2004

In vivo selection reveals combinatorial controls that define a critical exon in the spinal muscular atrophy genes

Natalia N. Singh, *University of Massachusetts Medical School*
Elliot J. Androphy, *University of Massachusetts Medical School*
Ravindra N. Singh, *University of Massachusetts Medical School*

In vivo selection reveals combinatorial controls that define a critical exon in the spinal muscular atrophy genes

NATALIA N. SINGH,[1] ELLIOT J. ANDROPHY,[1,2] and RAVINDRA N. SINGH[1,2]

[1]Department of Medicine and [2]Program in Molecular Genetics and Microbiology, University of Massachusetts Medical School, Worcester, Massachusetts 01605-2324, USA

ABSTRACT

Humans have two near identical copies of the *survival of motor neuron* (*SMN*) gene, *SMN1* and *SMN2*. In spinal muscular atrophy (SMA), *SMN2* is not able to compensate for the loss of *SMN1* due to an inhibitory mutation at position 6 (C6U mutation in transcript) of exon 7. We have recently shown that C6U creates an extended inhibitory context (Exinct) that causes skipping of exon 7 in *SMN2*. Previous studies have shown that an exonic splicing enhancer associated with Tra2 (Tra2-ESE) is required for exon 7 inclusion in both *SMN1* and *SMN2*. Here we describe the method of in vivo selection that determined the position-specific role of wild-type nucleotides within the entire exon 7. Our results confirmed the existence of Exinct and revealed the presence of an additional inhibitory tract (3′-Cluster) near the 3′-end of exon 7. We also demonstrate that a single nucleotide substitution at the last position of exon 7 improves the 5′ splice site (ss) such that the presence of inhibitory elements (Exinct as well as the 3′-Cluster) and the absence of Tra2-ESE no longer determined exon 7 usage. Our results suggest that the evolutionary conserved weak 5′ ss may serve as a mechanism to regulate exon 7 splicing under different physiological contexts. This is the first report in which a functional selection method has been applied to analyze the entire exon. This method offers unparallel advantage for determining the relative strength of splice sites, as well as for identifying the novel exonic *cis*-elements.

Keywords: SMN; alternative splicing; in vivo selection; SMA

INTRODUCTION

Alternative pre-mRNA splicing is an important means of gene regulation in higher eukaryotes (Black 2003). To serve the broader purpose of development and tissue-specific regulation, alternative splicing has co-evolved with the degeneration of discrete and defining signals: the 5′ splice site (ss), 3′ ss, polypyrimidine tract, and branch point. In addition to the spliceosome, which represents the most complex macromolecular machines known (Nilsen 2003), the control of exon definition resides with other factors. Serine–arginine-rich proteins (SR proteins), SR-like proteins, and heterogeneous nuclear ribonucleoproteins (hnRNPs) are among many nonspliceosomal factors that mediate exon definition (Gravely 2000; Hastings and Krainer 2001; Cartegni et al. 2002; Dreyfuss et al. 2002; Maniatis and Tasic 2002). These proteins bind to pre-mRNA sequences called exonic or intronic splicing enhancers (ESEs or ISEs) and silencers (ESSs or ISSs). Enhancers and silencers promote or suppress splice-site selection, respectively. The organization of these regulatory sequences within pre-mRNAs and the intracellular concentrations of the transacting factors that bind them modulate alternative splicing (Maniatis and Tasic 2002). In some cases, creation of nonsense mutations causes exon skipping, suggesting that pre-mRNA splicing coordinates with the nuclear translational machinery (Wang et al. 2002). Further, pre-mRNA splicing has been shown to communicate with the transcriptional (Proudfoot et al. 2002) and the transport machineries (Reed and Hurt 2002). Conditions of stress and the extracellular signals also affect the pattern of alternative splicing (Black 2003; Rothrock et al. 2003). Despite the complexities in pre-mRNA splicing including exon definition, two methods, one based on score matrices (Cartegni et al. 2003) and the other based on computational algorithms (Fairbrother et al. 2002), have been used to predict ESE elements. Although the latter predicts a wide range of ESEs as compared to the former method, it excludes the role of exonic sequences located toward the termini as well as those motifs that are more than 6 nt long. In both methods, the degree of par-

Reprint requests to: Ravindra N. Singh, Department of Medicine (LRB 326), University of Massachusetts Medical School, 364 Plantation Street, Worcester, MA 01605-2324, USA; e-mail: Ravindra.Singh@umassmed.edu; fax: (508) 856-6797.

Article and publication date are at http://www.rnajournal.org/cgi/doi/10.1261/rna.7580704.

ticipation of a particular ESE and the modulating role of an overlapping signal remains largely unpredictable. In some cases, modulatory activity is affected by RNA secondary structure that provides accessibility to the splice sites and/or constitutes binding sites for interacting proteins (Grover et al. 1999; Tu et al. 2000; Damgaard et al. 2002; Buratti et al. 2004). However, because of the dynamic nature of pre-mRNA splicing (Staley and Guthrie 1998) and noncanonical base-pairing (Leontis et al. 2002), the role of RNA structure is not easily predictable. Thus, functional assays remain the best means to determine the role of sequence motifs and/or RNA structure in a specific context.

Spinal muscular atrophy (SMA), the second most common autosomal recessive disorder, is caused by the absence of the *Survival of Motor Neuron 1* (*SMN1*) gene (Lefebvre et al. 1995), which encodes a ubiquitously expressed 38-kDa protein that plays an essential role in cellular metabolism (Meister et al. 2002; Pellizzoni et al. 2002). A nearly identical copy of the gene, *SMN2*, fails to compensate for the loss of *SMN1* because exon 7 is skipped, producing a truncated protein SMNΔ7, which is unstable (Lorson et al. 1998). *SMN1* and *SMN2* differ by a critical C-to-T substitution (C6U transition in transcript of *SMN2*) at position 6 of exon 7. This single nucleotide substitution determines exon 7 usage (Lorson et al. 1999; Monani et al. 1999). Based on ESE score matrices, C6U has been implicated in abrogation of an ESE (Fig. 1; Cartegni and Krainer 2002). According to this model, disruption of the ESE (motif "CAGACAA") associated with SF2/ASF causes exclusion of exon 7 in *SMN2*. Supporting this model, UV-cross-linking showed specific interaction of SF2/ASF with exon 7 of *SMN1* but not with *SMN2*. Further, SF2/ASF promoted exon 7 inclusion in an in vitro splicing reaction (Cartegni and Krainer 2002). Contradicting the ESE model, a recent report proposed that C6U creates an ESS motif ("UAGACA") associated with hnRNP A1, which causes skipping of exon 7 (Fig. 1; Kashima and Manley 2003). It was shown that reduction of hnRNP A1 in HeLa cells using RNA interference promoted exon 7 inclusion in *SMN2*. Through in vitro UV-cross-linking, hnRNP A1 was found to bind exon 7 of *SMN2* but not *SMN1*.

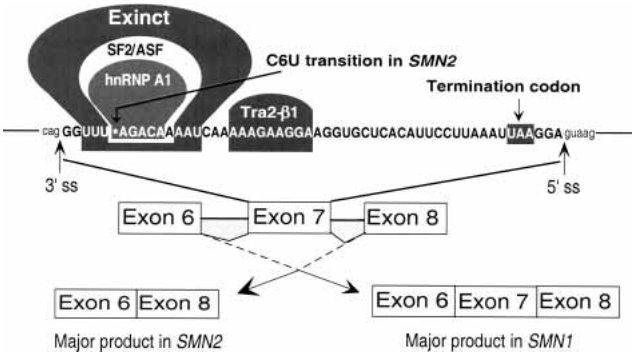

FIGURE 1. Schematic diagram of *SMN* construct containing C6U mutation within exon 7 and the preferred splicing pathways. Capital letters represent exon 7 sequence. Lowercase letters represent intronic sequences. Tra2-ESE is located in the middle of exon 7 (Hofmann et al. 2000), whereas SF2/ASF-ESE is located toward the 3′ ss of exon 7 (Cartegni and Krainer 2002) and overlaps with hnRNP A1-ESS (Kashima and Manley 2003). Earlier reports suggest that C6U (indicated by a star) abrogates SF2/ASF-ESE and/or creates hnRNP A1-ESS, producing exon 7-skipped product in *SMN2*. We have recently shown that C6U creates an extended inhibitory context (Exinct) that causes exon 7 skipping (Singh et al. 2004). (ss) Splice site.

Exon 7 spans 54 nt that also harbor the translation termination codon (positions 49–51; Fig. 1). The last position of exon 7 is an adenosine residue (A54), which is conserved between the two *SMN* genes. The presence of A54 places exon 7 in the minor group of internal exons that lack a 3′-end G residue (Burge et al. 1999). Exon 7 is known to have a weak 3′ ss (Lim and Hertel 2001), probably because of its suboptimal polypyrimidine tract. An improved polypyrimidine tract promoted *SMN2* exon 7 inclusion; however an AG-rich ESE (motif "AAAGAAGGA") associated with SR-like protein Tra2 was still required (Fig. 1; Lorson and Androphy 2000; Hofmann et al. 2000). Elevated expression of Tra2 and its associated proteins has been shown to promote exon 7 inclusion in *SMN2* (Hofmann et al. 2000; Hofmann and Wirth 2002; Young et al. 2002). Sodium butyrate, which induces SR proteins, has been shown to stimulate inclusion of exon 7 in *SMN2* (Chang et al. 2001). Treatment of cells with a high concentration of the phosphatase inhibitor sodium vanadate also promotes inclusion of exon 7 in *SMN2* (Zhang et al. 2001). Compounds such as aclarubicin (Andreassi et al. 2001) and valproic acid (Sumner et al. 2003) have also been shown to promote exon 7 inclusion in *SMN2*. In addition, *cis*-elements present in intron 6 and intron 7 have been shown to modulate exon 7 inclusion (Miyajima et al. 2002; Miyaso et al. 2003). Further, bifunctional oligos (Skordis et al. 2003) and modified oligos (Cartegni and Krainer 2003) that bind to exonic sequence at the 3′ ss have been shown to promote exon 7 inclusion in *SMN2*. These results suggest that regulation of exon 7 splicing could be controlled by multiple factors that interact with pre-mRNA. This hypothesis is consistent with the differential splicing of exon 7 recently reported in discordant SMA families (Helmken et al. 2003).

Iterative in vivo selection represents a high throughput screen that allows parallel sampling of a very large population of molecules for a desired property (Coulter et al. 1997). Using in vivo selection and a series of mutations, we have recently shown that C6U creates an extended inhibitory context (named "Exinct"; Fig. 1; Singh et al. 2004). Our results suggest that several overlapping sequence motifs collaborate to regulate the alternative splicing of exon 7. Here, we report the use of a modified version of iterative in vivo selection that identified the role of sequences beyond small motifs. Our method was designed to randomly create point or clustered point mutations that allowed testing the position-specific role of all wild-type nucleotides within the entire exon 7. These mutations were generated through partial

randomization. The distribution of mutations within enriched exon 7 identified the hyper- and hypomutable regions that reflected the negative and the positive nature of the wild-type nucleotides, respectively. This approach revealed the suboptimal nature of the 5′ ss, which was improved by an adenosine-to-guanosine substitution at the last position of exon. In addition to confirming the presence of Exinct, it also identified an additional inhibitory element located near the 3′ end of exon 7. Our results reveal many novel features of combinatorial control that defines a critical exon in the spinal muscular atrophy genes.

RESULTS

In vivo selection of entire exon 7

To analyze the position-specific role of nucleotides within exon 7 on its splicing, we adapted an in vivo approach that allowed selective amplification of exon 7 sequences that promoted its own inclusion (Fig. 2). To the best of our knowledge, in vivo selections of more than 25-nt-long exonic stretches have not been reported. This plan required partial randomization that preserved the wild-type character of the molecule and at the same time determined the mutability of the wild-type residues. There is no specific rule for the degree of partial randomization in selection experiments; overrandomization may create artificial sequence motifs, whereas underrandomization can produce false conserved positions. A consensus 30% randomization has been traditionally used in most if not all such selection experiments (Bartel et al. 1991; Baskerville et al. 1995; Saldanha et al. 1996; Guo et al. 2000; Singh et al. 2002). In this case, each position is doped with 70% wild-type and 10% each of the non-wild-type nucleotides. This approach generates an initial pool of sequences that harbors a range of substitutions with the maximum being close to the 30%. We chose to perform in vivo selection of the entire exon 7 of *SMN1* because this would determine novel *cis*-elements in addition to independently confirming the role of the earlier described *cis*-elements. This would also determine the relative strength of splice sites, as impact of the 5′ ss of exon 7 on alternative splicing of exon 7 is not known. Because C6U is the only critical difference between the two *SMN* genes (Lorson et al. 1999; Monani et al. 1999), our approach will identify mutations that compensate for the inhibitory effects of C6U.

One of the important highlights of our approach was the selective amplification of the entire exon 7 using spliced products (exons 6–8) as the template for successive rounds of in vivo selection (Fig. 2). Our experimental design allowed removal of exon 7 exclusively, followed by its reinsertion into the minigene-splicing cassette (pBxT2) without carrying over additional nucleotides from the flanking exons 6 and 8. This all became possible because of a second amplification step (Fig. 2B), during which a few substitutions on both sides of the amplified exon 7 were introduced to create the recognition motif for a new restriction endonuclease BsaXI that cleaves upstream and downstream of its recognition site. This strategy removed the flanking exon 6 and exon 8 sequences from the amplified DNA while retaining the entire exon 7. In the second amplification step, we also substituted 2 nt within the last three positions of exon 6 and 1 nt within the first three positions of exon 8, restoring the flanking intron 6 and intron 7 sequences, respectively. The exon 7 fragment generated after BsaXI digestion was inserted back into a specially designed cassette pBxT2 to restore the original minigene splicing cassette (Fig. 2B).

Sequencing of a statistically significant number of clones from the initial pool confirmed partial randomization at every position of exon 7 (not shown). Some positions were more random than the others (Fig. 3A). This is typical for any randomized oligo synthesis in which exact percentage of bases at any particular position cannot be controlled. The average randomization per molecule was 33%, close to the desired value. Consistent with the probability distribution, the initial pool harbored sequences with a broad range of substitutions (Fig. 3B). Of note, partial randomization produces only a fraction of the possible variants. For example, our experiment began with ~10^{11} unique molecules that represent a fraction of the "saturation of sequence space," which is defined as theoretically possible variants (~10^{32}) that could be generated upon complete randomization of 54-nt-long exon 7. However, synthesis of all variants is neither desirable nor possible.

The initial pool was subjected to four successive rounds of in vivo selection. Because of the partial rather than complete randomization, we did not encounter aberrant splicing owing to the creation of the cryptic splice sites. This was reflected by discrete bands (exon-included and exon-excluded) at each cycle (Fig. 4A, lanes 5–9). The final pool (pool 4) included exon 7 about 200-fold more efficiently than the initial pool, ~13-fold higher than the corresponding value for *SMN1*. The sequences of 59 randomly chosen clones from pool 4 are listed in Figure 5. Consistent with the initial pool (Fig. 3), the final pool harbored isolates with a broad range of substitutions. The in vivo splicing patterns of 11 representative clones are shown in Figure 4B, though we have determined the splicing pattern of all clones shown in Figure 5 (not shown). In agreement with the high inclusion efficiency of pool 4, we could not detect the exon 7-excluded product for any of the isolates. These results suggest that the inclusion of exon 7 was promoted by substitutions that improved the recognition of both splice sites.

Mutability of exonic nucleotides predicts the position-specific role

Results of our in vivo selection could be best explained by the mutability of exonic sequences. A highly mutable posi-

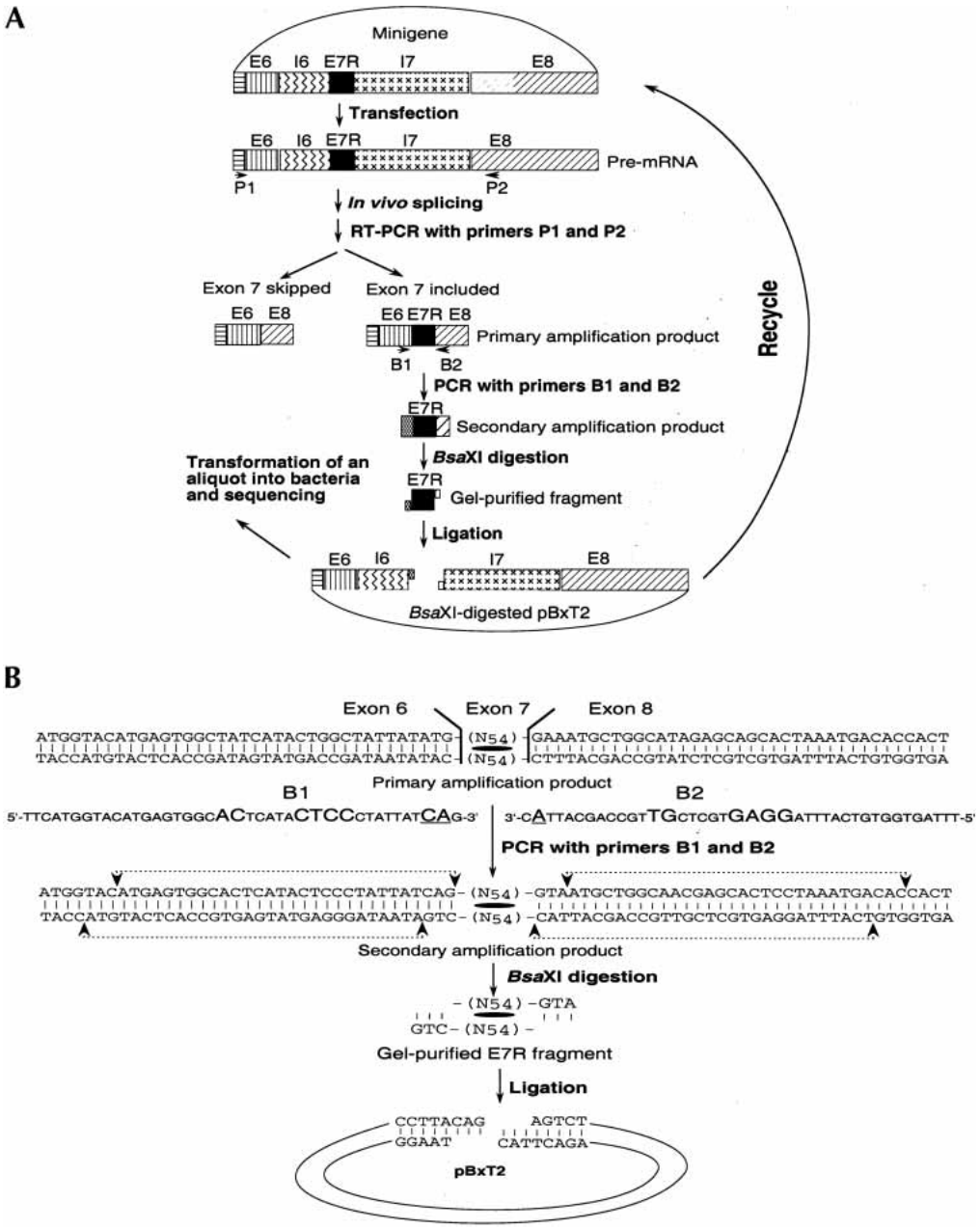

FIGURE 2. Strategy for the iterative in vivo selection of the entire exon 7. (*A*) Selection procedure (diagram not to the scale). pCI vector backbone is shown as a half circle. Exon 6 (E6), Intron 6 (I6), partially random exon 7 (E7R), intron 7 (I7), and exon 8 (E8) sequences are marked differently. (*B*) Details of the second PCR amplification step and ligation. Primers B1 and B2 anneal to constant sequences in E6 and E8, respectively. B1 and B2 contain mutations marked in capitalized and underlined letters that create BsaXI site and restore intronic sequences, respectively. Upon digestion with BsaXI, the intact E7R is released with the 3′ overhangs that correspond to the complementary intronic sequences of the 5′ overhangs of BsaXI-digested pBxT2. The insertion of E7R into pBxT2 plasmid restores the minigene splicing cassette.

tion is considered as inhibitory for exon 7 inclusion, whereas the least mutable position is considered as stimulatory (Deminoff et al. 1995; Guo et al. 2000). We calculated mutability of wild-type residues of exon using an equation that corrects selection results for the bias of the initial pool (Fig. 6; Deminoff et al. 1995). Negative and positive bars indicate the stimulatory and inhibitory nature of the wild-type nucleotides, respectively. Positions with mutability value zero are considered as neutral, although we did not have many such positions. A small number of neutral positions could be also obtained because of the lack of the saturation of sequence space (defined as theoretically possible variants upon complete randomization). Thus in a stretch of conserved positions, a neutral position could be considered as conserved. Likewise, in a stretch of mutable positions, a neutral position could be considered as mu-

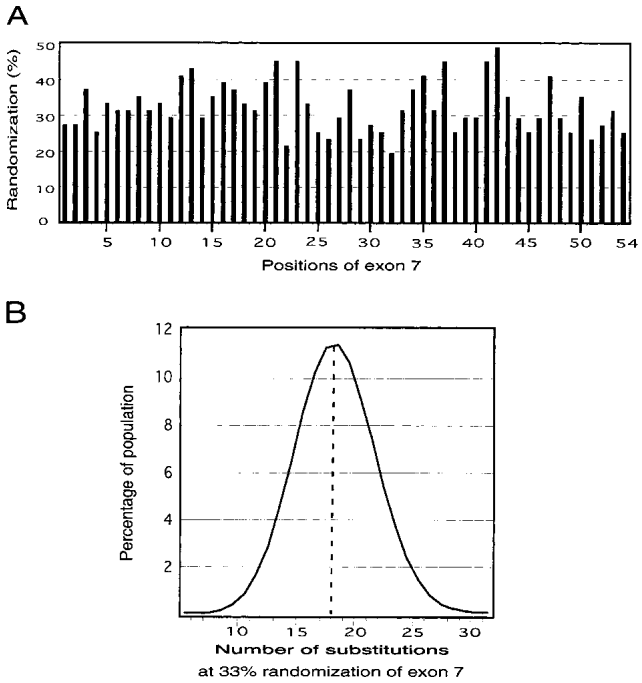

FIGURE 3. Analysis of the initial pool based on 51 sequences (not shown). (*A*) Percentage of randomization at every position of exon 7 has been shown. As evident from the bars, not every position was equally random. These values were taken into account for mutability calculations as shown in Figure 6. Average randomization of the initial pool was found to be 33%. (*B*) The initial pool will have a variable population of substitutions. This could be calculated using binomial distribution (Baskerville et al. 1995) as follows: $F = \{L!/[n!(L - n)!]\}p^n(1 - p)^{(L - n)}$, where F is the fraction of the population containing n mutations, L is the length of the molecule, p is the fraction of non-wild-type nucleotide/position, and n is the actual number of mutations per template. For example, at 33% randomization ($p = 0.33$) of 54-nt-long exon 7 ($L = 54$), only 11.4% of the total population will have 18 substitutions ($n = 18$). As evident from distribution, the population will encompass a bell-shaped distribution of lower and higher substitutions as well.

table. The highly conserved nature of position 1 is apparent by its mutability value close to −1, whereas the least conserved (or the highly mutable) position 54 has the mutability value of +17.6. The cutoff values for the conserved and the mutable positions have been taken as −0.2 and +0.2, respectively. Both the positive and the negative cutoff values are supported by mutations and/or deletions described below. Because of the exceptionally high mutability of position 54, other positive values may have been skewed. However, this did not affect the overall significance of the positive and the negative *cis*-element. In fact, the strength of in vivo selection with partial randomization lies in the fact that it tends to simultaneously reveal the inhibitory and the stimulatory stretches despite the skewed selection at certain positions.

Based on the mutability plot, more than 70% of the conserved residues are located in the middle of exon, forming a conserved tract that covers about ~50% of exon 7 (Fig. 6). This long conserved tract is more noticeable in isolates with less than 33% substitutions. Mutable residues (with values from +1 to +17.6) are located throughout the molecule, although the majority of them are concentrated toward the ends of the exon forming an Exinct near the 3′ ss and a cluster near the 5′ ss. This again is more noticeable in isolates with less than 33% substitutions. Consistent with the stimulatory and the inhibitory nature of *cis*-elements, mutations within the conserved and the mutable tracts promoted exon 7 exclusion and inclusion, respectively (discussed below).

Selection of terminal guanosine residues exceeds the statistical probability

Terminal exonic positions interact with a number of protein factors at different steps of splicing. However, because of the inherent difficulty in precise iterative amplification of an entire exon, functional selection assays have not been used to determine their relative significance. Despite the prevalence of G residues at both terminal exonic positions

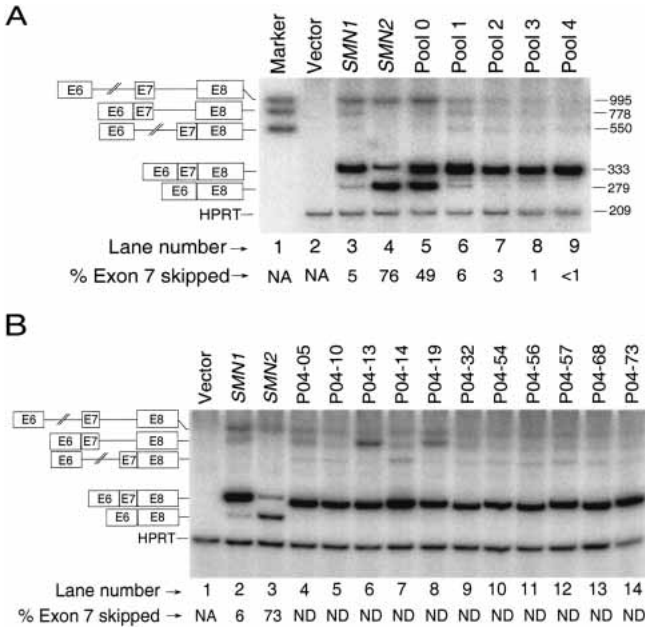

FIGURE 4. Comparison of splicing efficiency of different pools and selected sequences. (*A*) Comparative splicing patterns of different minigene variants and selected pools. Because of the shortened intron 6, precursor mRNA (995 bp) along with splicing intermediates (except lariats) were detectable. The 778-, 550-, 333-, and 279-bp products correspond to intron 7-retained, intron 6-retained, fully spliced, and exon 7-skipped products, respectively. The *HPRT* gene product was simultaneously amplified as an internal control (209-bp product). The percent of exon 7 skipping was calculated from the total value of exon 7-included and -excluded products. Abbreviations E6, E7, and E8 stand for exon 6, exon 7, and exon 8, respectively. (*B*) The in vivo splicing pattern of 11 representative clones from pool 4 (boxed in Fig. 5). The spliced intermediates and products are the same as marked in Figure 4A. No exon 7-skipped product was detected in any isolate.

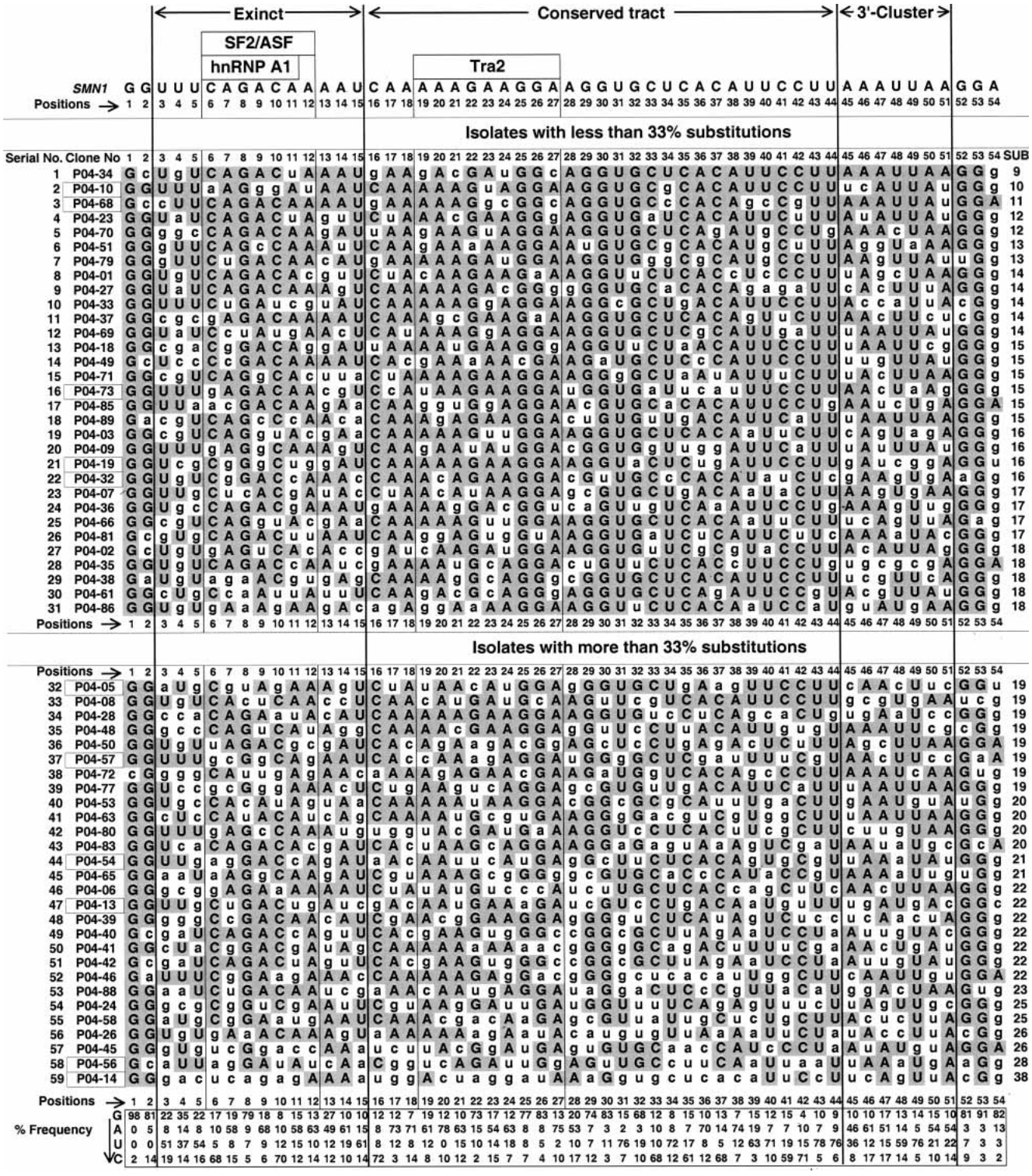

FIGURE 5. Sequences from in vivo selection. Fifty-nine clones from pool 4 are shown. The wild-type residues are written in capital letters that are also shadowed. Non-wild-type residues are written in lowercase letters. The number of substitutions per molecule is given on the *right* (marked as SUB). The percentage frequency of selected nucleotides at each position is given at the *bottom*. All clones showed exon 7 inclusion better than *SMN1* (not shown).

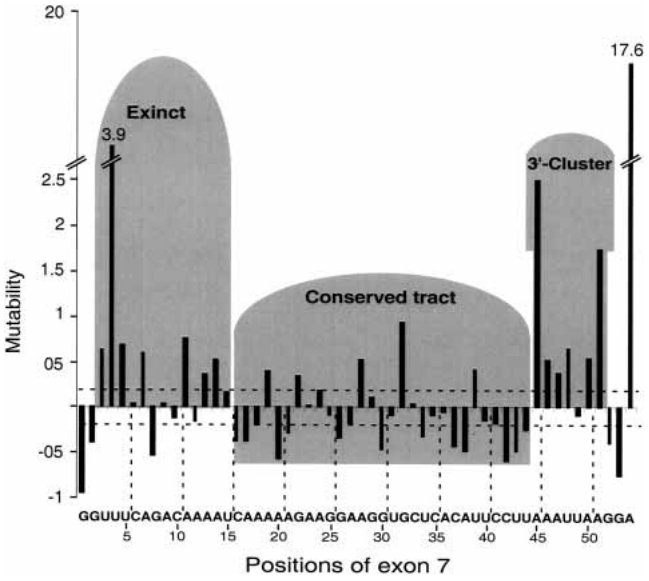

FIGURE 6. Mutability of residues based on the results of in vivo selection. The values of −1 and +17.6 represent the absolutely conserved and the least conserved residues, respectively. The dotted horizontal lines show the cutoff points with the mutability values of +0.2 and −0.2, corresponding to the mutable and the conserved residues, respectively. Based on the stretches of the mutable and the conserved residues, the core of Exinct, the 3′-Cluster, and the long conserved tract have been highlighted. The inhibitory nature of residues covering Exinct has been recently described (Singh et al. 2004). Consistent with the inhibitory nature of the 3′-Cluster, deletions and mutations in this region promoted exon 7 inclusion in *SMN2* (Fig. 7). Multiple mutations in the region of the conserved tract have been shown to cause exon 7 skipping in *SMN1* (Hofmann et al. 2000; Lorson and Androphy 2000). Novel mutations supporting the stimulatory role of the long conserved tract have been shown in Figure 8. The exceptionally high mutability of position 54 is consistent with the dominant effect of A54G substitution on exon 7 inclusion as shown in Figure 9.

(Burge et al. 1999), their impact on alternative splicing is not known. A complete randomization of position 1 with 25% of each of the four nucleotides should select 1G in ~50% of the instances for a constitutive exon (Burge et al. 1999). In this case 1A, 1C, and 1U should represent the remaining 50% of the molecules. Because our initial pool contained A, C, and U residues in ~28% of the cases (Fig. 3A), we should have selected these residues in ~14% of the instances if exon 7 were to fall into the category of a typical constitutive exon. However, the wild-type G at position 1 was retained in more than 98% of the isolates, suggesting the inhibitory role of other residues at this position. In agreement with the stimulatory role of 1G, in vivo selection of the first 6 nt of exon 7 retained 1G in all the isolates (Singh et al. 2004).

Similar to 1G, the preference of 2G also exceeded the statistical prevalence. Generally, the second position of constitutive exons has no preference for any particular nucleotide (Burge et al. 1999). We isolated 2G in ~81% of the instances, which is ~14% higher than the neutral position. The second highest preference was for a C residue at this position. Surprisingly, U had the lowest preference at this position, although an earlier in vitro selection with U2AF,[35] which interacts with the 3′ ss, had shown a strong preference for 2U (Wu et al. 1999). This may suggest interaction of other proteins that are specific to exon 7. The nature of residues selected at position 2 matched with the in vivo selection of the first 6 nt of exon 7 (Singh et al. 2004).

Another surprising result was the overwhelming selection of a non-wild-type G residue at the last position of the exon (54G). This is in agreement with the prevalence of the 3′ terminal G, which is represented in ~80% of the cases in humans (Burge et al. 1999). Because our initial pool contained 54G in ~11% of the variants, we should have selected 54G in ~35% of isolates if exon 7 were to behave like a typical constitutive exon. However, 54G was found in ~82% of the selected clones, suggesting that the wild-type 54A is inhibitory, which may be responsible for the weak 5′ ss of exon 7. This also implies that a weak 5′ ss is the limiting factor for exon 7 exclusion in *SMN2*.

In vivo selection confirms the presence of Exinct

Consistent with our recent report that revealed the presence of Exinct (Singh et al. 2004), in vivo selection of the entire exon 7 showed the longest tract of mutable residues at the 5′-end of exon 7. The core of Exinct covered 13 residues between positions 3 and 15 (Fig. 6). Eight positions (3, 4, 5, 7, 11, 13, 14, and 15) were highly mutable in this tract. Within Exinct, position 4 showed the highest mutability of 3.9 that was second to only position 54. In agreement with these results, mutations in this region have been previously shown to promote exon 7 inclusion in *SMN2* (Singh et al. 2004). Noticeably, a G residue at position 8 was retained in most of the selected clones, suggesting that wild-type 8G is not inhibitory. This residue could be important for providing the secondary contact to protein that interacts with upstream and/or downstream sequences. Residues at positions 10 and 12 could also provide similar secondary contacts. Consistent with the earlier results of single mutations (Lorson and Androphy 2000; Cartegni and Krainer 2002) and in vivo selection of a stretch of the first 6 nt of exon 7 (Singh et al. 2004), position 6 preferred the non-wild-type G residue to U or A residues. Guanosine was twice as preferred at position 4 than at position 6, although U4G mutation has been shown to be not as efficient as U6G (Singh et al. 2004). This highlights the nature of combinatorial control, in which other positions may have collaborated with 4G than could be achieved by single mutations. Our recent study has shown such collaborations through long-distance interactions within Exinct (Singh et al. 2004). Thus we infer that abrogation of Exinct through a combination of substitutions may be necessary for efficient inclusion of exon 7. This hypothesis is in agreement with the absence of selection of any specific ESE motif at the 3′ ss of exon 7.

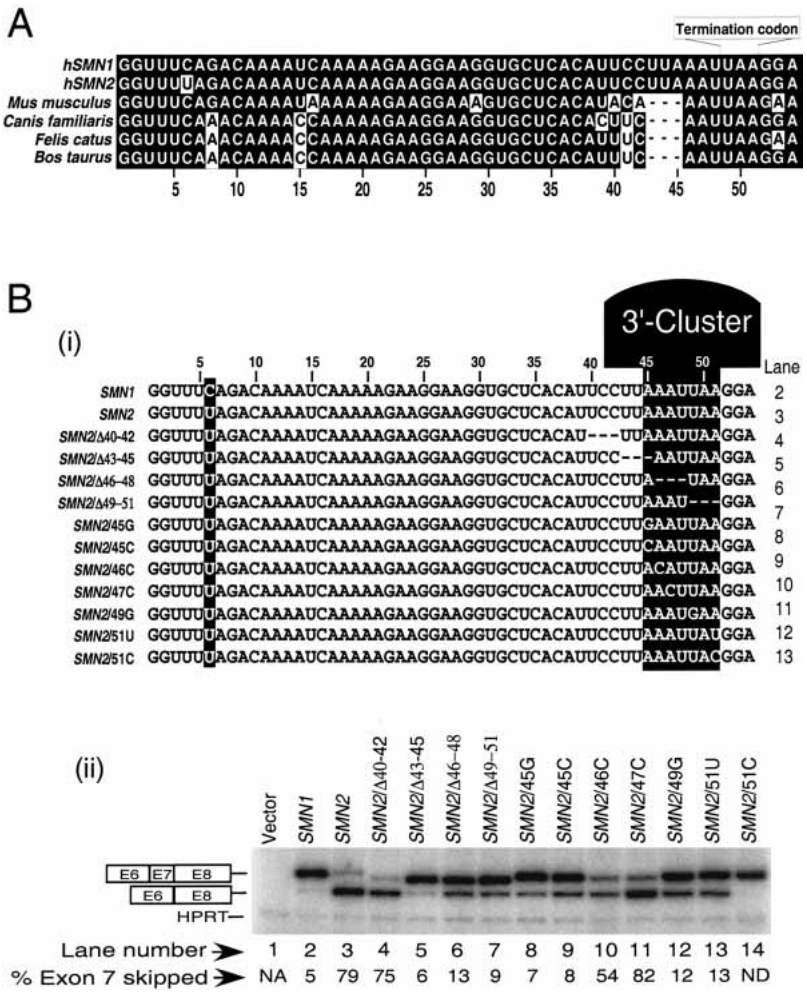

FIGURE 7. Effect of deletions or substitutions within the 3′-Cluster of exon 7 of *SMN2*. (*A*) A comparison of exon 7 sequences of mammalian *SMN*. The conserved positions are shaded. Triple dash (positions 43–45) indicates the deletion or absence of UUA codon in exon 7 of nonhuman *SMN* genes. Accession numbers of the above sequences are as follows: human *SMN1*, AC044797; human *SMN2*, AC010237; *Mus musculus*, AF131205; *Canis familiaris*, U50747; *Felis catus*, AY094503; *Bos taurus*, AF035323. (*B*) Mutations validating the inhibitory nature of the 3′-Cluster. (*i*) Sequences of *SMN2* mutants are shown. Dash indicates deletion. All mutations are located within the 3′-Cluster. (*ii*) Splicing pattern of mutants as shown in panel *i*. Numbers and nucleotides represent the positions and the types of substitution within exon 7. Spliced intermediates and products are the same as marked in Figure 4A.

The 3′-Cluster represents a novel inhibitory *cis*-element

The in vivo selection identified a number of mutable residues that form a 7-nt-long cluster toward the 3′-end of exon 7, which we call the "3′-Cluster." This cluster spans positions 45 through 51 and includes the translation termination codon. It also includes the last residue of a leucine codon (UUA between positions 43 and 45) that is not evolutionary conserved among mammalian exon 7 (Fig. 7A). Deletion of UUA promoted exon 7 inclusion in *SMN2*, whereas deletion of the preceding codon had no effect (Fig. 7B, lanes 4,5). This suggests a specific role of these nucleotides that were gained (or retained) during evolution. Validating the inhibitory nature of the 3′-Cluster, other triple deletions (or codon deletions) as well as many single mutations within this tract restored exon 7 inclusion in *SMN2* (Fig. 7B). Interestingly, a single A51C mutation restored exon 7 inclusion to the level of *SMN1*. Because of covariations in which selection at one position is driven by residues at other positions, the effect of all single mutations would not be the same. Consistently, some single mutations were more effective than the others. Supporting the inhibitory nature of the 3′-Cluster, an in vivo selection of a small stretch of sequence spanning this region preferred non-wild-type residues (data not shown).

The Tra2-β1-ESE is located within a long conserved tract

The double substitution G25U/G26U abrogates the Tra2-β1-ESE (between positions 19 and 27; Fig. 1) and has been shown to promote exon 7 skipping (Hofmann et al. 2000). In agreement with the role of Tra2-β1-ESE, none of the selected variants had G25U/G26U substitution (Fig. 5). In isolates with less than 33% substitutions, a purine-rich sequence in the middle of exon 7 was (mostly) conserved (Fig. 5). In some instances, pyrimidine residues within the Tra2-ESE were also selected, suggesting that pyrimidines are not inhibitory for Tra2 binding. Consistent with this observation, a recent report demonstrates that a Tra2-β1 binding site can contain a substantial number of pyrimidines (Stoilov et al. 2004). Validating the stimulatory role of the long conserved tract, many mutations within this tract led to exon 7 exclusion in *SMN1* (Fig. 8). Mutations used in our study are different than an earlier study, which also demonstrated the stimulatory role of sequences that overlap the long conserved tract (Lorson and Androphy 2000). The double mutations showed better effect than the single mutations, an observation also supported by Lorson and Androphy (2000). Surprisingly, the novel double mutation A36U/C37U completely eliminated exon 7 inclusion in *SMN1*, indicating the stimulatory role of the wild-type residues. Because A36/C37 fall outside the earlier described binding site of Tra2-β1 (Hofmann et al. 2000), another protein or protein complex may interact at this site. These results also suggest that the long conserved

tract contains many overlapping stimulatory *cis*-elements, some of which could be tissue specific.

Validation of the relative significance of the terminal positions

To validate the results of in vivo selection, we made substitution mutations of terminal residues that are generally ignored in computation programs (Fairbrother et al. 2002). The functional significance of G at position 1 was confirmed by G1H substitutions (H connotes A or C or U), which decreased exon 7 inclusion (Fig. 9B, lanes 4–6). The inhibitory effect was most pronounced with 1U, which produced less than 10% fully spliced products. The presence of 1A or 1C had an inhibitory effect as well, but to a lesser degree. The dominant effect of 54G was confirmed by its ability to compensate for the inhibitory effect of substitutions at different positions. For example, it overcame the negative effects of 1H substitutions. Furthermore, it compensated for the inhibitory effects of the C6U substitution in *SMN2*. Surprisingly, 54G compensated for the abrogation of Tra2-ESE in *SMN1* and *SMN2* (Fig. 9B, lanes 12,13). These results demonstrate that the 54G-mediated strong 5′ ss not only prevents exon 7 exclusion by promoting splicing of the downstream intron 7, but also ensures exon 7 inclusion by promoting splicing of the upstream intron 6. To the best of our knowledge, such an overriding effect of one exonic nucleotide has not been reported. Although our results confirmed the dominant effect of 54G, we also determined its limitations. For example, abrogation of the Tra2-ESE along with the G1U substitution in *SMN2* led to exon 7 exclusion despite the presence of 54G (Fig. 9B, lane 16). This suggests why none of the selected clones showed the simultaneous absence of both 1G and 54G. This also shows the influence of cooperative interactions between the two splice sites. Supporting the dispensable nature of 54G, we selected two clones with non-wild-type 54U (P04–19 and P04–05) and one clone with non-wild-type 54C (P04–13), all of which promoted exon 7 inclusion better than *SMN1* (Fig. 4B). Interestingly, all of these clones had multiple mutations within the 3′-Cluster (Fig. 5), supporting the inhibitory nature of other residues that contributed to the creation of the weak 5′ ss of exon 7.

DISCUSSION

Unique design allows in vivo selection of an entire exon

Here we report the successful application of the powerful method of iterative in vivo selection to identify the critical role of exonic positions that modulate the alternative splicing of exon 7 in *SMN* genes. Our work gains special significance because this is the first report of iterative in vivo selection of an entire exon. Our approach provides an unparalleled advantage in deciphering the impact of overlapping sequence motifs that interact with many factors during the dynamic process of pre-mRNA splicing. Following the rules of multitarget selection (Vant-Hull et al. 1998), the outcome showed complex properties in that preference for residues at a particular position was driven by a combination of events in which relative concentrations of interacting factors may have played a critical role. Our approach required not only the meticulous design of amplification and recycling protocols, but also the use of partial randomization that maintained the wild-type character of the molecule. This work was inspired by an earlier success in which a similar approach defined the position-specific role of nucleotides

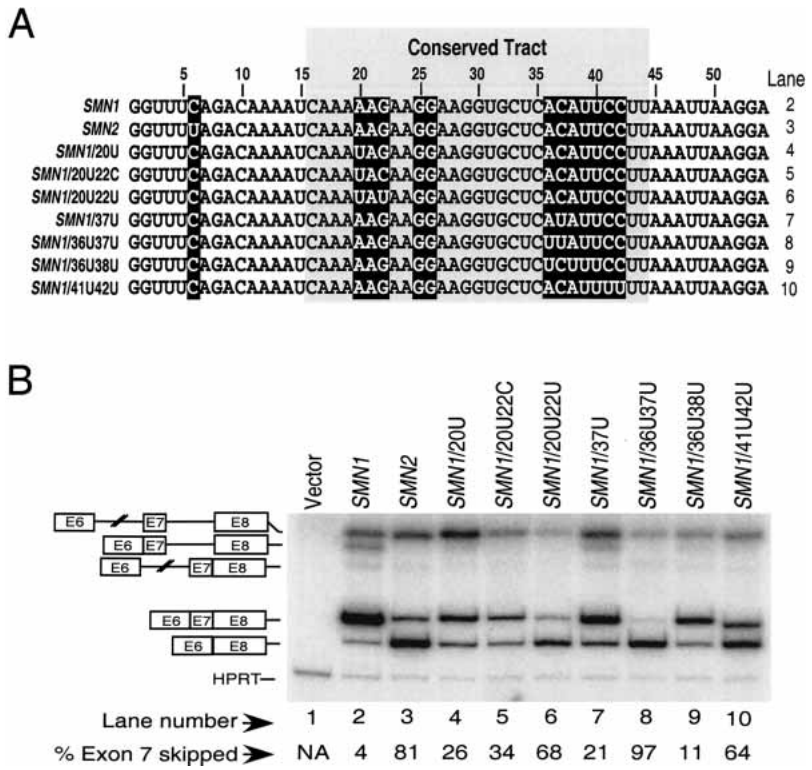

FIGURE 8. Effect of substitutions within the long conserved tract in the middle of exon 7 of *SMN1*. (*A*) Sequences of exon 7 mutants. All mutations were made in *SMN1*. Substitutions are restricted to the highlighted area. Based on an earlier study (Hofmann et al. 2000), critical positions (25G26G) of Tra2-β1 interaction have also been highlighted. (*B*) Splicing pattern of mutants as shown in panel *A*. Numbers and nucleotides represent the positions and the types of substitution within exon 7 of *SMN1*. Spliced intermediates and products are the same as marked in Figure 4A.

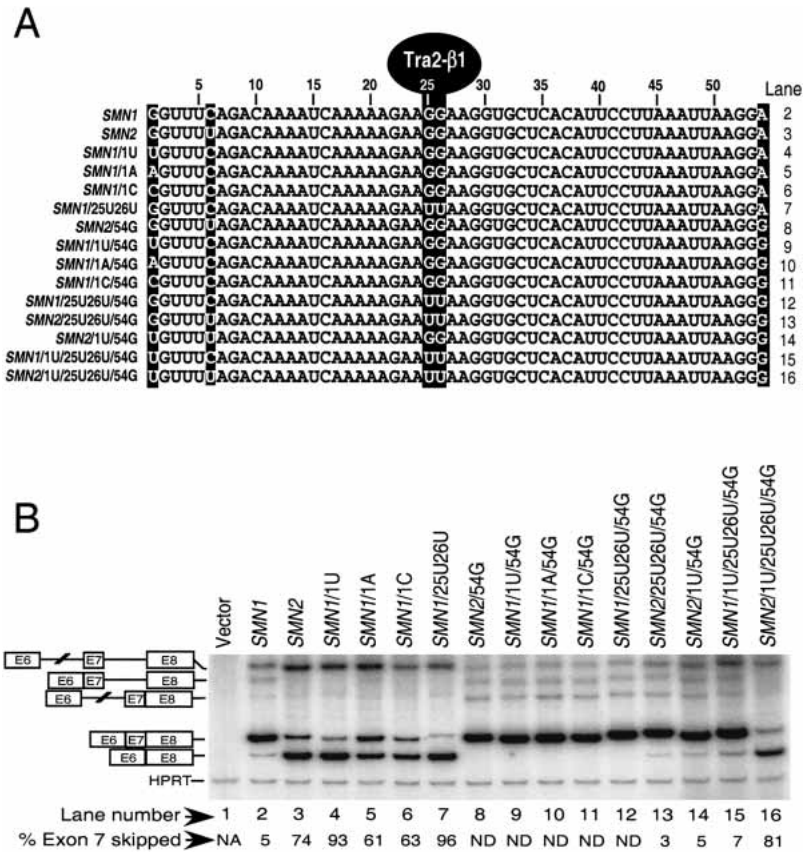

FIGURE 9. Mutations validating the dominant role of terminal positions. (*A*) Sequences of *SMN1* and *SMN2* mutants are shown. Substitutions are restricted to the highlighted area. Based on an earlier study (Hofmann et al. 2000), critical positions (25G26G) of Tra2-β1 interaction have been highlighted. (*B*) Splicing pattern of *SMN1* and *SMN2* mutants as shown in panel *A*. Numbers and nucleotides represent the positions and the types of substitution within exon 7. Spliced intermediates and products are the same as marked in Figure 4A.

widely separated on the primary RNA structure (Bartel et al. 1991). As expected from an in vivo experiment, the selected variants went through several checkpoints that are generally coupled with pre-mRNA splicing and RNA stability. These include transcription (Proudfoot et al. 2002), nonsense-altered splicing (Wang et al. 2002), nonsense-mediated decay (Maniatis and Reed 2002), and nuclear transport (Reed and Hurt 2002). Thus, the selected sequences, all of which showed no detectable amount of exon 7 excluded product (Fig. 5), represent a valuable archive for extracting the position-specific information within the entire exon 7.

Context-specific role of terminal exonic positions

In vivo selection of the entire exon 7 provided the first opportunity to simultaneously test the context-specific role of terminal exonic positions. The results of selection supported the conserved nature of the first nucleotide (1G) of exon 7, and, along with the site-directed mutations (Fig. 9), confirmed its critical role in the definition of this exon. In our previous report, we had shown the dominant role of 1G could not be replaced by any combination of substitutions within the first six positions of exon 7 (Singh et al. 2004). In addition to the interaction with U2AF^{35} (Wu et al. 1999), functions of 1G may include the interactions with proteins that participate in the second transesterification step (Collins and Guthrie 2000; Moore 2000; Chua and Reed 2001; Lallena et al. 2002). Control at the second step of splicing is consistent with the proofreading function of the spliceosomal machinery. We have previously shown that insertion of a G residue at intron 6/exon 7 boundary obliterates exon 7 inclusion (Singh et al. 2004). This clearly demonstrates that 1G executes a stimulatory role only in the context of a 54-nt-long exon 7, not the 55-nt-long exon 7. These results also highlight the context-specific features of *cis*-elements that are intimately tied to the size of exon 7. This further suggests that 1G in exon 7 does not merely perform a generic role as expected from an otherwise typical constitutive exon. We believe that absence of 1G in exon 7 creates a negative *cis*-element, and that is the reason why 1G could not be functionally replaced by mutations within the first six positions of exon 7 (Singh et al. 2004). This was also the reason why stimulatory 54G alone was not enough to drive selection of isolates containing G1H substitutions. Supporting this argument, a G1C isolate had 10 substitutions within the first 16 positions of exon 7, in addition to the A54G substitution (Fig. 5).

One of the important findings of in vivo selection was the role of the noncoding 54A, the last position of exon 7. The overwhelming selection of A54G substitution suggests the highly inhibitory nature of wild-type 54A. Despite guanosine comprising ~80% of the last position of internal exons (Burge et al. 1999), the role of this residue is not fully understood. In addition to its involvement in base-pairing with U1 snRNA, 54G may have other functions. For example, a recent study with the yeast homolog of the U1C protein showed its preference for a G residue at the last position of exon (Du and Rosbash 2002). Interaction of U1C protein precedes and possibly assists in U1 snRNA base-pairing at the 5′ ss. Further, 54G may break an RNA secondary structure that is inhibitory for the recruitment of U1 snRNP at the 5′ ss. All of these factors may have contributed to a context-specific dominant role of 54G in exon 7 splicing. The effect of 54G was so pronounced that abrogation of the Tra2-ESE in *SMN2* was tolerated (Fig. 9B, lane

13). Further, the inhibitory effects of G1H substitutions in combination with an abrogated Tra2-ESE were also compensated by 54G. The available computational programs cannot predict such a dominant effect of one exonic nucleotide on pre-mRNA splicing. A functional analysis of exons using a two-exon cassette also failed to detect the weak 5′ ss of exon 7 (Lim and Hertel 2001). We were able to assess the effect of 54G because a three-exon cassette, in which exons 6 and 7 competed for the same 3′ ss of intron 7, was used.

The presence of a G residue at the last position of an internal exon is not an absolute guarantee for its inclusion. For example, the abrogation of Tra2-ESE in the presence of inhibitory G1U substitution in *SMN2* led to exon 7 exclusion (Fig. 9B, lane 16). In several other instances, such as exon 11 of NF1 (Fang et al. 2001), exon 9 of human *ATP synthase* γ-subunit (Hayakawa et al. 2002) and exon 12 of *CFTR* (Pagani et al. 2003), the exons are alternatively spliced despite G at the last position. In addition, the tissue-specific skipping of exon 6 of *Caspase-3* gene has been reported despite the presence of guanosine residues at both ends of the exon (Huang et al. 2001). Selection of two clones with 54U and one with 54C further suggests that the presence of 54G is not an essential requirement for improving the 5′ ss of exon 7 (Fig. 5). This also demonstrates that the selected 54H variants were equally good if not better than the 54G containing isolates. These results clearly show that selection of 54G accompanied with the retention of the wild-type stimulatory motifs, as well as elimination of the wild-type inhibitory motifs. This was necessary to compete among 10^{11} variants that were initially used for in vivo selection. In other words, the strong stimulatory effect of 54G was exerted in association with other factors that also contributed to the selection. Our results suggest that the absence of 54G strengthens the inhibitory effects exerted by the residues of the 3′-Cluster that is strategically located in the vicinity of the 5′ ss. Supporting this argument, a number of mutations within the 3′-Cluster promoted exon 7 inclusion in *SMN2* (Fig. 7B).

Selection preserved the wild-type character of the molecule

Despite the predominant selection of G residues at the terminal positions, the wild-type nucleotides at many positions were found to be conserved. This suggests that only minimal changes were required to improve exon 7 inclusion by either abrogating the existing ESS elements and/or by strengthening the weak ESE elements. When we analyzed the variants with less than 33% substitutions, it was remarkable to observe a long tract of conserved positions in the middle of exon 7. This validated the rationale of our approach using 33% randomization, which was meant to determine the position-specific role of all exonic nucleotides. Supporting an essential role of Tra2, its associated ESE was found to be conserved. Consistent with the stimulatory role of the long conserved tract, mutations outside the boundary of Tra2-ESE also led to exon 7 exclusion in *SMN1* (Fig. 8). These results suggest that multiple proteins may contact the long conserved tract in the middle of exon 7.

Role of secondary contacts and RNA structure

We found position 8 (8G) relatively conserved within an otherwise inhibitory context. It is possible that 8G plays an important role by providing secondary contact to protein(s) or protein complexes that primarily bind to other sites. Secondary contacts may help stabilize a bridging protein–protein interaction critical for exon definition. We hypothesize that Tra2 or its interacting partners SRp30c and hnRNP G make such contacts to 8G to recruit U2AF^{35} at the 3′ ss. The role of an RNA secondary structure in reinforcing such contacts could not be ruled out. 8G falls within a loop of a predicted terminal stem-loop structure TSL1 at the 3′ ss of exon 7 (Singh et al. 2004). The specific orientation of 8G with respect to both the 3′ ss and Tra2-ESE may be critical for secondary contacts. Our earlier report in which addition of one base at the intron 6/exon 7 boundary completely obliterated exon 7 inclusion in *SMN1* supports this hypothesis (Singh et al. 2004). The important role of secondary contacts has been well established for a group II intron binding protein (Matsuura et al. 2001). Recently, it has been implicated for SR-proteins that show specificity to RNA conformations (Buratti et al. 2004). The effect of additional contacts has also been observed for U1 snRNP protein U1C, which produced different binding motifs in iterative selection experiment when flanking sequences varied (Du and Rosbash 2002). The above observations support a common theme in which distance and distribution of sequences that flank critical motifs are equally important.

Evolutionary considerations

One of the advantages of our strategy is the identification of *cis*-elements that are gained during evolution of the human genome. For instance, a comparison of mammalian exon 7 sequences revealed that human exon 7 contains an additional codon UUA between positions 43 and 45 (Fig. 7A). Supporting the inhibitory nature of this codon, A45 was found to be mutated by this selection protocol. Deletion of this particular codon restored exon 7 inclusion in *SMN2* to the level of *SMN1* (Fig. 7B). This is a very interesting observation, suggesting that the effects of the C6U transition is not inhibitory in the absence of a codon that was gained/retained during evolution. From the mechanistic point of view, this observation clearly demonstrates the interconnectivity of events that ensue at both ends of exon 7. Because of the critical difference in the sequence, it is most likely that exon 7 splicing in humans is regulated differently than in other mammalian counterparts. We believe that specific requirement in humans may have arisen due to duplication of *SMN* genes. Such a requirement may include putting a

cap on the levels of SMN synthesis. Thus, the roles played by 54A and the 3′-Cluster may be an essential component of the fine regulation of *SMN* gene expression. Also, this could be one of the most plausible reasons why evolution of human *SMN* did not opt for 54G, which would have assured exon 7 inclusion without any consequence to the protein sequence.

Future implications

Having successfully established the in vivo selection of an entire exon, our approach has enormous potential for analysis of other alternatively spliced exons. Our results demonstrated the "power of partial randomization," which revealed the context-specific positive and negative motifs, even in the presence of skewed selection at a few positions. With the advancement in oligo synthesis and DNA sequencing technology, our approach offers an attractive alternative to the random mutations that provide only limited information. In the absence of a general method to predict ESS and the inherent difficulty in performing the iterative in vivo selection of terminal exonic sequences, our approach offers a significant advantage. However, like any other method, all elements may not be discovered in one experiment. Additional experiments will be required to confirm specific features that could not be predicted by a mutability plot. These features include but are not limited to covariations, secondary contacts, and RNA structure. Based on the results of in vivo selection of the entire exon 7, future experiments that will further define a particular sequence stretch would be easier to plan.

Beyond helping understand the exon-specific mechanism of alternative splicing, our approach will have use in predicting disease severity. As evident from the available expressed sequence tags (ESTs), there are variations in exon 7 sequences among the human population. Our results suggest that people with *SMN1* deletions will have little or no risk of SMA if their *SMN2* exon 7 has appropriate mutations that abrogate inhibitory elements such as Exinct and/or the 3′-Cluster. Similar studies will be helpful in predicting the risk of other diseases that are directly associated with alternative splicing.

In summary, our approach established the feasibility of iterative in vivo selection of an entire exon. Using this approach we were able to show the relative significance of many positions within a critical exon that is associated with a debilitating disease of infants and children. Our results demonstrated the dominant effect of one exonic position that creates an extremely weak 5′ ss. It also confirmed the presence of Exinct and identified a long conserved tract in the middle of exon. It revealed a novel inhibitory *cis*-element at the 3′-end of exon 7 (3′-Cluster). Understanding alternative splicing of *SMN* genes is critical for considering novel treatment strategies. Most SMA patients carry copies of *SMN2*. Thus, improving exon 7 inclusion in *SMN2* holds the promise for a cure. Our results provide additional avenues in which the potential drugs could be designed to target the 5′ ss, leading to the increased inclusion of exon 7 in *SMN2*.

MATERIALS AND METHODS

Plasmid construction

Unless otherwise stated, all minigenes used in this study were constructed in the pCI mammalian expression vector (Promega), using a strategy described earlier (Singh et al. 2004). Minigene splicing cassettes pSMN1ΔI6 and pSMN2ΔI6 were constructed by deleting ~6 kb within intron 6 from pSMN1 and pSMN2, respectively (Singh et al. 2004), allowing detection of partially spliced intermediates along with fully spliced products (Fig. 4A). For all our experiments performed in this report, pSMN1ΔI6 and pSMN2ΔI6 refer to *SMN1* and *SMN2*, respectively. The minigene splicing cassette pBxT2 was described earlier (Singh et al. 2004). All mutations were confirmed by sequencing.

Cell culture, transfection, and in vivo splicing analysis

Human cervical carcinoma (C33a) cells were grown in Dulbecco's modified Eagle medium (DMEM) supplemented with 10% fetal bovine serum (GIBCO-BRL), 100 units/mL penicillin, and 100 μg/mL streptomycin (Mediatech). Transfections, reverse transcription, and amplification were performed similarly as described by Singh et al. (2004). Amplification products representing the in vivo spliced intermediates were separated on a native 5.8% polyacrylamide gel (w/v). Analysis and quantifications were performed using a FPL-5000 Image Reader and ImageGauge software (Fuji Photo Film Inc.). Splicing efficiency was calculated after normalization to the internal HPRT control by computing the fraction of spliced product counts over the sum of spliced and unspliced counts. The percentages of exon 7-included and exon 7-excluded products were calculated from an average of three independent experiments.

Generation of a partially random exon 7

The initial pool with partially random exon 7 was generated using a 90-mer oligonucleotide (E7Rand; 5′-CCTTTATTTTCCTTACAG ggtttcagacaaaatcaaaaagaaggaaggtgctcacattccttaaattaaggaGTAAGTCT GCCAGCATTA-3′). Lowercase letters represent partially random 54-nt-long exon 7; capital letters represent two stretches of flanking intronic sequences (18 nt each). The randomization (doping) was performed with 70% wild-type and 10% each of the three non-wild-type nucleotides (Singh et al. 2002). Using high fidelity PCR with P*fx*, E7Rand as an upstream primer, PCI-DN as a downstream primer, and pSMN1ΔI6 as a template, a 1.1-kb DNA fragment was generated. This fragment contained a partially random exon 7, the entire intron 7, and most of exon 8. Approximately 4 μg (~4×10^{12} unique molecules) of the gel-purified fragment were used as a mega-primer in the second PCR with PCI-UP as a forward primer and pSMN1ΔI6 as a template. The resultant 1.5-kb-long amplification product (containing the entire exon 6, short intron 6, a partially random exon 7, the entire intron 7, and most

of exon 8) was gel purified and digested with NotI and XhoI. The digestion product was ligated into NotI-XhoI-digested pSMN1ΔI6. The ligation mixture served as the initial pool (pool 0), which was used for in vivo selection experiments. Sequences of the clones from the initial pool confirmed partial randomization (not shown).

In vivo selection

The in vivo selection experiment started with transfection of C33a cells with the initial pool of splicing cassettes (~2×10^{11} unique molecules per 4×10^6 cells) and continued with the selection and amplification steps outlined in Figure 2A: (1) amplification of the in vivo spliced products by RT-PCR; (2) gel isolation of the product with included exon 7; (3) amplification of exon 7 in a second PCR reaction; (4) ligation of amplified exon 7 back into a splicing cassette; and (5) transfection of C33a cells with the ethanol-precipitated ligated products. Steps from 1 to 5 comprised one round of selection. Repeated rounds enriched exon 7 sequences that promoted exon 7 inclusion. At each round, the transfection was done directly with the ligation mixture to maintain pool diversity (Singh et al. 2004). Ligation was performed in a 100-μL reaction containing ~1 μg of BsaXI-digested pBxT2 and ~50 ng of enriched exon 7 fragment. Approximately 1μg of ligated plasmid (~2×10^{11} molecules) was used to transfect two 60-mm plates of C33a cells (~2×10^6 cells per plate) using a standard calcium phosphate coprecipitation procedure. About 20 h posttransfection, cells were harvested, total RNA was isolated, and the in vivo spliced intermediates amplified by RT-PCR essentially as described earlier except that ~3 μg of total RNA were used per 20 μL RTase reaction and the number of PCR cycles did not exceed 20. The 333-bp DNA fragment, which corresponded to the exon 7-included product, was gel purified and used as a template for the second PCR with primers B1 and B2 (Fig. 2B). The resulting 134-bp DNA fragment was gel purified, digested with BsaXI and inserted back into the BsaXI-linearized pBxT2. At each round, a fraction of ligation mixture was transformed into *Escherichia coli* to isolate the individual clones. The splicing activity of ~30 unique clones was analyzed at each round. Four rounds of selection were performed and the splicing pattern of 59 unique clones from pool 4 was determined.

Analysis of sequences

We analyzed the statistically significant number of clones from the final round of in vivo selection. To determine the position-specific significance of nucleotides, sequences were forced aligned from one end to the other (Bartel et al. 1991; Singh et al. 2002). The "mutability values" shown in Figure 6 were calculated by comparing the ratios (R) of mutant (mut) to wild-type (wt) nucleotides of exon 7 in selected (pool-4) and initial pools (pool-0), using the equation $[(R_{(\mathrm{mut/wt})\mathrm{pool\text{-}4}})/(R_{(\mathrm{mut/wt})\mathrm{pool\text{-}0}})] - 1$ (Deminoff et al. 1995). Motif analyses were done using MacVector software. Covariation analysis was done manually.

ACKNOWLEDGMENTS

We express our appreciation to Drs. Phillip Zamore, Stephan Luecke, and Thoru Pederson for valuable suggestions and critical comments on the manuscript. This work was supported by grants from National Institutes of Health to E.J.A. (RO1 NS40275) and from Families of SMA to R.N.S.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

Received April 6, 2004; accepted May 24, 2004.

REFERENCES

Andreassi, C., Jarecki, J., Zhou, J., Coovert, D.D., Monani, U.R., Chen, X., Whitney, M., Pollok, B., Zhang, M., Androphy, E., et al. 2001. Aclarubicin treatment restores SMN levels to cells derived from type I spinal muscular atrophy patients. *Hum. Mol. Genet.* **10:** 2841–2849.

Bartel, D.P., Zapp, M.L., Green, M.R., and Szostak, J.W. 1991. HIV-1 Rev regulation involves recognition of non-Watson–Crick base pairs in viral RNA. *Cell* **67:** 529–536.

Baskerville, S., Zapp, M., and Ellington, A.D. 1995. High-resolution mapping of the human T-cell leukemia virus type 1 Rex-binding element by in vitro selection. *J. Virol.* **69:** 7559–7569.

Black, D.L. 2003. Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.* **72:** 291–336.

Buratti, E., Muro, A.F., Giombi, M., Gherbassi, D., Iaconcig, A., and Baralle, F.E. 2004. RNA folding affects the recruitment of SR proteins by mouse and human polypurinic enhancer elements in the fibronectin EDA exon. *Mol. Cell. Biol.* **24:** 1387–1400.

Burge, C.B., Tuschl, T., and Sharp, P.A. 1999. Splicing of precursors to mRNAs by the spliceosomes. In *The RNA world*, 2nd ed. (eds. R.F. Gesteland et al.), pp. 525–560. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Cartegni, L. and Krainer, A.R. 2002. Disruption of an SF2/ASF-dependent exonic splicing enhancer in *SMN2* causes spinal muscular atrophy in the absence of *SMN1*. *Nat. Genet.* **30:** 377–384.

———. 2003. Correction of disease-associated exon skipping by synthetic exon-specific activators. *Nat. Struct. Biol.* **10:** 120–125.

Cartegni, L., Chew, S.L., and Krainer, A.R. 2002. Listening to silence and understanding nonsense: Exonic mutations that affect splicing. *Nat. Rev. Genet.* **3:** 285–298.

Cartegni, L., Wang, J., Zhu, Z., Zhang, M.Q., and Krainer, A.R. 2003. ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res.* **31:** 3568–3571.

Chang, J.G., Hsieh-Li, H.M., Jong, Y.J., Wang, N.M., Tsai, C.H., and Li, H. 2001. Treatment of spinal muscular atrophy by sodium butyrate. *Proc. Natl. Acad. Sci.* **98:** 9808–9813.

Chua, K. and Reed, R. 2001. An upstream AG determines whether a downstream AG is selected during catalytic step II of splicing. *Mol. Cell. Biol.* **21:** 1509–1514.

Collins, C.A. and Guthrie, C. 2000. The question remains: Is the spliceosome a ribozyme? *Nat. Struct. Biol.* **7:** 850–854.

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.

Damgaard, C.K., Tange, T.O., and Kjems, J. 2002. hnRNP A1 controls HIV-1 mRNA splicing through cooperative binding to intron and exon splicing silencers in the context of a conserved secondary structure. *RNA* **8:** 1401–1415.

Deminoff, S.J., Tornow, J., and Santangelo, G.M. 1995. Unigenic evolution: A novel genetic method localizes a putative leucine zipper that mediates dimerization of the *Sacharomyces cerevisiae* regulator Gcr1p. *Genetics* **141:** 1263–1274.

Dreyfuss, G., Kim, V.N., and Kataoka, N. 2002. Messenger-RNA-binding proteins and the messages they carry. *Nat. Rev. Mol. Cell. Biol.* **3:** 195–205.

Du, H. and Rosbash, M. 2002. The U1 snRNP protein U1C recognizes

the 5′ splice site in the absence of base pairing. *Nature* **419:** 86–90.
Fairbrother, W.G., Yeh, R.F., Sharp, P.A., and Burge, C.B. 2002. Predictive identification of exonic splicing enhancers in human genes. *Science* **297:** 1007–1013.
Fang, L.J., Simard, M.J., Vidaud, D., Assouline, B., Lemieux, B., Vidaud, M., Chabot, B., and Thirion, J.P. 2001. A novel mutation in the neurofibromatosis type 1 (NF1) gene promotes skipping of two exons by preventing exon definition. *J. Mol. Biol.* **307:** 1261–1270.
Graveley, B.R. 2000. Sorting out the complexity of SR protein functions. *RNA* **6:** 1197–1211.
Grover, A., Houlden, H., Baker, M., Adamson, J., Lewis, J., Prihar, G., Pickering-Brown, S., Duff, K., and Hutton, M. 1999. 5′ splice site mutations in τ associated with the inherited dementia FTDP-17 affect a stem-loop structure that regulates alternative splicing of exon 10. *J. Biol. Chem.* **274:** 15134–15143.
Guo, H., Karberg, M., Long, M., Jones 3rd, J.P., Sullenger, B., and Lambowitz, A.M. 2000. Group II introns designed to insert into therapeutically relevant DNA target sites in human cells. *Science* **289:** 452–457.
Hastings, M.L. and Krainer, A.R. 2001. Pre-mRNA splicing in the new millennium. *Curr. Opin. Cell. Biol.* **13:** 302–309.
Hayakawa, M., Sakashita, E., Ueno, E., Tominaga, S., Hamamoto, T., Kagawa, Y., and Endo, H. 2002. Muscle-specific exonic splicing silencer for exon exclusion in human *ATP synthase* γ-subunit pre-mRNA. *J. Biol. Chem.* **277:** 6974–6984.
Helmken, C., Hofmann, Y., Schoenen, F., Oprea, G., Raschke, H., Rudnik-Schoneborn, S., Zerres, K., and Wirth, B. 2003. Evidence for a modifying pathway in SMA discordant families: Reduced SMN level decreases the amount of its interacting partners and Htra2-β1. *Hum. Genet.* **114:** 11–21.
Hofmann, Y. and Wirth, B. 2002. hnRNP-G promotes exon 7 inclusion of survival motor neuron (SMN) via direct interaction with Htra2-β1. *Hum. Mol. Genet.* **11:** 2037–2049.
Hofmann, Y., Lorson, C.L., Stamm, S., Androphy, E.J., and Wirth, B. 2000. Htra2-β1 stimulates an exonic splicing enhancer and can restore full-length SMN expression to survival motor neuron 2 (*SMN2*). *Proc. Natl. Acad. Sci.* **97:** 9618–9623.
Huang, Y., Shin, N.H., Sun, Y., and Wang, K.K. 2001. Molecular cloning and characterization of a novel *caspase-3* variant that attenuates apoptosis induced by proteasome inhibition. *Biochem. Biophys. Res. Commun.* **283:** 762–769.
Kashima, T. and Manley, J.L. 2003. A negative element in *SMN2* exon 7 inhibits splicing in spinal muscular atrophy. *Nat. Genet.* **34:** 460–463.
Lallena, M.J., Chalmers, K.J., Llamazares, S., Lamond, A.I., and Valcarcel, J. 2002. Splicing regulation at the second catalytic step by Sex-lethal involves 3′ splice site recognition by SPF45. *Cell* **109:** 285–296.
Lefebvre, S., Burglen, 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:** 1–5.
Leontis, N.B., Stombaugh, J., and Westhof, E. 2002. The non-Watson–Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res.* **30:** 3497–3531.
Lim, S.R. and Hertel, K.J. 2001. Modulation of survival motor neuron pre-mRNA splicing by inhibition of alternative 3′ splice site pairing. *J. Biol. Chem.* **276:** 45476–45483.
Lorson, C.L. and Androphy, E.J. 2000. An exonic enhancer is required for inclusion of an essential exon in the SMA-determining gene *SMN. Hum. Mol. Genet.* **9:** 259–265.
Lorson, C.L., Strasswimmer, J., Yao, J.M., Baleja, J.D., Hahnen, E., Wirth, B., Le, T., Burghes, A.H., and Androphy, E.J. 1998. SMN oligomerization defect correlates with spinal muscular atrophy severity. *Nat. Genet.* **19:** 63–66.
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.* **96:** 6307–6311.
Maniatis, T. and Reed, R. 2002. An extensive network of coupling among gene expression machines. *Nature* **416:** 499–506.
Maniatis, T. and Tasic, B. 2002. Alternative pre-mRNA splicing and proteome expansion in metazoans. *Nature* **418:** 236–243.
Matsuura, M., Noah, J.W., and Lambowitz, A.M. 2001. Mechanism of maturase-promoted group II intron splicing. *EMBO J.* **20:** 7259–7270.
Meister, G., Eggert, C., and Fischer, U. 2002. SMN-mediated assembly of RNPs: A complex story. *Trends Cell Biol.* **12:** 472–478.
Miyajima, H., Miyaso, H., Okumura, M., Kurisu, J., and Imaizumi, K. 2002. Identification of a *cis*-acting element for the regulation of SMN exon 7 splicing. *J. Biol. Chem.* **277:** 23271–23277.
Miyaso, H., Okumura, M., Kondo, S., Higashide, S., Miyajima, H., and Imaizumi, K. 2003. An intronic splicing enhancer element in Survival Motor Neuron (*SMN*) pre-mRNA. *J. Biol. Chem.* **278:** 15825–15831.
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.
Moore, M.J. 2000. Intron recognition comes of AGe. *Nat. Struct. Biol.* **7:** 14–16.
Nilsen, T.W. 2003. The spliceosome: The most complex macromolecular machine in the cell? *Bioessays* **25:** 1147–1149.
Pagani, F., Stuani, C., Tzetis, M., Kanavakis, E., Efthymiadou, A., Doudounakis, S., Casals, T., and Baralle, F.E. 2003. New type of disease causing mutations: The example of the composite exonic regulatory elements of splicing in *CFTR* exon 12. *Hum. Mol. Genet.* **12:** 1111–1120.
Pellizzoni, L., Yong, J., and Dreyfuss, G. 2002. Essential role for the SMN complex in the specificity of snRNP assembly. *Science* **298:** 1775–1779.
Proudfoot, N.J., Furger, A., and Dye, M.J. 2002. Integrating mRNA processing with transcription. *Cell* **108:** 501–512.
Reed, R. and Hurt, E. 2002. A conserved mRNA export machinery coupled to pre-mRNA splicing. *Cell* **108:** 523–531.
Rothrock, C., Cannon, B., Hahm, B., and Lynch, K.W. 2003. A conserved signal-responsive sequence mediates activation-induced alternative splicing of CD45. *Mol. Cell* **12:** 1317–1324.
Saldanha, R., Ellington, A., and Lambowitz, A.M. 1996. Analysis of the CYT-18 protein binding site at the junction of stacked helices in a group I intron RNA by quantitative binding assays and in vitro selection. *J. Mol. Biol.* **261:** 23–42.
Singh, R.N., Saldanha, R.J., D'Souza, L.M., and Lambowitz, A.M. 2002. Binding of a group II intron-encoded reverse transcriptase/maturase to its high affinity intron RNA binding site involves sequence-specific recognition and autoregulates translation. *J. Mol. Biol.* **318:** 287–303.
Singh, N.N., Androphy, E.J., and Singh, R.N. 2004. An extended inhibitory context causes skipping of exon 7 of *SMN2* in spinal muscular atrophy. *Biochem. Biophys. Res. Commun.* **315:** 381–388.
Skordis, L.A., Dunckley, M.G., Yue, B., Eperon, I.C., and Muntoni, F. 2003. Bifunctional antisense oligonucleotides provide a *trans*-acting splicing enhancer that stimulates *SMN2* gene expression in patient fibroblasts. *Proc. Natl. Acad. Sci.* **100:** 4114–4119.
Staley, J.P. and Guthrie, C. 1998. Mechanical devices of the spliceosome: Motors, clocks, springs and things. *Cell* **92:** 315–326.
Stoilov, P., Daoud, R., Nayler, O., and Stamm, S. 2004. Human tra2-β1 autoregulates its protein concentration by influencing alternative splicing of its pre-mRNA. *Hum. Mol. Genet.* **13:** 509–524.
Sumner, C.J., Huynh, T.N., Markowitz, J.A., Perhac, J.S., Hill, B., Coovert, D.D., Schussler, K., Chen, X., Jarecki, J., Burghes, A.H., et al. 2003. Valproic acid increases SMN levels in spinal muscular atrophy patient cells. *Ann. Neurol.* **54:** 647–654.
Tu, M., Tong, W., Perkins, R., and Valentine, C.R. 2000. Predicted changes in pre-mRNA secondary structure vary in their association with exon skipping for mutations in exons 2, 4, and 8 of the *Hprt* gene and exon 51 of the fibrillin gene. *Mutat. Res.* **432:** 15–32.

Vant-Hull, B., Payano-Baez, A., Davis, R.H., and Gold, L. 1998. The mathematics of SELEX against complex targets. *J. Mol. Biol.* **278:** 579–597.

Wang, J., Chang, Y.F., Hamilton, J.I., and Wilkinson, M.F. 2002. Nonsense-associated altered splicing: A frame-dependent response distinct from nonsense-mediated decay. *Mol. Cell* **10:** 951–957.

Wu, S., Romfo, C.M., Nilsen, T.W., and Green, M.R. 1999. Functional recognition of the 3′-splice site AG by the splicing factor U2AF^{35}. *Nature* **402:** 832–835.

Young, P.J., DiDonato, C.J., Hu, D., Kothary, R., Androphy, E.J., and Lorson, C.L. 2002. SRp30c-dependent stimulation of survival motor neuron (*SMN*) exon 7 inclusion is facilitated by a direct interaction with hTra2-β1. *Hum. Mol. Genet.* **11:** 577–587.

Zhang, M.L., Lorson, C.L., Androphy, E.J., and Zhou, J. 2001. An in vivo reporter system for measuring increased inclusion of exon 7 in SMN2 mRNA: Potential therapy of SMA. *Gene Ther.* **8:** 1532–1538.

In vivo selection reveals combinatorial controls that define a critical exon in the spinal muscular atrophy genes

NATALIA N. SINGH, ELLIOT J. ANDROPHY and RAVINDRA N. SINGH

RNA 2004 10: 1291-1305