Alternative splicing in spinal muscular atrophy underscores the role of an intron definition model

Natalia N. Singh and Ravindra N. Singh
Department of Biomedical Sciences; Iowa State University; Ames, IA USA

Humans have two nearly identical copies of the Survival Motor Neuron (SMN) gene: SMN1 and SMN2. The two SMN genes code for identical proteins; however, SMN2 predominantly generates a shorter transcript due to skipping of exon 7, the last coding exon. Skipping of SMN2 exon 7 leads to production of a truncated SMN protein that is highly unstable. The inability of SMN2 to compensate for the loss of SMN1 results in spinal muscular atrophy (SMA), the second most prevalent genetic cause of infant mortality. Since SMN2 is almost universally present in SMA patients, correction of SMN2 exon 7 splicing holds the promise for cure. Consistently, SMN2 exon 7 splicing has emerged as one of the best studied splicing systems in humans. The vast amount of recent literature provides a clue that SMN2 exon 7 splicing is regulated by an intron definition mechanism, which does not require cross-exon communication as prerequisite for exon inclusion. Our conclusion is based on the prominent role of intronic cis-elements, some of them have emerged as the frontrunners among potential therapeutic targets of SMA. Further, the widely expressed T-cell-restricted intracellular antigen-1 (TIA1), a member of the glutamine rich domain containing RNA-binding proteins, has recently been found to regulate SMN exon 7 splicing by binding to intron 7 sequences away from the 5’ splice site (ss). These findings make a strong argument for an “intron definition model,” according to which regulatory sequences within a downstream intron are capable of enforcing exon inclusion even in the absence of a defined upstream 3’ ss of an alternatively spliced exon.

Introduction

Spinal muscular atrophy (SMA) is a leading genetic cause of infant mortality.1-3 The disease is attributed to deletion or mutation within Survival Motor Neuron 1 (SMN1) gene.4 SMN2, a nearly identical copy of SMN1, fails to prevent SMA due to a critical C-to-T mutation at the 6th position (C6U mutation in RNA) of exon 7.5-7 C6U mutation leads to a predominant exon 7 skipping from SMN2, producing mostly truncated protein SMNΔ7, which is unstable.5 High SMN2 copies capable of producing cumulative high levels of SMNΔ7 as well as full-length SMN fully rescue SMA mice, suggesting that SMNΔ7 does not play a dominant negative role.7 Full-length SMN protein (hereafter referred to as “SMN”) has multiple functions that include snRNP biogenesis, transcription, splicing, cell signaling, macromolecular trafficking and stress granule formation.10-18 Consistent with the essential role of SMN in cellular metabolism, complete loss of SMN is embryonic lethal.19

SMA is a model genetic disease in which correction of aberrant splicing of a single exon (i.e., SMN2 exon 7) holds the promise for cure of more than 90% of the affected population.20 Consistently, SMN2 exon 7 splicing has been extensively studied and remains one of the highly contested topics among genetic diseases linked to defective splicing. Many excellent reviews published recently highlight approaches of SMN2 exon 7 splicing.

Key words: SMN1, SMN2, SMA, TIA1, Splicing, ISS-N1, hnRNPA1

Submitted: 03/30/11
Revised: 04/26/11
Accepted: 05/02/11
DOI: 10.4161/rna.8.4.16224
*Correspondence to: Ravindra Singh; Email: singhr@iastate.edu
As per new hypothesis supporting the exon definition model, intronic sequences immediately upstream and downstream of an exon may enforce exon definition through additional cross-communication.32 However, such intronic communications may also serve the opposite purpose as has been documented in case of hnRNP A1-mediated exon skipping event through a looping out mechanism.33

Figure 1. Organization of human (SMN) and mouse (Smn) genes covering the last 3 exons. (A) Comparison of exon 7 and introns 6 and 7 of SMN and Smn genes. Exons are shown as colored boxes, whereas introns are shown as broken lines. Sizes of introns and exon 7 are given. (B) Results of RT-PCR for transcripts derived from SMN1 and SMN2 minigenes.26 Note the presence of intron 7 retained intermediate due to a weak 5’ ss in SMN1. As per intron definition model, splicing of SMN1 exon 7 was mainly driven by intron 6 definition due to a strong 3’ ss. C6U mutation causes a loss on intron 6 definition in SMN2 leading to skipping of exon 7. (C) Partial view of factors responsible for intron definition. In vivo selection of entire SMN exon 7 revealed three major regulatory regions: Exinct, conserved tract and 3’-cluster.26 SF2/ASF, hnRNP A1, Sam68 and hnRNP Q interact within the region corresponding to Exinct.46,47 Binding site of Tra2-β1 and associated factors hnRNP G, SRp30c and TDP-43 fall within conserved tract.49-52 The 3’-cluster falls within the terminal stem loop 2 (TSL2) structure that also sequesters the 5’ ss of exon 7.44 ISS-N1 is a major intronic inhibitory elements that harbors 2 hnRNP A1 motifs.27, 56 ISS-N1 also overlaps with a GC-rich sequences that provides the shortest antisense target for correction of SMN2 exon 7 splicing in SMA.55 The cytosine residue at the 10th intronic position (10C) has been implicated in a long-distance interaction (LDI) with deep intronic sequences.57 TIA1 binds to an intronic region downstream of ISS-N1.27 Cis-elements and factors that are likely to define either intron 6 or intron 7 have been grouped. (+) and (-) indicates that a given cis-element or splicing factor promotes and suppresses exon 7 inclusion, respectively. Exon 7 sequences are shown in capital letters, whereas intronic sequences are shown in small case letters.

According to conventional point of view, the exon definition model is applicable to the splicing of small internal exons that are flanked by large intronic sequences.29-31 The exon definition model requires simultaneous recognition of both splice sites through precise cross-talk involving both ends of an exon. As per new hypothesis supporting the exon definition model, intronic sequences immediately upstream and downstream of an exon may enforce exon definition through additional cross-communication.32 However, such intronic communications may also serve the opposite purpose as has been documented in case of hnRNP A1-mediated exon skipping event through a looping out mechanism.33 The exon definition model requires a highly regulated transition process from splice site recognition across short exons to assembly of the spliceosome across long
introns. However, regulation of this transition process is poorly understood.

Considering both splice sites are simultaneously selected in the exon definition model, it is rather unlikely that a disproportionate intron retention product will accumulate during splicing via exon definition mechanism. Thus, exon definition model provides a basis for challenge if disproportionate intron retention products are easily detectable in a reliable RT-PCR reaction. Also, the effect of mutations within deep intronic sequences on splicing could not be explained by the simplistic view of cross-talk between two ends of an exon (Fig. 1; described below). The intron definition model is often used to describe mechanism of splicing of large internal exons flanked by at least one small intron. Based on the wealth of information accumulated to date, we propose that splicing of SMN2 exon 7 is driven by the intron definition mechanism. Of note, much of the results described here pertain to the widely studied SMN intron 7, which is only 444-nt long. However, this size exceeds the length of a typical intron (~250 nts) defined by an intron definition model. Our argument supporting the intron definition model may be relevant to other short exons flanked by large intronic sequences.

**Exon Skipping is Linked to a Critical Exonic Position**

The initial finding that C6U mutation alone is sufficient to cause SMN2 exon 7 skipping uncovered the significance of a single position within exon 7. Close proximity of C6U mutation to the 3' ss led to an early hypothesis that a weak 3' ss of SMN2 exon 7 is the probable cause of exon 7 skipping. Subsequently, C6U mutation was linked to either loss of an SF2/ASF-associated enhancer or gain of an hnRNP A1-associated silencer (Fig. 1). Based on an in vitro study, SF2/ASF and hnRNP A1-associated hypotheses may not be mutually exclusive. Depletion of hnRNP A1 has been shown to restore SMN2 exon 7 inclusion in an engineered cell line lacking SF2/ASF. These findings support C6U-associated gain of hnRNP A1 motif and clearly demonstrate the prominent role of hnRNP A1 in SMN2 exon 7 splicing in vivo. Recent studies revealed interactions of hnRNP A1 with additional sites located within SMN intron 7 (Fig. 1 and described below). Supporting the C6U-associated gain of an inhibitory element coupled with the loss of a stimulatory element, bifunctional antisense oligonucleotides (ASOs) that blocked C6U region and presumably recruited stimulatory factors promoted SMN2 exon 7 inclusion. Recently, Sam68 has been proposed to be an additional C6U-interacting factor responsible for SMN2 exon 7 skipping. These findings are in agreement with the role of an extended inhibitory context (Exinct) that spans from 3rd to 15th positions of exon 7. Interestingly, hnRNP Q was shown to stimulate SMN2 exon 7 inclusion by binding to C6U region that falls within Exinct. It is possible that low levels of exon 7 inclusion in SMN2 mRNA is maintained by factors such as hnRNP Q through interaction with a hotspot like Exinct.

In vivo selection of entire SMN1 exon 7 further supported the presence of Exinct and revealed two additional regulatory regions within exon 7. These are: conserved tract (a positive cis-element) in the middle of exon 7 and 3'-cluster (a negative cis-element) located toward the end of the exon (Fig. 1). Presence of Exinct, conserved tract and 3'-cluster was independently corroborated by a complementary approach of an antisense microRNA that creates a weak 3' ss. Indeed, results of in vivo selection of entire exon 7 affirmed intron definition model as a majority of selected sequences carried an A-to-G substitution (A54G) at the last position on exon 7. Further, A54G mutation in SMN2 promoted exon 7 inclusion despite the loss of additional positive cis-elements at the 3' ss. The likely cause of the stimulatory effect of A54G mutation was the strengthened 5' ss that allowed an efficient U1 snRNP recruitment through a strong RNA:RNA duplex formed between U1 snRNA and the 5' ss (abbreviated as “U1:5' ss duplex”). Of note, A54G mutation increases the size of U1:5' ss duplex by two base pairs: G:G and G:U. In a counter experiment, a mutant U1 snRNA that increased U1:5' ss duplex by five base pairs also restored SMN2 exon 7 inclusion. Similar to A54G mutation, mutant U1 snRNA promoted SMN2 exon 7 inclusion despite the simultaneous loss of additional positive cis-elements responsible for strengthening of the 3' ss. Notably, mutant U1 snRNA that effectively defined intron 7 produced a disproportionate increase in intron 6-retained intermediate due to the existing weak 3' ss. These results are perfectly in line with the intron definition model,

**Intron Retention Supports Intron Definition Model**

Minigenes containing shortened intron 6 of SMN1 and SMN2 recapitulate the splicing pattern of endogenous genes and are useful tools for detection of intron-retained intermediates. In addition to predominant exon 7-included products, SMN1 minigene produces detectable level of intron 7-retained intermediates (Fig. 1B). These observations suggest that the 5’ ss of exon 7 is weak and inclusion of exon 7 from SMN1 is primarily (if not exclusively) driven by the removal of intron 6, supporting the intron definition model. Thus, it is possible that the loss of intron 6 definition due to C6U mutation (that creates a weak 3' ss) in SMN2 coupled with an inherently weak 5' ss causes exon 7 skipping.

If the intron definition mechanism were to hold true, strengthening of the 5’ ss alone would guarantee SMN2 exon 7 inclusion even in the presence of a very weak 3' ss. Indeed, results of in vivo selection of entire exon 7 affirmed intron definition model as a majority of selected sequences carried an A-to-G substitution (A54G) at the last position on exon 7. Further, A54G mutation in SMN2 promoted exon 7 inclusion despite the loss of additional positive cis-elements at the 3' ss. The likely cause of the stimulatory effect of A54G mutation was the strengthened 5' ss that allowed an efficient U1 snRNP recruitment through a strong RNA:RNA duplex formed between U1 snRNA and the 5' ss (abbreviated as “U1:5' ss duplex”). Of note, A54G mutation increases the size of U1:5' ss duplex by two base pairs: G:C and G:U. In a counter experiment, a mutant U1 snRNA that increased U1:5' ss duplex by five base pairs also restored SMN2 exon 7 inclusion. Similar to A54G mutation, mutant U1 snRNA promoted SMN2 exon 7 inclusion despite the simultaneous loss of additional positive cis-elements responsible for strengthening of the 3' ss. Notably, mutant U1 snRNA that effectively defined intron 7 produced a disproportionate increase in intron 6-retained intermediate due to the existing weak 3' ss. These results are perfectly in line with the intron definition model,
Long-distance interactions (LDIs) involving deep intronic sequences could play pivotal role in pre-mRNA splicing. Depending upon the nature, such interactions could enhance or slow intron definition process. First LDI within SMN2 intron 7 was discovered through an ultra-refined antisense microwalk in which two 14-mer ASOs (F14 and L14) annealing to targets differing by a single nucleotide produced antagonistic effects on SMN2 exon 7 splicing. F14 promoted SMN2 exon 7 inclusion by sequestering a 14-nt long sequence that starts with a cytosine residue at the 10th intronic position (10C), whereas L14 that did not block 10C promoted SMN2 exon 7 skipping. Further experiments confirmed that 10C makes a negative LDI with downstream intronic sequences. Deletion of those downstream sequences reversed the negative effect of L14. Interestingly, 10C falls within the GC-rich sequence and occupies the first ISS-N1 position that is not a part of hnRNP A1 motif (Fig. 1C). In any case, 10C-associated LDI suggests a tightly regulated intron definition process in which motifs across intron 7 show mutual dependency.

Distributed Signals of Intron Definition

A number of reported cis-elements are likely to prevent intron 7 definition by weakening the 5’ ss of exon 7. These include an inhibitory terminal stem-loop structure (TSL2), intronic splicing silencer N1 (ISS-N1) and a GC-rich sequence (Figs. 1 and 2). Deletion of ISS-N1 fully restores SMN2 exon 7 inclusion even in the absence of positive exonic cis-elements at the 3’ ss of exon 7. Further, ASO-mediated blocking of ISS-N1 fully restores SMN2 exon 7 inclusion in SMA patient cells. ISS-N1 contains two putative hnRNP A1 binding sites (Fig. 2). It is likely that hnRNP A1 in complex with ISS-N1 renders the 5’ ss of exon 7 inaccessible. It is also possible that ISS-N1 is a complex regulatory element with multiple overlapping motifs. For instance, ISS-N1 partially overlaps with an 8-nt long inhibitory GC rich sequence that spans from 7th to 14th positions of intron 7. An 8-mer ASO targeting GC-rich sequence fully restores SMN2 exon 7 inclusion in SMA patient cells. Although mechanism by which GC-rich sequence exerts its negative impact is yet to be determined, the stimulatory effect due to blocking of GC-rich sequence supports the intron definition model. Of note, SMN exon 7 lacks specific intronic motif pairs that have been recently implicated in exon definition.

Long-distance interactions (LDIs) involving deep intronic sequences could play pivotal role in pre-mRNA splicing. Depending upon the nature, such interactions could enhance or slow intron definition process. First LDI within SMN2 intron 7 was discovered through an ultra-refined antisense microwalk in which two 14-mer ASOs (F14 and L14) annealing to targets differing by a single nucleotide produced antagonistic effects on SMN2 exon 7 splicing. F14 promoted SMN2 exon 7 inclusion by sequestering a 14-nt long sequence that starts with a cytosine residue at the 10th intronic position (10C), whereas L14 that did not block 10C promoted SMN2 exon 7 skipping. Further experiments confirmed that 10C makes a negative LDI with downstream intronic sequences. Deletion of those downstream sequences reversed the negative effect of L14. Interestingly, 10C falls within the GC-rich sequence and occupies the first ISS-N1 position that is not a part of hnRNP A1 motif (Fig. 1C). In any case, 10C-associated LDI suggests a tightly regulated intron definition process in which motifs across intron 7 show mutual dependency.
One may envision several other scenarios of LDIs involving sequences of intron 7, intron 6 and exon 7. For example, a SMN2 specific A-to-G mutation at the 100th position (A100G) of intron 7 creates an hnRNP A1 motif.29 Of note, A100G is a deep intronic mutation that does not create a cryptic splice site and yet imparts a significant impact on SMN2 exon 7 splicing. It is likely that hnRNP A1 bound to this motif (at the 100th position of intron 7) communicates with another hnRNP A1 molecule interacting with ISS-N1 and/or C6U location. Further, intron 6 has been reported to contain inhibitory Element 1 that encompasses binding site for PTB.59,60 It is probable that PTB bound to Element 1 communicates with another PTB molecule interacting with intron 7. The intron definition model would necessitate abrogation of these and other negative LDIs through stimulatory functions of intron-interacting factors. We have recently reported TIA1, a glutamine-rich domain containing RNA-binding protein, as the first splicing factor to activate SMN2 exon 7 inclusion by binding to intron 7.25 The stimulatory effect of TIA1 was able to counteract the inhibitory effect of PTB. Interestingly, site of TIA1 interaction within intron 7 is distinct from the traditional TIA1 binding sites located immediately downstream of 5′ ss. TIA1 interacts with U-rich clusters (URC1 and URC2) downstream of ISS-N1 (Fig. 2). Such interaction is not consistent with the exon definition model that generally requires a host of stimulatory cis-elements adjacent to splice sites. Of note, there is substantial sequence conservation within URC1 and URC2 between human and mouse genes. However, non-conserved residues seem to enforce TIA1 specificity in SMN intron 7.25 Consistently, a TIA1 Knockout mouse cell line showed no effect on splicing of Smn exon 7.25 Binding site of TIA1 is situated adjacent to Element 2, an earlier reported stimulatory element within intron 7.62 It is likely that splicing activators interacting with Element 2 as well as with other yet undefined motifs within intron 7 communicate with TIA1. Once again, such communications would be consistent with intron definition model.

Concluding Remarks

The exon definition model has been considered as the conventional mechanism of splicing of small exons flanked by large intronic sequences.29-32 The simultaneous recognition of both splice sites of an internal exon is the minimum prerequisite for this model. Consistently, a mutation within an exon leading to its skipping provides a likely example of the exon definition model. However, the competing intron definition model could be applicable in situations where intronic sequences appear to play a significant role in splice site selection. Employing SMN exon 7 splicing as one of the best-studied systems, we propose that a small exon flanked by large intronic sequences could be spliced by the intron definition mechanism. The strongest support for the intron definition model came from intronic motifs, whose blocking promoted SMN2 exon 7 inclusion.27 The independently validated stimulatory impact of the ASO-mediated intron definition is so enormous that the approach has a potential for the treatment of SMA.27,55,63-65 Additional support of the intron definition model came from the observation that TIA1 restored SMN2 exon 7 inclusion by binding to intronic sequences away from the 5′ ss. The intron definition model is further supported by the presence of the disproportionately trapped intron-retained intermediates that are uncharacteristic of the exon definition model. Based on the evidence of a complex regulatory network and LDIs within SMN2 intron 7, one may envision intronic motifs as the powerhouse of splicing regulation.

Splicing is coupled with transcription, capping, micro-RNA processing, polyadenylation and nuclear export.66-71 Large introns may encompass critical information for several layers of regulation that links splicing to the above mentioned processes. Many unrelated small compounds have been reported to modulate SMN exon 7 splicing.23 Mechanisms of action of these compounds remain unknown. Not limited by the constraints of a size, intron definition provides an attractive model to accommodate the role of a large number of factors, small compounds and structural rearrangements in pre-mRNA splicing. It remains to be seen if compelling evidence supporting the intron definition model for a small exon in SMN is part of a larger trend of pre-mRNA splicing regulation in higher eukaryotes. Overall, the intron and exon definition models may not be mutually exclusive. It is likely that an intron definition process in SMN leads to an exon definition process. However, there are several missing links and we are rather at the very early stages of understanding the mechanism of interconnectivity between these two very attractive and simplistic models.

Acknowledgments

Work in Singh group is supported by a grant from United States National Institutes of Health (R01NS055925). R.N.S. acknowledges support of Salisbury Endowment at Iowa State University, Ames, IA.

References

3. Wei CD, Kong L, Sumner CJ. The genetics of spinal muscular atrophies. Curr Opin Neurol 2010; 23:450-8; DOI: 10.1097/WCO.0b013e32833e1765.
5. Lorson CL, Halhnen E, Androphy EJ, Wirch B. A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy. Proc Natl Acad Sci USA 1999; 96:6307-11; DOI: 10.1073/pnas.96.11.6307.
In winkler c, egger g, gradl d, meister g, giegerich m, wedlich d, et al. reduced u snrnp assembly causes motor axon degeneration in an animal model for spinal muscular atrophy. genes dev 2005; 19:2320-30; doi: 10.1016/j.sgl.2005.04.002.


Shababi M, Mattis VB, Lorson CL. Therapeutics that directly increase the expression to treat spinal muscular atrophy. Drug News Perspect 2010; 23:475-82.


cartegni l, krainer AR. Disruption of an SF2/ASF-dependent exonic splicing enhancer in SMN2 causes spinal muscular atrophy in the absence of SMN1. Nat Genet 2002; 30:377-84; doi: 10.1038/ng454.

kashima T, manley JL. A negative element in SMN2 exon inhibits splicing in spinal muscular atrophy. Nat Genet 2003; 34:460-3; doi: 10.1038/ng1027.


