory elements within *SMN* exon 7. Secondly, the C→T disruption may be subtle enough that in the absence of a downstream competitive splice site and exon (i.e. *SMN* exon 8), inclusion of exon 7 occurs essentially by default. Thirdly, the *dsx* cassette is designed to assess the relative strength of enhancers (22,24,31) and in a single copy neither *SMN1* nor *SMN2* is sufficient and the C→T transition represents a disruption that was not measured in this context. Designing an *in vitro* template that recapitulates the *in vivo SMN1* and *SMN2* splicing patterns represents an important step in dissecting the biochemistry of exon 7 regulation.

### The exon-skipped protein product is less stable and cannot provide protection from development of SMA

SMA develops as a consequence of the lack of a single functioning copy of *SMN1*, which expresses abundant full-length transcripts and no detectable levels of Δ7. Interestingly, even multiple copies of *SMN2* that arise through partial recombination events cannot fully protect from SMA (39). The primary product of *SMN2* is the exon-skipped product Δ7 RNA (1,9). We have reported that the Δ7 product is defective in self-association (37) and has been shown to be unable to stimulate *in vitro* splicing reactions like full-length SMN (7). This work identified a critical genetic regulatory mechanism for the production of Δ7 RNA. The protein product of the aberrant splicing event was much less stable than full-length SMN (Fig. 5). The stability of each protein form may be linked to their abilities to self-associate or to associate with additional cellular factors such as the Sm proteins (5,6,37). Potentially a more dramatic difference in protein stability would have been seen within a cellular background that lacked full-length endogenous SMN. SMN self-association is mediated by the peptide encoded by exon 6. Since the endogenous full-length SMN would be able to associate with the transiently expressed Δ7 product through the intact self-association domain encoded by exon 6 (37), the hetero-oligomers may form more stable complexes than Δ7 homo-oligomers. To date, all cell types examined contain full-length SMN, including type I SMA patients, even though endogenous SMN expression is typically decreased in cells derived from these types of patient (35,36). The relative instability of Δ7 may also contribute to disease development by binding to and subsequently destabilizing full-length SMN in a hetero-oligomer, thus explaining why even four copies of *SMN2* fail to protect fully from SMA. Transiently expressed SMNΔ7 protein has recently been detected 48 h post-transfection (40); however, this observation was based on steady-state protein levels, not cycloheximide treatment or pulse–chase analysis.

*SMN1* and *SMN2* are dramatically distinct functionally, yet at the nucleotide level they are nearly identical. Therefore, gaining insight into the mechanism(s) that regulate exon 7 inclusion versus exclusion is important in understanding SMA development. We have previously shown that the exon 7+6 nucleotide is critical for exon 7 inclusion, and now identify an essential ESE within exon 7. Interestingly, a very small number of SMA patients are not homozygously deleted for *SMN1* (39), raising the possibility that SMA-modifying genes other than *SMN2* exist (41). In further support of SMA-modifying genes (in addition to *SMN2*), SMA patients have been identified who contained the same mutated SMA allele; however, the clinical course of the disease was different. These results suggest that the factor(s) regulating exon 7 inclusion, including regulation through the AG-rich element, may be candidates for SMA-modifying genes. Identifying the factor(s) mediating SE2 responsiveness may lead to new insights into the pathogenesis of SMA.

## MATERIALS AND METHODS

### Plasmids

Previously described *pSMN1* and *pSMN2* mini-genes (8) were used as templates to generate oligo-mediated site-specific mutations with Thermo Pol Vent polymerase (New England Biolabs, Beverly, MA) using the splicing of overlapping extensions strategy, and cloned into the appropriate plasmid backbone between the *Bcl*I (*SMN* intron 6) and *Not*I (pCI polylinker) sites. The following primer sets were used to generate the indicated mutations (in bold):

*pSMN1*(C/A): 5′-CCT TAC AGG GTT T**A**A GAC AAA ATC-3′;
5′-GAT TTT GTC T**T**A AAC CCT GTA AGG-3′;
*pSMN1*(C/G): 5′-CCT TAC AGG GTT T**G**A GAC AAA ATC-3′;
5′-GAT TTT GTC T**C**A AAC CCT GTA AGG-3′;
SE1: 5′-CCT TAC AGG GTT TCA **TT**C **TT**A ATC AAA AAG AAG GAA GG-3′;
5′-CCT TCC TTC TTT TTG ATT **AA**G **AA**T GAA ACC CTG TAA GG-3′;
SE2: 5′-CAG ACA AAA TCA A**TT** A**TT** A**TT** AAG GTG CTA C-3′;
5′-GTG AGC ACC TT**A A**T**A A**T**A A**TT GAT TTT GTC TG-3′;
SE3: 5′-GGA AGG TGC TC**T** C**T**T TCC TT**T TT**T TAA GGA GTA AGT CTG CCA GC-3′;
5′-GCT GGC AGA CTT AC TAC TCC TTA A**AA A**AA GGA AA**G A**GA GCA CCT TCC-3′;
12Tppt: 5′-CTT CCT TT**T** TTT T**TT** TTA CAG GGT TTT/C AGA-3′;
5′-TCT G/AAA ACC CTG TAA **AA**A AAA **A**AA AGG AAG-3′;
SE2a: 5′-CAA AAT CAA **TTT** GAA GGA AGG TGC-3′;
5′-GCA CCT TCC TTC **AAA** TTG ATT TTG-3′;
SE2b: 5′-CAA AAT CAA AAA **TTT** GGA AGG TGC TCA C-3′;
5′-GTG AGC ACC TTC C**AA A**TT TTT GAT TTT G-3′;
SE2c: 5′-CAA AAA GAA **TTT** AGG TGC TCA CAT TCC-3′;
5′-GGA ATG TGA GCA CCT **AAA** TTC TTT TTG-3′.

*In vitro* splicing cassettes were constructed by annealing the following pairs of complementary oligonucleotides that created *Xba*I and *Hin*dIII overhanging ends between the *Xba*I and *Hin*dIII sites of 3014, a plasmid that contains the previously described *Drosophila double-sex* splicing cassette (25):

*dsx*:*SMN1* and *SMN2*: 5′-CTA GAG GTT T**C/T**A GAC AAA ATC AAA AAG AAG GAA GGT GCT CAC TT CCT TTA AAT TAA GGA A-3′;
5′-AGC TTT CCT TAA TTT AAG GAA TGT GAG CAC CTT CCT TCT TTT TGA TTT TGT CT**G/A** AAA CCT-3′;
3xSE2: 5′-CTA GAC AAA AGA AGG AAG GCA AAA AGA AGG AAG GCA AAA AAG AAG GAA GGA A-3′;

5′-AGC TTT CCT TCC TTC TTT TTGCC TTCCTT CTT TTT GCC TTC CTT CTT TTG T-3′;
5xSE2: 5′-CTA GAA GAA GGA AGG AGA AGG AAG GAG AAG GAA GGA GAA GGA AGG AGA AGG AAG GAA-3′;
5′-AGC TTT CCT TCC TTC TCC TTC CTT CTC CTT CCT TCT CCT TCC TTC TCC TTC CTT CTT-3′.

Mammalian expression vectors pCI:HA-SMN and -SMNΔ7 were constructed by cloning the HA:SMN or Δ7 containing fragment from pcDNA3:HA:SMN or Δ7 (*Kpn*I and a Klenow-filled *Xho*I fragment) into pCI (*Kpn*I and *Sma*I).

### Reverse transcription–PCR (RT–PCR)

U2OS cells, a human osteosarcoma line, were transiently transfected by standard $CaPO_4$ procedures with 10 μg of the indicated plasmid. Thirty-six hours post-transfection, total RNA was isolated in guanidine thiocyanate lysis buffer and purified over a CsCl gradient. First-strand cDNA synthesis and amplification of plasmid-derived cDNAs were performed as previously described (8) except that pCI:Fwd#2 (5′-CGA CTC ACT ATA GCC TAG CC-3′) was used in the PCR step, only 15 cycles were performed, and 40 μM dATP supplemented with 1 μl of [$^{32}$P]dATP (3000 Ci/mmol; NEN, Boston, MA) were used. Reaction products were boiled and resolved in a 6% sequencing gel. Quantitations were performed on a BioRad (Hercules, CA) PhosphorImager; full-length and Δ7 transcripts were quantitated and expressed as a percent relative to full-length expression within the same reaction.

### *In vitro* splicing

The Promega (Madison, WI) RNA Splicing System was used as recommended by the manufacturer. Briefly, *in vitro* splicing reactions were performed at 30°C for 3 h using 10 μl of HeLa nuclear extract and ~10 ng of gel-purified [$^{32}$P]dUTP-labeled RNA templates. Reaction products were phenol extracted twice, precipitated and resolved on a 6% sequencing gel. Positive control reactions resulted in similarly sized splice products and DNA sequencing reactions were used to identify the final products.

## ACKNOWLEDGEMENTS

We thank Tom Maniatis and laboratory members, Mark Roth, Adrian Krainer and David Pintel for helpful discussions, and Carl Baker for the *double-sex* construct and advice. C.L.L. was supported by a fellowship from the Muscular Dystrophy Association. Funding for these studies was provided by the Muscular Dystrophy Association and Families of SMA to E.J.A.

## REFERENCES

1. Lefebvre, S., Burglin, L., Reboullet, S., Clermont, O., Burlet, P., Viollet, L., Benichou, B., Cruaud, C., Millasseau, P., Zeviani, M. *et al*. (1995) Identification and characterization of a spinal muscular atrophy determining gene. *Cell*, **80**, 155–165.
2. Burglen, L., Lefebvre, S., Clermont, O., Burlet, P., Viollet, L., Cruaud, C., Munnich, A. and Melki, J. (1996) Structure and organization of the human survival motor neuron (SMN) gene. *Genomics*, **32**, 479–482.
3. Monani, U.R., Lorson, C.L., Parsons, D.W., Prior, T.W., Androphy, E.J., Burghes, A.H. and McPherson, J.D. (1999) A single nucleotide difference that alters splicing patterns distinguishes the SMA gene *SMN1* from the copy gene *SMN2*. *Hum. Mol. Genet.*, **8**, 1177–1183.
4. Lorson, C.L. and Androphy, E.J. (1998) The domain encoded by exon 2 of the survival motor neuron protein mediates nucleic acid binding. *Hum. Mol. Genet.*, **7**, 1269–1275.
5. Liu, Q., Fischer, U., Wang, F. and Dreyfuss, G. (1997) The spinal muscular atrophy disease gene product, SMN, and its associated protein SIP1 are in a complex with spliceosomal snRNP proteins. *Cell*, **90**, 1013–1021.
6. Fischer, U., Liu, Q. and Dreyfuss, G. (1997) The SMN–SIP1 complex has an essential role in spliceosomal snRNP biogenesis. *Cell*, **90**, 1023–1029.
7. Pellizzoni, L., Kataoka, N., Charroux, B. and Dreyfuss, G. (1998) A novel function for SMN, the spinal muscular atrophy disease gene product, in pre-mRNA splicing. *Cell*, **95**, 615–624.
8. Lorson, C.L., Hahnen, E., Androphy, E.J. and Wirth, B. (1999) A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy. *Proc. Natl Acad. Sci. USA*, **96**, 6307–6311.
9. Gennarelli, M., Lucarelli, M., Capon, F., Pizzuti, A., Merlini, L., Angelini, C., Novelli, G. and Dallapiccola, B. (1995) Survival motor neuron gene transcript analysis in muscles from spinal muscular atrophy patients. *Biochem. Biophys. Res. Commun.*, **213**, 342–348.
10. Wirth, B., Herz, M., Wetter, A., Moskau, S., Hahnen, E., Rudnik-Schoneborn, S., Wienker, T. and Zerres, K. (1999) Quantitative analysis of survival motor neuron copies: identification of subtle SMN1 mutations in patients with spinal muscular atrophy, genotype–phenotype correlation, and implications for genetic counseling. *Am. J. Hum. Genet.*, **64**, 1340–1356.
11. Black, D.L. (1995) Finding splice sites within a wilderness of RNA. *RNA*, **1**, 763–771.
12. Reed, R. (1996) Initial splice-site recognition and pairing during pre-mRNA splicing. *Curr. Opin. Genet. Dev.*, **6**, 215–220.
13. Manley, J.L. and Tacke, R. (1996) SR proteins and splicing control. *Genes Dev.*, **10**, 1569–1579.
14. Fu, X.D. (1995) The superfamily of arginine/serine-rich splicing factors. *RNA*, **1**, 663–680.
15. Xiao, S.H. and Manley, J.L. (1997) Phosphorylation of the ASF/SF2 RS domain affects both protein–protein and protein–RNA interactions and is necessary for splicing. *Genes Dev.*, **11**, 334–344.
16. Wu, J.Y. and Maniatis, T. (1993) Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell*, **75**, 1061–1070.
17. Tian, M. and Maniatis, T. (1993) A splicing enhancer complex controls alternative splicing of doublesex pre-mRNA. *Cell*, **74**, 105–114.
18. Tian, M. and Maniatis, T. (1994) A splicing enhancer exhibits both constitutive and regulated activities. *Genes Dev.*, **8**, 1703–1712.
19. Sun, Q., Mayeda, A., Hampson, R.K., Krainer, A.R. and Rottman, F.M. (1993) General splicing factor SF2/ASF promotes alternative splicing by binding to an exonic splicing enhancer. *Genes Dev.*, **7**, 2598–2608.
20. Coulter, L.R., Landree, M.A. and Cooper, T.A. (1997) Identification of a new class of exonic splicing enhancers by *in vivo* selection. *Mol. Cell. Biol.*, **17**, 2143–2150 [Erratum. *Mol. Cell. Biol.*, **17**, 3468].
21. Tian, H. and Kole, R. (1995) Selection of novel exon recognition elements from a pool of random sequences. *Mol. Cell. Biol.*, **15**, 6291–6298.
22. Tian, M. and Maniatis, T. (1992) Positive control of pre-mRNA splicing *in vitro*. *Science*, **256**, 237–240.
23. Wang, Z., Hoffmann, H.M. and Grabowski, P.J. (1995) Intrinsic U2AF binding is modulated by exon enhancer signals in parallel with changes in splicing activity. *RNA*, **1**, 21–35.
24. Watakabe, A., Tanaka, K. and Shimura, Y. (1993) The role of exon sequences in splice site selection. *Genes Dev.*, **7**, 407–418.
25. Zheng, Z.M., He, P.J. and Baker, C.C. (1997) Structural, functional, and protein binding analyses of bovine papillomavirus type 1 exonic splicing enhancers. *J. Virol.*, **71**, 9096–9107.
26. Lynch, K.W. and Maniatis, T. (1995) Synergistic interactions between two distinct elements of a regulated splicing enhancer. *Genes Dev.*, **9**, 284–293.
27. Xu, R., Teng, J. and Cooper, T.A. (1993) The cardiac troponin T alternative exon contains a novel purine-rich positive splicing element. *Mol. Cell. Biol.*, **13**, 3660–3674.
28. Buvoli, M., Mayer, S.A. and Patton, J.G. (1997) Functional crosstalk between exon enhancers, polypyrimidine tracts and branchpoint sequences. *EMBO J.*, **16**, 7174–7183.
29. Gersappe, A. and Pintel, D.J. (1999) CA- and purine-rich elements form a novel bipartite exon enhancer which governs inclusion of the minute virus of mice NS2-specific exon in both singly and doubly spliced mRNAs. *Mol. Cell. Biol.*, **19**, 364–375.

30. Berget, S.M. (1995) Exon recognition in vertebrate splicing. *J. Biol. Chem.*, **270**, 2411–2414.
31. Tanaka, K., Watakabe, A. and Shimura, Y. (1994) Polypurine sequences within a downstream exon function as a splicing enhancer. *Mol. Cell. Biol.*, **14**, 1347–1354.
32. Maquat, L.E. (1995) When cells stop making sense: effects of nonsense codons on RNA metabolism in vertebrate cells. *RNA*, **1**, 453–465.
33. Pearn, J. (1980) Classification of spinal muscular atrophies. *Lancet*, **i**, 919–922.
34. Parsons, D.W., McAndrew, P.E., Iannaccone, S.T., Mendell, J.R., Burghes, A.H.M. and Prior, T.W. (1998) Intragenic telSMN mutations: frequency, distribution, evidence of a founder effect, and modification of the spinal muscular atrophy phenotype by cenSMN copy number. *Am. J. Hum. Genet.*, **63**, 1712–1723.
35. Coovert, D., Le, T., McAndrew, P., Strasswimmer, J., Crawford, T., Mendell, J., Coulson, S., Androphy, E.J., Prior, T. and Burghes, A.H.M. (1997) The survival motor neuron protein in spinal muscular atrophy. *Hum. Mol. Genet.*, **6**, 1205–1214.
36. Lefebvre, S., Burlet, P., Liu, Q., Bertrandy, S., Clermont, O., Munnich, A., Dreyfuss, G. and Melki, J. (1997) Correlation between severity and SMN protein level in spinal muscular atrophy. *Nature Genet.*, **16**, 265–269.
37. Lorson, C.L., Strasswimmer, J., Yao, J.-M., Baleja, J.D., Hahnen, E., Wirth, B., Thanh, L., Burghes, A.H.M. and Androphy, E.J. (1998) SMN oligomerization defect correlates with spinal muscular atrophy severity. *Nature Genet.*, **19**, 63–66.
38. Lynch, K.W. and Maniatis, T. (1996) Assembly of specific SR protein complexes on distinct regulatory elements of the *Drosophila doublesex* splicing enhancer. *Genes Dev.*, **10**, 2089–2101.
39. Burghes, A. (1997) When is a deletion not a deletion? When it is converted. *Am. J. Hum. Genet.*, **61**, 9–15.
40. Mohaghegh, P., Rodrigues, N.R., Owen, N., Ponting, C.P., Le, T.T., Burghes, A.H. and Davies, K.E. (1999) Analysis of mutations in the tudor domain of the survival motor neuron protein SMN. *Eur. J. Hum. Genet.*, **7**, 519–525.
41. Scharf, J.M., Endrizzi, M.G., Wetter, A., Huang, S., Thompson, T.G., Zerres, K., Dietrich, W.F., Wirth, B. and Kunkel, L.M. (1998) Identification of a candidate modifying gene for spinal muscular atrophy by comparative genomics. *Nature Genet.*, **20**, 83–86.