

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
29 December 2004 (29.12.2004)

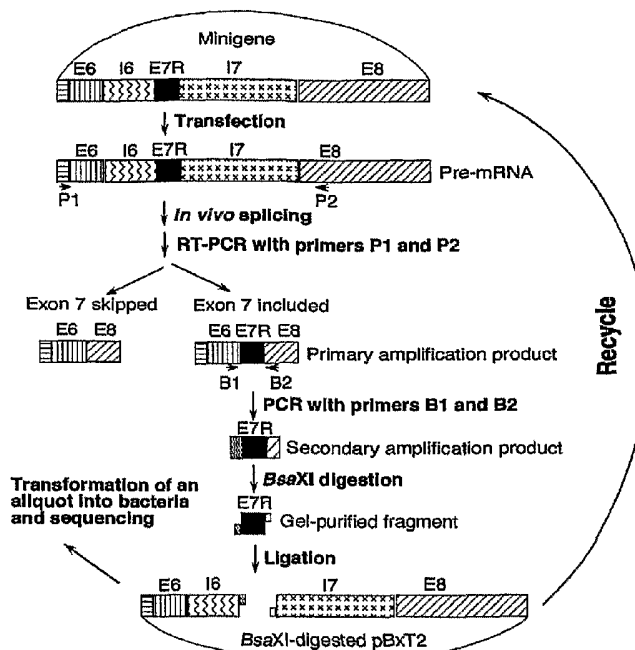
PCT

(10) International Publication Number
WO 2004/113867 A2

- (51) International Patent Classification⁷: **G01N**
- (21) International Application Number:
PCT/US2004/019081
- (22) International Filing Date: 16 June 2004 (16.06.2004)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
60/478,469 16 June 2003 (16.06.2003) US
60/576,380 1 June 2004 (01.06.2004) US
- (71) Applicant (for all designated States except US): **UNIVERSITY OF MASSACHUSETTS** [US/US]; One Beacon Street, 26th Floor, Boston, MA 02108 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): **SINGH, Ravindra, N.** [IN/US]; 15 Sherwood Avenue, Shrewsbury, MA 01545 (US). **SINGH, Natalia, N.** [RU/US]; 15 Sherwood Avenue, Shrewsbury, MA 01545 (US). **ANDROPHY, Elliot, J.** [US/US]; 2 Mark Street, Natick, MA 01760 (US).
- (74) Agent: **KUGLER, Janice, L.**; Fish & Richardson P.C., 225 Franklin Street, Boston, MA 02110-2804 (US).
- (81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.
- (84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) Title: EXON ANALYSIS



(57) Abstract: The present invention provides methods of analyzing an entire exon or a portion thereof using a novel cell-based SELEX procedure. Described is a novel and accurate means of determining the location of exonic splicing enhancers (ESEs) and exonic splicing silencers (ESSs). The methods described have wide application, e.g., for determining the molecular etiology of human diseases associated with splicing errors, and provides new therapeutic targets for such diseases. Nucleotide sequences (aptamers) identified by the present invention are also provided.



Published:

— without international search report and to be republished
upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

EXON ANALYSIS

CLAIM OF PRIORITY

This application claims priority under 35 USC §119(e) to U.S. Provisional Patent Application Serial No. 60/478,469, filed on June 16, 2003, and a Provisional Patent
5 Application titled Exon Analysis, filed June 1, 2004 and assigned Serial No. 60/NNN,NNN, the entire contents of which are hereby incorporated by reference.

TECHNICAL FIELD

The present invention relates to methods to identify positive and negative cis-splicing elements found within an exon.

BACKGROUND

10 Removal of intron sequences by splicing occurs by two sequential transesterification steps that are catalyzed by the components of a large (about 60S) RNA-protein complex, termed the spliceosome (Kramer, *Annu. Rev. Biochem.* 65:367-409 (1996)). The major spliceosome is formed by step-wise assembly of four small
15 nuclear ribonucleoprotein particles (snRNPs U1, U2, U4/U6, and U5) and non-snRNP splicing factors on a pre-mRNA (Burge et al., pp. 525-560 in *The RNA World*, Cold Spring Laboratory Press, Cold Spring Harbor, NY (1999)). The non-snRNP factors important in constitutive and alternative splicing comprise the serine-arginine (SR) proteins (Fu, *RNA* 1:663-680 (1995); Manley and Tacke, *Genes Dev.* 10:1569-1579
20 (1996); Gravely, *Trends Genet.* 17:100-107 (2000)), SR-like proteins (Blencowe et al., *Biochem. Cell. Biol.* 77:277-291 (1999)), hnRNP (Dreyfuss et al., *Rev. Mol. Cell. Biol.* 3:195-205 (2002); Smith and Valcarcel, *Trends Biochem. Sci.* 25:381-388 (2000)), and other exon-specific proteins (Burge et al., pp. 525-560 in *The RNA World*, Cold Spring Laboratory Press, Cold Spring Harbor, NY (1999); Smith and Valcarcel, *Trends Biochem.*
25 *Sci.* 25:381-388 (2000)). Most of these proteins contain RNA recognition motifs (RRMs) that recognize specific sequences on exons called exonic splicing enhancers (ESE) or exonic splicing silencers (ESS) (Smith and Valcarcel, *Trends Biochem. Sci.* 25:381-388 (2000)).

30 Generally, exon inclusion is modulated by a combinatorial interplay of positive (ESE) and negative (ESS) *cis*-elements. Because of the promiscuous nature of SR and SR-like proteins, accurate prediction of ESE and ESS elements has been very difficult.

Use of weight matrices based on functional SELEX (Systematic Evaluation of Ligands through EXponential Enrichment) methods (Liu et al., *Nat. Genet.* 27:55-58) (2001)) and a recently developed computational method, RESCUE-USE (Fairbrother et al., *Science* 297:1007-1013 (2002)), have been useful in identifying such elements. However, these
5 methods have their own limitations in predicting all *cis*-elements, due to complexity of the context, which is diverse and dissimilar at every exon.

The SELEX technique and various modifications have been described in numerous U.S. patents including U.S. Patent Nos. 5,270,163; 6,344,318; 6,441,158;
10 6,482,594; 6,465,189; 6,387,620; 6,376,190; 6,291,184; 6,261,774; 6,048,698; 6,030,776; 5,998,142; and 5,792,613. This technique remains the most unbiased experimental method to predict the relative impacts of individual and multiple substitutions under the physiological concentration of splicing factors.

SUMMARY

15 In part, the present invention is based on new methods for identifying nucleic acid sequences that are crucial to the gene splicing process. The methods can be used, for example, to determine the contribution to pre-mRNA splicing of every exonic nucleotide, and to identify context-specific *cis*-interactions that affect splicing, including long-distance *cis*-interactions. The cell-based SELEX methods described herein can be used to
20 scan an entire exon or nucleotide sequences flanking exon boundaries for splicing enhancers and silencers. The present invention provides methods of predicting and identifying positive and negative splicing elements within an entire exon. In some embodiments, the present method can be used to simultaneously identify positive and negative splicing elements over a large stretch of sequences including an entire exon (e.g.,
25 about 50-300 bases) for genes involved in alternative splicing or associated with splicing defect diseases. The methods described herein can be used to identify compounds that affect splicing in genes associated with splicing defect diseases, which can be used to treat the diseases. In addition, the methods can be used to identify nucleic acid sequences, e.g., aptamers as described herein, that modulate (e.g., enhance or inhibit)
30 splicing and can be used as therapeutic compounds to treat splicing defect diseases.

The methods described herein include providing a pool of nucleic acids, wherein each nucleic acid includes partially randomized exonic sequences, e.g., comprising the sequence of an entire exon of interest; subjecting the pool of nucleic acids to one or more rounds of cell-based selection; selecting those nucleic acids with pre-selected splicing

characteristics, i.e., inclusion or exclusion of an exon; and determining the distribution of mutations that are associated with the pre-selected splicing characteristics. In some embodiments, the pre-selected splicing characteristics include the presence or exclusion of a selected exon. The distribution of mutations in hyper- and hypo-mutable regions corresponds to negative (not crucial to splicing) and positive (crucial to splicing) wild type nucleotides, respectively. In some embodiments, the nucleic acids contain additional wild-type or partially randomized intronic sequences and exonic sequences in addition to partially randomized exonic sequences comprising the sequence of an entire exon of interest.

In some embodiments, the new methods include techniques based in part on cell-based SELEX (Systematic Evolution of Ligands by Exponential Enrichment) methodology. The methods can include some or all of the following:

- (a) expressing in a cell, e.g., a eukaryotic cell, a pool of partially random oligonucleotides including an exon or a portion thereof,
- (b) isolating spliced mRNA from the transfected eukaryotic cell,
- (c) amplifying an exon-included product from the isolated mRNA,
- (d) isolating the exon-included product, which is typically an oligonucleotide comprising an included exon,
- (e) inserting double cutter or quadruple cutter restriction enzyme sites outside the boundary of the included exon to create a double cutter- or quadruple cutter-engineered oligonucleotide,
- (f) digesting the double cutter- or quadruple cutter-engineered oligonucleotide with a double cutter or quadruple cutter restriction enzyme and purifying the exon,
- (g) ligating the purified exon into a double cutter- or quadruple cutter-linearized plasmid; and
- (h) transfecting cells, e.g., eukaryotic cells, with the ligated plasmid of step (g) containing the exon .

In some embodiments, the method further includes:

- (j) transfecting a bacterial cell with the ligated plasmid of step (g) including the included exon and analyzing the sequence of the exon to determine where wild-type nucleotides are retained or where non-wild-type nucleotides are substituted, and

(k) designating a position where wild-type nucleotides are retained as a portion of a potential positive splicing element and positions where non-wild-type nucleotides are substituted as a portion of a potential negative splicing element.

While the above steps are labeled for ease of reference (*e.g.*, step (a)), the methods of the present invention are not intended to be limited to a particular order of steps. Variations on the order of the steps will be apparent to one skilled in the art to achieve the goals of the present invention. In some embodiments, step (a) is achieved by transfecting the cells with plasmids suitable for expressing partially random nucleotides. In some embodiments, the amplification steps employ polymerase chain reaction (PCR), particularly RT-PCR. Steps (b)-(g) can be performed more than once, *e.g.*, two, three, four or more times. In some embodiments, steps (j) and (k) are performed before step (h).

In some embodiments, the quadruple cutter restriction enzyme is *BsaXI*. In some embodiments, the plasmid including partially random oligonucleotides in step (a) is human expression vector pCI plasmid.

In some embodiments, the exonic sequences each comprise at least about 50, 75, 100, 150, 200, or 250 nucleotides. In some embodiments, the exonic sequences comprise an entire exon.

Any bacterial cell that can conveniently be grown in culture and transfected, such as *E. coli*, can be used in the methods described herein. Among eukaryotic cells, human cells can be used in the present invention, *e.g.*, C33a human cervical cancer cells, motor neuron-like NSC34 cells, and fibroblast cells from a subject with SMA. In one embodiment, the eukaryotic cells are neuronal cells.

As used herein, a “cell-based method” refers to a method that makes use of a living cell, which can be a tissue culture cell, a cultured primary cell, or a cell in an animal.

As used herein, a “partially random oligonucleotide” refers to a nucleic acid in which each position has a greater chance of being wild-type than of being non-wild type. Given that there are four possible nucleotides (adenosine (A), guanine (G), thymine (T) or cytosine (C)), a fully random oligonucleotide would be one in which each position has a roughly equal chance of being any one of the possible nucleotides (so that about 25% = A, about 25% = G, about 25% = T, and about 25% = C). A partially random oligonucleotide is one in which each position is “doped” with the wild type nucleotide, so that there is a greater chance that the nucleotide will be wildtype than not. Thus, for

example, if the wild type sequence includes a G at position 6, the partially random nucleotides will be at least about 30% = G, and less of each of the others (A, T, and C); these are referred to as “70% substituted.” In some embodiments, the partially random oligonucleotides are about 70%, 60%, 50%, 40%, 30%, 20%, or 10% substituted. Given the difficulties in controlling the process by which the partially random oligonucleotides are synthesized, “about” in this context includes variations of up to about $\pm 5\text{-}7\%$. In some embodiments, partially random oligonucleotides for use in the methods described herein can be doped with about 70% wild-type nucleotides, and about 10% each of non-wild-type nucleotides (e.g., 30% substituted).

In some embodiments, the number of partially random oligonucleotides is at least 10^{12} (of which about 10^{11} are unique). In one embodiment, the quadruple cutter enzyme used is *BsaXI*. In another embodiment, the library of partially random doped oligonucleotides is inserted into human expression vector pCI plasmid. The methods described herein can be used for any gene involved in alternative splicing to identify positive and negative splicing elements in an exon or boundary portion thereof. For example, the methods described herein can be used to analyze exons of any of the genes listed in Table 1.

The amplification steps can employ PCR, particularly RT-PCR. Exon-included products and/or oligonucleotides can be isolated by gel-purification.

In some embodiments, the method further comprises analyzing exon-included sequences by amplification and sequencing of individual clones.

Double cutter enzymes cut two times towards the 3' end of the enzyme recognition motif. Double cutter enzymes can be used singly or in combination with another double cutter enzyme. For instance, one double cutter enzyme can be used at the 5' end of the exon and a different double cutter enzyme can be used at the 3' end of the exon. A number of double-cutters are known in the art and described herein.

Quadruple cutters can also be used. These restriction enzymes cut two times towards the 3' end and two times towards the 5' end of the recognition motifs. Exemplary quadruple cutters are *BaeI* (5 nucleotide overhang), *BcgI* (2 nucleotide overhang), and *BsaXI* (3 nucleotide overhang). In some embodiments, *BsaXI* is used. A number of quadruple-cutters are known in the art and described herein.

Also described herein are nucleotide sequences (aptamers), identified by the cell-based methods described herein, that are derived from (e.g., a variant identified via a cell-

based selection method described herein) the splicing disease-associated genes, compositions including these aptamers, and compositions and methods for targeting (e.g., inhibiting the binding of proteins to) these aptamers. These aptamers, or fragments thereof, bind to proteins and can be used, *inter alia*, to promote or inhibit *cis*- or *trans*-splicing of exons as appropriate, and thus can be used either as compounds useful in treating splicing diseases such as spinal muscular atrophy (SMA), or to screen for such compounds. Such compounds can be, *inter alia*, aptamers, small molecules, ribozymes, antisense (including morpholino oligos), or small interfering RNAs (siRNAs) that bind to the aptamer sequence and inhibit or promote *cis*- or *trans*-splicing of exons. In some embodiments, the aptamers bind to regulatory elements that interact with *cis*-elements.

The present invention also provides methods of screening for candidate compounds that enhance splicing of a Survival of Motor Neuron (SMN) RNA. The methods include providing a ribonucleic acid including the RNA equivalent of SEQ ID NO:247 or a fragment thereof, e.g., nucleotides 1-22, 23-40, or 35-54, e.g., the Exinct region (nucleotides 3-15, including the binding site of SF2/ASF (Splicing Factor 2/Alternative Splicing Factor), a highly conserved nuclear protein essential for splicing of pre-mRNA that can influence selection of alternative splice sites); the conserved central region (nucleotides 16-44, including the Tra2 (Transformer 2) binding site), or the 3'-Cluster (nucleotides 45-51); contacting the ribonucleic acid with a test compound; and measuring binding of the test compound to the ribonucleic acid. Binding indicates that the test compound is a candidate compound that can potentially be used to enhance splicing of an SMN RNA.

In another aspect, the invention provides methods of screening for candidate therapeutic compounds for the treatment of SMA. The method includes expressing a deoxyribonucleic acid including a sequence encoding a SMN2 (SEQ ID NO:247) ribonucleic acid in a cell; contacting the cell with a candidate compound identified by a method described herein; and detecting the presence of properly spliced ribonucleic acid transcribed from the deoxyribonucleic acid in the cell; wherein the presence of the ribonucleic acid indicates that the compound is a candidate therapeutic compound for the treatment of SMA.

The invention also provides methods of identifying a candidate therapeutic compound for the treatment of spinal muscular atrophy (SMA), by administering a candidate compound identified by a method described herein to an animal model of SMA,

and determining the effect of the candidate compound on a symptom of SMA in the animal model. An improvement in a symptom of SMA indicates that the compound is a candidate therapeutic compound for the treatment of SMA.

The test compounds utilized in the assays and methods described herein can be, inter alia, nucleic acids, small molecules, organic or inorganic compounds, immunoglobulins or immunoglobulin fragments, proteins, or polypeptides. For example, SMN1 or SMN2 polynucleotides (e.g., variants including truncation mutants, deletion mutants, and point mutants as described herein; sense, antisense, and small inhibitory RNAs (siRNAs) including short hairpin RNAs (shRNAs); aptamers; and ribozymes) can be used as test compounds in the methods described herein. Alternatively, compounds or compositions that mimic the RNA structure of SMN1/SMN2 can be used.

A test compound that has been screened by a cell-based method described herein and determined to have a desired activity, e.g., enhancement of exon 7 inclusion, can be considered a candidate compound. A candidate compound that has been screened, e.g., in an *in vivo* model, and determined to have a desirable effect, e.g., on one or more symptoms of a disorder associated with gene splicing as described herein (such as SMA), can be considered a candidate therapeutic agent. Candidate therapeutic agents, once screened in a clinical setting, are therapeutic agents, which can then be optionally optimized and/or derivatized, and formulated with physiologically acceptable excipients to form pharmaceutical compositions. In some embodiments, the test compound, candidate compound, or candidate therapeutic compound is an aptamer, e.g., an aptamer including the sequence UUUUUGAUUUUGAUCGUAUGAUC (SEQ ID NO:253).

In some embodiments, the compounds are optimized to improve their therapeutic index, i.e., increase therapeutic efficacy and/or decrease unwanted side effects. Thus, in some embodiments, the methods described herein include optimizing the test or candidate compound. In some embodiments, the methods include formulating a therapeutic composition including a test or candidate compound (e.g., an optimized compound) and a pharmaceutically acceptable carrier therefor. In some embodiments, the compounds are optimized by derivatization.

The invention further includes methods of diagnosing a subject with SMA, by obtaining a sample of genomic DNA from the patient, and determining the presence of one or more sequences that negatively affect splicing of exon 7 of SMN. The presence of a sequence that negatively affects splicing of exon 7 of SMN indicates that the subject

has SMA. In some embodiments, the sequences that negatively affect splicing include one or more of Exinct and/or the 3'-Cluster.

As detailed below, the cell-based methods described herein were used to analyze exon 7 of the *SMN* gene to identify splicing enhancing and silencing elements. Thus the invention also provides nucleotide sequences (aptamers) identified by the cell-based SELEX of the *SMN1/SMN2* genes and compositions including these aptamers. These aptamers, or fragments thereof, can be used, *inter alia*, to promote *cis*- or *trans*-splicing of exon 7 either as a research tool or as a therapeutic for treating SMA. The present invention further provides methods of treating SMA by administering a therapeutically effective amount of an aptamer including a polynucleotide sequence of SEQ ID NOs:253 to a subject with SMA. In one embodiment, the aptamer is modified to enhance stability or transfer into a cell. For example, the nucleotide sequence of SEQ ID NOs:253 can have a 2'-O-methyl, phosphorothioate, or 5'-PEG backbone modification. Other modifications will be apparent to one skilled in the art. In one embodiment, the aptamer is bifunctional, i.e., includes an antisense sequence to facilitate annealing to the pre-mRNA. In some embodiments, the aptamer includes antisense sequence targeted to the SMN2 pre-mRNA to facilitate annealing to the SMN2 pre-mRNA.

The invention provides powerful methods for iterative cell-based selection to identify exonic positions that modulate the alternative splicing of exons. This is the first report of iterative cell-based selection of an entire exon. The methods described herein provide an 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 multi-target selection (Vant-Hull et al., J Mol Biol. 278:579-597 (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. As expected from a cell-based 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., Cell 108:501-512 (2002)), nonsense-altered splicing (Wang et al., Mol Cell 10:951-957 (2002)), nonsense-mediated decay (Maniatis & Reed, Nature 416:499-506 (2002)) and nuclear transport (Reed & Hurt, Cell 108:523-531 (2002)). Thus, the selected sequences, all of which showed no detectable amount of exon 7 excluded product, represent a valuable archive for extracting position-specific information within the entire exon.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

Other features and advantages of the invention will be apparent from the following detailed description, and from the claims.

DESCRIPTION OF DRAWINGS

FIG. 1A is a schematic illustration showing the construction of minigene cassettes of *SMN1* (*pNT1*) and *SMN2* (*pNC4*) containing truncated intron 6. The amplified product as shown deleted the major portion of intron 6.

FIG. 1B is a schematic illustration showing the effect of intron 6 truncation on splicing. X indicates a band of unknown origin. Numbers on the right correspond to the products that are mentioned on the left. Lane-to-lane differences in the level of expression are either due to the differences in transfection efficiency or due to the shorter half-life of the longer transcripts. Because of the shortened intron 6, the unspliced precursors as well as all of the splicing intermediates are visible in the gel. The intron 7 spliced products are not visible in the longer constructs (*SMN1* and *SMN2*) because of the large size of intron 6. *SMN2* and *NC4* did not show the spliced products of intron 6 because of the weak 3' splice site (3'-ss). The defective nature of the 3'-ss has been confirmed by others. The low levels of the spliced intron 7 products in the case of *NC4* could be due to the faster rate of exon 7 skipping. The relative levels of splicing products were not affected by shortening of intron 6. Loading was adjusted based on the similar amount of simultaneously amplified HPRT as an internal control.

FIG. 2 is a schematic illustration showing steps involved in the introduction of random exon 7 sequences into the splicing cassette (not to scale). The construct pNT1 as discussed in Figure 1 was used as the backbone plasmid. The primer E7Rand contains the 54-nucleotide partially random sequence corresponding to exon 7 (shown as a black box). Incorporation of this sequence into construct E6-I6-E7-I7-E8 was achieved by 2-step PCR and subsequent cloning into pBxT2 vector. The pBxT2 vector is similar to

pNT1 except exon 7 is reduced to 27 nucleotides containing the *Bsa*X1 restriction site and the original *Bsa*X1 restriction site from the pCI backbone has been deleted. The resulting construct represents a pool of DNA templates containing randomized exon 7 sequences. I, E, and R stand for intron, exon, and random, respectively.

5 Fig. 3 shows the sequences of 42 clones (SEQ ID NOS:2-43) from the initial random pool (pool-0). The wild-type sequence (SEQ ID NO:1) of exon 7 of the *SMN1* and *SMN2* genes is given at the top of the figure. The number at the top as well as at the bottom denotes the nucleotide position in exon 7. The number of substitutions per molecule is given on the right. Some molecules are more random than the others as
10 expected from the probability calculation. The distribution of nucleotides at every position of exon 7 is given at the bottom. All positions were confirmed as random. Clones that had deletions in exon 7 are not included.

 Fig. 4A is a schematic illustration showing one embodiment of an experimental procedure for cell-based SELEX (not to scale). A minigene containing the randomized
15 internal exon 7 is transfected into C33a cells. Splicing leads to exon 7 inclusion or skipping. Only fully spliced product (E6-E7R-E8) is isolated - the exon 7 contained within is enriched in sequences that promote exon inclusion. The exon 7 sequence of the fully spliced product is re-amplified by a second PCR step and subcloned into the splicing cassette for a subsequent round of selection.

20 FIG. 4B is a schematic illustration of the amplification steps of one embodiment of the cell-based selection methods described herein. The first PCR step amplified the spliced exon 7 with flanking 5' exon-6 and 3' exon-7 sequences. The second PCR step substitutes some bases to introduce the *Bsa*X1 restriction site. Upon digestion with *Bsa*X1 endonuclease, exon 7 is excised in a single fragment. This fragment is gel-
25 purified and ligated to the *Bsa*X1 linearized plasmid BsaXT2. The ligated plasmid is directly used for further cycling through transfection of C33a cells. A portion of the ligated mixture is used for checking the characteristics of sequences of individual clones, using bacterial amplification.

 FIG. 5A is a reproduction of a gel showing a comparative analysis of splicing
30 patterns of four different SELEX pools. The relative amounts of exon 7 included products are indicated at the bottom. Numbers at the right indicate the size of the products. An intermediate band corresponding to 306 nucleotides between exon 7 skipped and fully spliced products results from the splicing of undigested BsaXT2

plasmid. The initial pool (lane 3) showed more than 50% spliced products as exon 7-skipped. Pool-4 has enriched sequences that promote exon 7 inclusion. As shown, barely any detectable skipped products are found in pool-4 (lane 6).

FIG. 5B is a reproduction of a gel showing the splicing pattern of clones from pool-4. Reflecting the characteristics of pool-4, all clones showed no detectable amount of the exon 7 skipped products.

FIGs. 6A-C are representations of selected variants (SEQ ID NOs:44-236) of all 54 nucleotides of exon 7 of the *SMN1* gene, identified by primary cell-based SELEX, and variants of the first 52 nucleotides of exon 7 of the *SMN1* gene identified by secondary cell-based SELEX.

FIG. 6D is a schematic illustration showing the mutability of residues in exon 7, based on the results of cell-based 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. Inhibitory nature of residues covering Exinct has been recently described (Singh et al., Biochem Biophys Res Commun 315:381-388 (2004)). Consistent with the inhibitory nature of the 3'-Cluster, deletions and mutations in this region promoted exon 7 inclusion in *SMN2* (Fig. 11A-C). Novel mutations supporting the stimulatory role of the long-conserved-tract are shown in Fig. 12A-B. The exceptionally high mutability of position 54 is consistent with the dominant effect of A54G substitution on exon 7 inclusion as shown in Fig. 13A-B.

Fig. 6E is a representation of sequences (SEQ ID NOs:290-348) resulting from cell-based selection of exon 7 sequences. 59 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 small-case letters. The number of substitutions per molecule is given in the right-most column (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*.

FIGs. 7A-7B are schematic illustrations of aptamer-mediated inclusion of exon 7 during pre-mRNA splicing of the *SMN2* gene. Possible interactions with transacting factors are indicated. All numbers in boxed areas refer to exons. Wavy arrows represent

transacting factors (unknown) that promote exon 7 inclusion. FIG. 7A is a diagrammatic representation of annealing of aptamers to exon 7. The position of annealing can be optimized. FIG. 7B is a diagrammatic representation of exon 7 inclusion by the *SMN2* gene through spliceosome-mediated *trans*-splicing (SMaRT). Two possible positions of annealing have been shown in part I and II. In scheme I, the PTM will have annealing sequences that will block the wild-type polypyrimidine tract and 3'-ss. In scheme II, the PTM will have annealing sequences that will be in the more accessible region of intron 6.

FIG. 8 is a diagrammatic representation of the RNA secondary structure of exon 7 of the *SMN2* transcript (SEQ ID NO:349). The numbering starts from exon 7. The 54-nucleotide long exon 7 is represented by upper case letters. The lower case letters at the 5' and 3' end of the molecule represent intron 6 and intron 7 sequences, respectively. The RNA structure was probed by chemical and enzymatic methods (unpublished results). The local RNA secondary structure of exon 7 consists of two terminal stem loops (TSL1 and TSL2), three internal loops (IL1, IL2, and IL3) and one internal stem (IS1). TSL1 is exclusively formed by exonic sequences. The results of cell-based selection predicted the inhibitory nature of terminal stem (TS1); the highly mutable residues are shown in circles. The internal loop 1 (IL1) connecting TSL1 with IS1 is formed by intron 6 sequences. Intron 6 and exon 7 form the internal stem IS1 by extended base pairing of 10 nucleotides. TSL2 is mostly formed by exon 7 sequences, though the first two bases of intron 7 are also buried in TS2 (last two lower case letters at the 3' end of molecule).

FIG. 9A is a diagrammatic representation of the *trans*-splicing of *SMN* gene(s) in C33a cells. The feasibility of *trans*-splicing was checked by co-transfecting the *trans*-splicing cassettes pEX6Int6 (5'-L) and pInt6Ex7 (PTM). The 5'-L contains the entire exon 6 sequence followed by the first 100 nucleotides of intron 6 of *SMN1*. The PTM contains the last 100 nucleotides of intron 6 followed by the entire exon 7. The pre-mRNA derived from these two constructs should anneal through base pairing (as shown in scheme I) as predicted by mfold analysis. The PTM should also base pair with *SMN2* pre-mRNA (NC4 construct) as shown in scheme II.

FIG. 9B is a reproduction of a gel showing the results of the experiment described in 9A. The expected size of the amplified bands is marked. The bands were generated through specific primers that anneal in exon 6 (of 5'-L or NC4 or *SMN1*/*SMN2* transcripts) and vector sequences of PTM construct. As expected, the *trans*-spliced products are absent when cells were transfected with 5'-L or PTM or NC4 alone. The

thick bands in lanes 1 and 3 represent unspliced precursors. Band X in lanes 4 and 5 is non-specific. The *trans*-spliced bands were confirmed by sequencing. The efficiency of *trans*-splicing was about 5% to that of the *cis*-splicing. The *trans*-splicing was achieved strictly through natural base pairing of 5'-L (or NC4) and PTM, as no guide sequences and secondary structure were added. Addition of aptameric sequences and other modification (such as guide sequences and secondary structure) will increase the efficiency of *trans*-splicing.

FIG. 10A is a diagrammatic representation of the probable location of the annealing site of a bifunctional aptamer.

FIG. 10B shows the sequence and structure of a representative bifunctional aptamer. A 24 base long bifunctional aptamer sequence has been shown. The annealing sequence is underlined, where as the flanking sequence is in bold letters. The annealing sequence will break the inhibitory TSL1 as shown in Figure 10A. The size and composition of annealing sequence can be changed for effective pairing. The flanking sequence will form a GNRN (SEQ ID NO: 350) tetra-loop, which is known to be stable. The flanking sequence was derived by cell-based selection. This aptamer should promote exon 7 inclusion from the *SMN2* gene.

FIG. 11A is 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 non-human *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).

FIG. 11B is a comparison of sequences *SMN2* mutants showing mutations validating the inhibitory nature of the 3'-Cluster. Dashes indicate deletions. All mutations are located within the 3'-Cluster.

FIG. 11C is a reproduction of a gel, showing the splicing pattern of the mutants shown in FIG. 11B. Numbers and nucleotides represent the positions and the types of substitution within exon 7. Spliced intermediates and products are the same as marked in Fig. 5A.

FIG. 12A shows the sequences of seven exon 7 mutants. All mutations were made in *SMN1*. Substitutions are restricted to the highlighted area. Critical positions (25G26G) of Tra2- β 1 interaction have also been highlighted.

FIG. 12B is a reproduction of a gel showing the splicing pattern of mutants as shown in panel A, illustrating the effect of substitutions within the long conserved tract in the middle of exon 7 of SMN1. 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 Fig. 4A.

FIG. 13A shows the sequences of *SMN1* and *SMN2* mutants. Substitutions are restricted to the highlighted area. Critical positions (25G26G) of Tra2- β 1 interaction have been highlighted.

FIG. 13B is a reproduction of a gel showing the splicing pattern of *SMN1* and *SMN2* mutants as shown in panel A, illustrating mutations that validate the dominant role of terminal positions. Numbers and nucleotides represent the positions and the types of substitution within exon 7. Spliced intermediates and products are the same as marked in Fig. 5A.

FIG. 14 is a schematic diagram of *SMN* construct containing C6U mutation within exon 7 and the preferred splicing pathways. Capital letters represent exon 7 sequence. Small-case letters represent intronic sequences. Tra2-ESE is located in the middle of exon 7, whereas SF2/ASF-ESE is located towards the 3' ss of exon 7 and overlaps with hnRNP A1-ESS. C6U is indicated by a star. ss stands for splice site.

DETAILED DESCRIPTION

Alternative splicing is a major predominant source of protein diversity. The cell-based functional selection methods described herein are valuable for understanding the mechanisms behind alternative splicing. The methods can simultaneously predict positive as well as negative elements, over a broad span of sequence. The present method can include all exonic nucleotides, which form different contexts in different exons. This method predicts not only the presence of ESS/ESE elements but also the relative impact of single nucleotides and groups of nucleotides, due to their positioning with respect to splice sites. The present invention provides, *inter alia*, methods to identify positive and negative splicing elements in long sequences via a cell-based SELEX technique.

General Methodology

Described herein are methods for identifying exonic regions and sequences that modulate, e.g., enhance or inhibit, splicing of pre-mRNA. These methods generally include:

- (i) providing a pool of nucleic acids including partially randomized exonic sequences;
- (ii) subjecting the pool of nucleic acids to one or more rounds of selection in a living cell;
- (iii) selecting those nucleic acids with pre-selected splicing characteristics, i.e., inclusion or exclusion of an exon; and
- (iv) determining the distribution of mutations that are associated with the pre-selected splicing characteristics.

The distribution of mutations in hyper- and hypo-mutable regions corresponds to negative (not crucial to splicing) and positive (crucial to splicing) wild type nucleotides, respectively.

In some embodiments, the new methods include techniques based in part on cell-based SELEX (Systematic Evolution of Ligands by Exponential Enrichment) methodology. The methods, which are exemplified in Figure 4A, can include some or all of the following.

First, a pool of partially random oligonucleotides comprising an exon or a portion thereof is expressed in cells, e.g., eukaryotic cells, where the oligonucleotides undergo splicing. This can be achieved by transfecting the cells with plasmids suitable for expressing the partially random oligonucleotides.

The partially random oligonucleotides are then subject to selection as follows. Spliced mRNA is isolated from the transfected eukaryotic cells. An exon-included product is amplified from the isolated mRNA, and the exon-included product including an oligonucleotide is isolated. Double cutter or quadruple cutter restriction enzyme sites are inserted outside the boundary of the included exon, to create double cutter- or quadruple cutter-engineered oligonucleotides (e.g., oligonucleotides flanked on at least one end with a double- or quadruple-cutter recognition site). The engineering of double cutter and quadruple cutter restriction enzyme sites into splicing cassettes allows selective excision of an entire exon under study, without altering the boundary sequence. The engineered oligonucleotides are digested with the appropriate double cutter or quadruple

cutter restriction enzyme, and the exon is purified. The purified exon is then ligated into a double cutter- or quadruple cutter-linearized plasmid; and the plasmid is used to transfect cells, e.g., eukaryotic cells.

After transfection, the cells are incubated for a sufficient amount of time to allow for transcription and splicing to occur (e.g., 24 hours), total RNA is isolated, and the spliced or unspliced intermediaries are amplified, e.g., using RT-PCR. If desired, the amplified products are digested with the appropriate double cutter or quadruple cutter restriction enzyme and ligated back into a linearized plasmid, and subjected to another round of selection. The pool can be subjected to sufficient rounds of selection until a desired percentage of exon inclusion is reached, e.g., about 70%, 80%, 90% or 100% exon inclusion. The sequences of the selected clones are then determined using known sequencing techniques.

In some embodiments, the methods also include transfecting bacterial cells with the ligated plasmid described above, which include an exon, and analyzing the sequence of the exon to determine where wild-type nucleotides are retained or where non-wild-type nucleotides are substituted, to determine the distribution of mutations, and designating a position where wild-type nucleotides are retained as a portion of a potential positive splicing element and positions where non-wild-type nucleotides are substituted as a portion of a potential negative splicing element.

The methods of the present invention are not intended to be limited to a particular order of steps. Variations on the order of the steps will be apparent to one skilled in the art, to achieve the goals of the present invention. In some embodiments, the amplification steps employ polymerase chain reaction (PCR), particularly RT-PCR. The pool of oligonucleotides can be performed more than once, e.g., two, three, four or more times.

In the context of the present invention, "positive splicing element" means an exonic splicing enhancer (ESE). In the context of the present invention, "negative splicing element" means an exonic splicing silencer (ESS). ESE and ESS sequences are short sequences (8-12 nucleotides) within or near the exon that facilitate or inhibit splicing. RNA-processing proteins bind to ESE sequences and enhance splicing. ESE elements are preserved by cell-based SELEX. RNA-processing proteins bind to ESS sequences and inhibit splicing. Cell-based SELEX favors the removal of ESS elements. By "element" is meant *cis*-elements, defined as a sequence or order of sequences on one RNA molecule. These sequences or parts of sequences can constitute one or many RNA

secondary or higher order structures that are important only when their relative positioning is conserved in a given RNA molecule.

This method of cell-based SELEX can be used to scan an entire exon or
 5 nucleotide sequences flanking exon boundaries for splicing enhancers and silencers. The present invention provides a method of identifying positive and negative splicing elements within an entire exon.

The methods and results described herein establish the feasibility of iterative cell-
 based selection of an entire exon, and identify the relative significance of many positions
 10 within a critical exon of *SMN* that is associated with SMA, a debilitating disease of infants and children.

While the present method is exemplified below for an entire exon, the present method is also applicable to shorter sequences, particularly the 5' and 3' exonic boundary
 15 sequences. With currently known techniques, there is difficulty in inserting exons into the cassette without altering the boundary sequences. The present invention was facilitated by the use of double cutter and quadruple cutter restriction enzymes. The engineering of double cutter and quadruple cutter restriction enzyme sites into splicing cassettes allows selective excision of an entire exon under study, without altering the
 20 boundary sequence. While *BsaXI* (a quadruple cutter restriction enzyme) was used in the Examples below, other enzymes can also be used. Double cutter and quadruple cutter enzymes can be used in various embodiments in the present invention. Double cutter enzymes cut two times towards the 3' end of the enzyme recognition motif. Double cutter enzymes can be used singly or in combination with another double cutter enzyme. For
 25 instance, one double cutter enzyme can be used at the 5' end of the exon and a different double cutter enzyme can be used at the 3' end of the exon. Exemplary double cutter enzymes are: *AlwI* (1 nucleotide 5' overhang), *BbsI* (4 nucleotide 5' overhang), *BbvI* (4 nucleotide 5' overhang), *BciVI* (1 nucleotide 3' overhang), *BmrI* (1 nucleotide 3' overhang), *BpmI* (2 nucleotide 3' overhang), *BsaI* (4 nucleotide 5' overhang), *BseRI* (2
 30 nucleotide 3' overhang), *BsgI* (2 nucleotide 3' overhang), *BsmAI* (4 nucleotide 5' overhang), *BsmBI* (4 nucleotide 5' overhang), *BsmFI* (4 nucleotide 5' overhang), *BspMI* (4 nucleotide 5' overhang), *BsrI* (2 nucleotide 3' overhang), *BsrDI* (2 nucleotide 3' overhang), *BstF5I* (2 nucleotide 3' overhang), *BtsI* (2 nucleotide 3' overhang), *EarI* (3 nucleotide 5' overhang), *EciI* (2 nucleotide 3' overhang), *FauI* (2 nucleotide 5' overhang),
 35 *FokI* (4 nucleotide 5' overhang), *HgaI* (5 nucleotide 5' overhang), *HphI* (1 nucleotide

3' overhang), *Mbo*II (1 nucleotide 3' overhang), *Mly*I (blunt end), *Mn*II (1 nucleotide 3' overhang), *Ple*I (1 nucleotide 5' overhang), *Sap*I (3 nucleotide 5' overhang), and *Sfa*NI (4 nucleotide 5' overhang).

Quadruple cutters can also be used. These restriction enzymes cut two times
5 towards the 3' end and two times towards the 5' end of the recognition motifs. Exemplary quadruple cutters are *Bae*I (5 nucleotide overhang), *Bcg*I (2 nucleotide overhang), and *Bsa*XI (3 nucleotide overhang). In some embodiments, *Bsa*XI is used.

While plasmid pBxT2 was used in the examples below, any plasmid with appropriate sites can be utilized; methods for engineering such plasmids are known in the
10 art. For example, if *Bsa*XI is used, any plasmid with *Bsa*XI sites can be used. If restriction enzymes other than *Bsa*XI are used (e.g., other double cutter or quadruple cutter enzymes known in the art, as discussed above), other appropriate plasmids can be used. Such plasmids are commercially available or can be readily generated from commercially available plasmids.

15 In one embodiment of the present invention, a library of partially random (doped) oligonucleotides is created in splicing cassettes, which are inserted into a plasmid for transfection into a eukaryotic cell. In one embodiment, the pCI plasmid human expression vector is used to insert the partially random exonic sequences. The degree of randomization can be confirmed by sequencing a number of clones. In some
20 embodiments, 40 or more clones are sequenced to confirm the desired level of randomization. In the examples shown below, human cells were transfected with about 10^{12} molecules (of which about 10^{11} are unique). After splicing takes place, mRNA, with the included exon, is isolated and purified. *Bsa*XI sites are inserted into the mRNA and digested with *Bsa*XI. The *Bsa*XI-digested exon sequence is inserted into a *Bsa*XI-
25 digested plasmid. This *Bsa*XI-engineered plasmid is then transfected back into a eukaryotic cell to go through subsequent rounds of cell-based SELEX. Plasmids containing sequences that promote the exon of interest are selectively amplified and can be analyzed via amplification in bacterial cells.

30 In the examples of the present invention, human cells are transfected with the splicing cassettes for the splicing event. Human cervical cancer C33a cells can be used in the present invention due to their high transfectability. However, other human or animal cell lines can be used as well. Other suitable human and animal cells are motor neuron-like NSC34 cells (Usuki et al., *Neurochem. Res.* 24:281-286 (1999)) and fibroblast cells

from SMA patients (Andreassi et al., *Human Mol. Genet.* 10:2841-2849 (2001); DiDonato et al., *Hum. Gene Ther.* 14:179-188 (2003); Skordis et al., *Proc. Natl. Acad. Sci. USA* 100:4114-19 (2003)). For analysis of sequences, individual clones were purified from *E. coli* and sequenced. While *E. coli* is commonly used for plasmid purification, other bacterial or fungal cells could also be used for amplifying clone sequences for analysis. Clones that produce inclusion of the exon of interest are analyzed to determine which nucleotides (or residues) within the exon are preserved and which are not. This analysis can be performed by multiple sequence alignment (Higgins et al., *Methods Enzymol.* 266:383-402 (1996); Thompson et al., *Nucl. Acid Res.* 25:4876-4882 (1997)), score matrices (Liu et al., *Genes Dev.* 12:1998-2012 (1998)) and RESCUE-USE (Fairbrother et al., *Science* 297:1007-1013 (2002)). Independent algorithmic programs based on the nature of the selected sequences can also be developed. Sequences identified by cell-based SELEX will also provide new insights for the genome-wide analysis of alternative splicing. Such studies have been done with only a limited number of naturally expressed sequences (Burset et al., *Nucl. Acid Res.* 28:4364-4375 (2000); Modrek et al. *Nucl. Acid Res.* 29:5850-5859 (2001); Xu et al., *Nucl. Acid Res.* 30:3754-3766 (2002)).

Once primary cell-based SELEX is performed, secondary cell-based SELEX can be performed by randomizing smaller regions of the exon. One advantage of the present invention is that it is possible to observe covariation of nucleotide position in the context of the entire exon. It should be mentioned that available methods do not allow determination of covariation of exonic nucleotides that are separated by more than 6 nucleotides. However, the cell-based SELEX method possesses the power to predict such covariation. For example, this method has predicted that the inhibitory effect of position 1 of exon 7 could be compensated for by another mutation at position 16. The present method can also be used to determine the effects of mutation on RNA secondary structure as exemplified below.

“Saturation of sequence space” relates to the number of possibilities if all positions in an oligonucleotide are totally randomized. For example, a 4 nucleotide sequence would have a saturation space of 4^4 , or 256. In the examples shown below, it is not practical or desirable to totally randomize an entire exon, including exon 7 of *SMN2*, which is 54 nucleotides in length. However, it is possible to randomize the sequence enough to check mutability.

The present method utilizes a mixture of about 70% wild-type and about 30% non-wild type nucleotides, which gives no cryptic splice sites. The use of less than about 70% wild-type can yield cryptic splice sites. Overrandomization can produce a completely different exon. Underrandomization cannot test the mutability of every
5 position within the exon. With about 70% wild-type nucleotides, a set of 10 clones will, on average, produce 7 that are wild-type.

Generally, one exon of about 100 nucleotides can have 6 or more *cis* elements (ESE or ESS) that influence splicing. The present method enables one skilled in the art to determine which of these elements are more important. That is, the new methods can be
10 used to determine the relative impact of elements, and which elements are positive and which are negative.

The methods described herein have numerous applications. Once one determines which sequences in an exon are positive or negative, the sequence can be manipulated to enhance splicing. For instance, ESS sequences can be blocked by antisense
15 oligonucleotides to prevent the binding of splicing-inhibitory proteins. Alternatively, ESE sequences identified in the methods described herein can be provided in *trans* to facilitate binding of splice-enhancing proteins. Such oligonucleotides would have at least two components (e.g., be bi-functional) - one part of the oligonucleotide would anneal to the exon itself (as "antisense"), while the second part would bind to a splice-enhancing
20 protein (as an aptamer). Use of bi-functional oligonucleotides has recently been demonstrated (Skordis et al., *Proc. Natl. Acad. Sci. USA* 100:4114-4119 (2003)). However, the methods described herein would substantially increase the efficacy of bi-functional oligonucleotides by providing methods for selecting sequences that are the best targets for manipulation in this manner.

25 Additionally, the identification of specific ESE elements will enhance the choice of targets for gene therapy, e.g., replacement of mutated ESE elements with normal wild-type or even enhanced sequences. This approach is particularly elegant because the therapy can be utilized without knowing the identity of the protein factor or factors that bind to the element.

30 Exonic *cis*-elements are known to play a role in many human diseases. Thus, the identification of these elements by the method of the present invention will facilitate the understanding of the etiology and provide new therapeutic targets for treating disease.

Conditions Associated with Aberrant Splicing

The methods described herein have application in the study and treatment of many conditions associated with RNA splicing errors. While the methods are exemplified with proximal spinal muscular atrophy (SMA), numerous other splicing error diseases are known. Exemplary RNA splicing diseases and the exons and introns involved are shown in Table 1. Exons from any of these genes listed in Table 1 can be used in the new methods.

TABLE 1. SPLICING-ASSOCIATED DISEASES

Gene/disorder	Mutation	Effect	Reference
Adenosine deaminase Severe combined immunodeficiency disease	R142X G>A and C>T in the same codon	Skipping of exon 5.	Santisteban et al., <i>Hum. Mol. Genet.</i> 4:2081-2087 (1995)
Arylsulfatase A Metachromatic leukodystrophy	Thr409Ile C>T	Activates a cryptic splice site.	Hasegawa et al., <i>Hum. Genet.</i> 93:415-420 (1994)
Beta-globin Beta-Thalassemia	T>G transversion in position 705 of intron 2	Activates cyptic 39 splice site in intron 2.	Dobkin and Bank, <i>Prog. Clin. Biol. Res.</i> 134:127-128 (1983) Dobkin et al., <i>Proc. Natl. Acad. Sci. USA</i> 80:1184-1188 (1983)
Beta-hexaminidase Beta-subunit Sandhof disease	Exon 11, C8T (P417L) Intron 10, A17G	Disrupts an ESE. Causes use of a cryptic splice site at nucleotide 1112; Causes usage of a cryptic splice site at positon 237. Disrupts ISE. The conversion also disrupts a putative branch point.	Fujimaru et al., <i>Hum. Genet.</i> 103:462-469 (1998)
BRCA1 Breast and ovarian cancer	E1694X G>T	Disrupts ESE and causes skipping of exon 18.	Liu et al., <i>Nat. Genet.</i> 27:55-58 (2001)
CD45 Individuals do not suffer from obvious immunodeficiency, but it is notable that no homozygotes have been described. Associated with Multiple sclerosis in the American population.	Silent C>G at position 77 of the alternative exon 4	Disrupts an ESE. Constitutive inclusion of exon 4.	Jacobsen et al., <i>Nat. Genet.</i> 26:495-499 (2000) Lynch and Weiss, <i>J. Biol. Chem.</i> 276:24341-24347 (2001) Zilch et al., <i>Eur. J. Immunol.</i> 28:22-29 (1998)
CYP 27 Cerebrotendinous xanthomatosis	Silent G112G, G>T in exon 2	Creates a cryptic splice site. Causes skipping of the entire exon 2.	Chen et al., <i>Biochem.</i> 37:4420-4428 (1998)

Gene/disorder	Mutation	Effect	Reference
DMPK Myotonic dystrophy	Expanded (CUG)>40 in the 3' UTR of DMPK	Two possibilities: CUG repeats sequester CUG-BP and therefore prevent it from its normal function; or expanded CUG repeats alter the alternative splicing of the DMPK-mRNA by developing a new 3' splice site.	Phillips et al., <i>Science</i> 280:737-741 (1998) Tiscornia and Mahadevan, <i>Mol. Cell</i> 5:959-967 (2000)
Fibrillin-1 Marfan syndrome	Silent C>T in exon 51	Skipping of exon 51.	Liu et al., <i>Nat. Genet.</i> 16:328-329 (1997)
Fumarylacetoacetate hydrolase Hereditary tyrosinemia type 1	N232N C>T	Skipping of exon 8.	Ploos van Amstel et al., <i>Hum. Genet.</i> 97:51-59 (1996)
Integrin GPIIIa Glanzmann thrombasthenia	C116A and silent mutation G1134A in exon 9	Skipping of exon 9.	Jin et al., <i>J. Clin. Invest.</i> 98:1745-1754 (1996)
MNK Menkes disease	Gly>Arg G>A	Skipping of exon 8.	Das et al., <i>Am. J. Hum. Genet.</i> 55:883-889(1994)
NF-1 Neurofibromatosis type 1	Y2264X C>A, C>G R304X C>T Q756X C>T	Skipping of exon 37. Skipping of exon 7. Skipping of exon 14.	Ars et al., <i>Nucleic Acids Res.</i> 28:1307-1312 (2000a), <i>Hum. Mol. Genet.</i> 9:237-247 (2000b) Hoffmeyer et al., <i>Am. J. Hum. Genet.</i> 62:269-277 (1998) Messiaen et al., <i>Hum. Genet.</i> 101:75-80(1997)
Porphobilinogen deaminase Acute intermittent porphyria	R28R, C>G	Skipping of exon 3.	Llewellyn et al., <i>J. Med. Genet.</i> 33:437-438 (1996)
Pyruvate dehydrogenase E1 alpha Leigh's encephalomyelopathy	Silent A>G	Aberrant splicing of exon 6.	De Meirleir et al., <i>Pediatr. Res.</i> 36:707-712 (1994)
SMN1 Spinal muscle atrophy (SMA)	425del5 W102X	Skipping of exon 3.	Sossi et al., <i>Eur. J. Hum. Genet.</i> 9:113-120 (2001)
SMN2 Spinal muscle atrophy (SMA)	Silent C6T substitution in exon 7	Disrupts ESE, skipping of exon 7.	Covert et al., <i>Hum. Mol. Genet.</i> 6:1205-1214 (1997) Jablonka et al., <i>Hum. Mol. Genet.</i> 9:, 341-346 (2000) Lefebvre et al., 1995, <i>Cell</i> 80:155-165 (1995); <i>Nat. Genet.</i> 16:265-269 (1997) Lorson and Androphy, <i>Hum. Mol. Genet.</i> 9:259-265 (2000) Lorson et al., <i>Proc. Natl. Acad. Sci. USA</i> 96:6307-6311 (1999) Monani et al., <i>Hum. Mol. Genet.</i> 9:333-339 (2000) Vitali et al., <i>Hum. Mol. Genet.</i> 8:2525-2532 (1999)

Gene/disorder	Mutation	Effect	Reference
Tau Frontotemporal Dementia with Parkinsonism linked to chromosome 17	Intron 10: 113 A>G, 114 C>T, 116 C>T, IVS10103 G>A	Disrupt ISE IVS1013 G>A improves slightly the splice site score (from 6.8 to 7.0)	D'Souza et al., <i>Proc. Natl. Acad. Sci. USA</i> 96:5598-5603 (1999) D'Souza and Schellenberg, <i>J. Biol. Chem.</i> 275:17700-17709 (2000)
	Exon 10: L284L T>C, S305S T>C, S305N G>A	Disrupt ESE S305N G>A improves the splice site. Disrupts ESE.	Hasegawa et al., <i>FEBS Lett.</i> 443:93-96 (1999) Hutton et al., <i>Nature</i> 393:702-705 (1998) Iijima et al., <i>Neuroreport</i> 10:497-501 (1999)
	Exon 10: N297K T>G, del280K (AAG deletion)		Spillantini et al., <i>Proc. Natl. Acad. Sci. USA</i> 95:7737-7741 (1998) Stanford et al., <i>Brain</i> 123:880-893 (2000)

These diseases and defects in pre-mRNA processing are reviewed in Stoilov et al. (*DNA & Cell Biol.* 21(11):803-818 (2002)). These genes are merely exemplified for use in the present invention. However, the new methods are not limited in application to these diseases, but can be used for the diagnosis of other diseases associated with pre mRNA processing/splicing defects, and for the analysis of other genes that undergo alternative splicing. The methods described herein can also be used as a powerful research tool to analyze alternatively splicing genes, even if they are not associated with a known disease.

Spinal Muscular Atrophy

Proximal spinal muscular atrophy (SMA) is the second most common autosomal recessive disorder characterized by loss of motor neurons in the anterior horn of the spinal cord. (Pearn, *Lancet* 8174:919-922 (1980)). Linkage mapping identified the Survival of Motor Neuron 1 (*SMN1*) gene as the genetic locus of SMA (Lefebvre et al., *Cell* 80:1-5 (1995)). In humans, two nearly identical *SMN* genes (*SMN1* and *SMN2*) exist in a duplicated region on chromosome 5q13 (Lefebvre et al., *Cell* 80:1-5 (1995)). Deletions or mutations within *SMN1* but not the *SMN2* gene cause all forms of proximal SMA (Lefebvre et al., *Cell* 80:1-5 (1995)). Although the two genes code for identical proteins, *SMN2* produces predominantly transcripts lacking in exon 7 due to a critical C to T substitution at position 6 (C6T) within exon 7 (Lorson et al., *PNAS* 96:6307-6311 (1999); Moniani et al., *Hum. Mol. Genet.* 8:1177-1183 (1999)). This leads to production of mostly truncated protein SMN Δ 7, which is incapable of compensating for the loss of the full-length protein (Lorson et al., *Nat. Genet.* 19:63-66 (1998); Lorson and Androphy, *Hum. Molec. Genet.* 9:259-265 (2000)). The SMN protein is vital for survival as it participates in the assembly of essential ribonucleoprotein particles present in cytoplasm

and nucleus (Liu and Dreyfuss, *EMBO J.* 15:3555-3565 (1996); Liu et al., *Cell* 90:1013-21 (1997); Wang et al., *Proc. Natl. Acad. Sci. USA* 99:13583-13588 (2002); Meister et al., *Trends Cell. Biol.* 12:472-478 (2002); Pellizzoni et al., *Science* 298:1775-1779 (2002)).

The function of SMN Δ 7 protein is not known. However, the deletion of the *SMN2* gene has been recently linked to a number of neuronal diseases (Veldink et al., *Neurol.* 56:749-752 (2001); Echaniz-Laguna et al., *J. Neurol.* 249:290-293 (2002); Rojas-Garcia et al., *Neurol.* 59:1112-1113 (2002)).

In the *SMN1* gene, a method of score matrices was successfully applied to identify an SF2/ASF-associated ESE within exon 7 (Cartegni and Krainer, *Nat. Genet.* 30:377-384 (2002)). Confirming this observation, disruption of SF2/ASF-associated ESE resulted in exon 7 exclusion. This led to the conclusion that C6T transition, which abrogates ASF/SF2-associated-ESE, is the primary cause of SMA. However, an equally plausible and mutually inclusive hypothesis, which will allow a simultaneous gain of an ESS as a consequence of C6T transition, was not ruled out. Until recently, the loss of ASF/SF2-associated ESE was thought to be the only distinguishing feature between the two SMN genes. The data described herein demonstrate that both splice sites in exon 7 are inherently weak. Strengthening either splice site eliminates the differential inclusion of exon 7 by the two SMN genes. The lack of selection of a strong ASF/SF2-associated ESE combined with high-density contiguous substitutions at positions 3, 4, and 5 strongly suggest the presence of an ESS at the 3'-ss.

Alternatively, an already present ESS element can function to suppress splice-site selection in the absence of an active ASF/SF2-associated-ESE, such as in the case of IgM exons M1 and M2 (Kan and Green, *Genes Dev.* 13:462-471 (1999)). The role of ASF-SF2-associated-ESE seems to be dispensable, as elevated expressions of hTra2- β 1 (Hofmann et al., *Proc. Natl. Acad. Sci. USA* 97:9618-9623 (2000)), SRp30c (Young et al., *J. Biol. Chem.* 277:2852-2859 (2002)) and hnRNP-G (Hofmann and Wirth, *Hum. Mol. Genet.* 11:2047-2049 (2002)) have been shown to promote exon 7 inclusion from *SMN2* gene. While hTra2- β 1 directly interacts with an ESE within exon 7, the effect of SRp30c and hnRNP-G interacts via hTra2- β 1.

The genomic sequences of the *SMN1* and *SMN2* genes are found in GenBank Accession Nos. AC04497 and AC010237. Most SMA patients have the *SMN2* gene and, therefore, improving the levels of exon 7 inclusion by the *SMN2* gene can be used to cure or treat SMA. The present invention provides aptamers that can be used to promote exon

7 inclusion from the *SMN2* gene during pre-mRNA splicing. Aptamers are defined as sequences that are selected by SELEX for a particular function. RNA-aptamer-mediated gene therapy is a rapidly growing field because of the strong specificity of aptamers to the targeted proteins (Conrad et al., *J. Biol. Chem.* 269:32051-32054 (1994); Wilson and Szostak, *Annu. Rev. Biochem.* 68:611-648 (1999); Brody and Gold, *J. Biotechnol.* 74:5-13 (2000); Siewert et al., *Chem. Biol.* 7:833-843 (2000); Sullenger and Gilboa, *Nature* 418:252-258 (2002); Rusconi et al., *Nature* 419:90-94 (2002)). Aptamers have been generated against one or more proteins. (Fitzwater and Polisky, *Methods Enzymol.* 267:275-301 (1996); Vant-Hull, *J. Mol. Biol.* 278:579-597 (1998)) using *in vitro* SELEX. In the present methods, cell-based SELEX has been used to isolate sequences that promote exon 7 inclusion. Using cell-based SELEX, exon 7 sequences were randomized and only those sequences that included exon 7 during pre-mRNA splicing were selected. The selected sequences (aptamers) reflect ideal combinations of nucleotides for productive interaction of splicing factors. Interactions of the aptamers with splicing factors can be preserved even if the aptamer sequences are not flanked by adjacent introns. This is particularly true for interactions with SR and SR-related proteins, whose average binding sites are less than ten nucleotides (Manley and Tacke, *Genes Dev.* 10:1569-1579 (1996)). Supporting this argument, small stretches of exon 7 sequence have been shown to compete for binding to SR-like protein hTra2- β 1 in UV-crosslinking experiments. (Hofmann et al., *Proc. Natl. Acad. Sci. USA* 97:9618-23 (2000)).

Animal models of SMA are known in the art, including those described in Frugier et al., *Hum. Mol. Genet.* 9(5):849 (2000); Hsieh-Li et al., *Nat Genet* 24:66-70 (2000); Monani et al., *Hum Mol Genet* 9:2451-2457 (2000).

RNA Structure

The present invention provides the information that can be used to determine secondary or higher order RNA structure that could be the subject of drug targeting. Non-favorable RNA structures tend to be discarded during cell-based selection. For example, if the variants identified by the cell-based selection methods described herein have increased substitution in a particular region, one can surmise that the substitutions bring about a favorable change in secondary or higher order structure in the RNA, *e.g.*, breaking a stem can result in increased splicing.

Recent studies by others demonstrated that using chimeric oligonucleotides or

bifunctional oligonucleotides targeting sequences corresponding to TSL1 also improved exon 7 inclusion (Cartegni and Krainer, *Nat. Struct. Biol.* 10:120-125 (2003); Skordis et al., *Proc. Natl. Acad. Sci. USA* 100:4114-4119 (2003)). Although these investigators did not seek to study the role of RNA secondary structure in particular, the strategy invariably broke the structure of Terminal Stem Loop 1 (TSL1). As shown in the examples below, the present invention allows the identification of RNA secondary structures that can be targeted by antisense oligonucleotides to improve exon 7 inclusion (e.g., as shown in Figure 10A, in which the Terminal Stem Loop 1 (TSL1) is targeted). Thus, the present invention provides a method for probing and identifying differences in secondary or higher order RNA structure of an exonic sequence, by comparing the secondary or higher order structure of the wild type exon sequence with the secondary or higher order structure of an exonic sequence variant identified by cell-based SELEX. The secondary or higher order structure of the wild-type and cell-based SELEX sequences can be compared based on standard techniques known in the art, such as chemical and enzymatic digestion. For instance, unpaired adenosines and guanosines, and all paired bases can be probed by dimethyl sulfate (DMS), ribonuclease T1, and ribonuclease V1, respectively. Ehresmann et al. provide exemplary structure probing techniques. (Ehresmann et al., *Nucl. Acids Res.* 15:9109-9128 (1987)). Additional techniques are well-known in the art.

Methods of Promoting Splicing

The present invention also provides methods for promoting *cis*- or *trans*-splicing of exon 7 of the *SMN2* gene, by supplying a polynucleotide sequence of one or more nucleotide sequences described herein, e.g., one or more of SEQ ID NOs:2-236 and 247-283, to a eukaryotic cell, to promote *cis*- or *trans*-splicing of exon 7. These sequences can be used in their entirety or smaller fragments can be used. These fragments can include lengths of about 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 51, 52, and 53 nucleotides, and any in-between lengths. In one embodiment, the one or more nucleotide sequences (e.g., of SEQ ID NOs:2-236 and 247-283) further comprise an antisense sequence targeted to an appropriate region of the *SMN2* pre-mRNA, or a fragment thereof, to facilitate annealing to the *SMN2* pre-mRNA. The one or more nucleotide sequences of SEQ ID NOs:2-236 and 247-283 can be modified to enhance stability or transfer into the cell. In one embodiment, the one or more nucleotide sequences of SEQ ID NOs:2-236 and 247-283 have 2'-O-methyl, phosphorothioate, or 5'-PEG backbone modifications.

Other possible modifications will be apparent to those skilled in the art. The *cis*- or *trans*-splicing can be performed, *e.g.*, in a eukaryotic cell such as C33a human cervical cancer cells, motor-neuron-like NSC34 cells, and fibroblast cells from subjects with SMA. In one embodiment, the eukaryotic cell is a neuronal cell.

Aptamers

The present invention also provides polynucleotide sequences (aptamers) identified by the cell-based SELEX of the *SMN1/SMN2* genes. These sequences are exemplified in Figures 3, 6A-C and E, 10B, 11B, 12A, and 13A. The complement of any of the novel polynucleotide sequences identified herein is also encompassed within the invention. A "complement" of a nucleic acid as used herein refers to an anti-parallel or antisense sequence that participates in Watson-Crick base pairing with the original sequence. The wild-type sequence of exon 7, shown at the top of Figure 3, is SEQ ID NO:1. The polynucleotide sequences that have been optimized by cell-based SELEX include SEQ ID NOs:2-236, particularly SEQ ID NOs:44-236; additional sequences include 247-283. These sequences can be used in their entirety or smaller fragments can be used. These fragments can include lengths of about 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 51, 52, and 53 nucleotides, and any intermediate lengths. Moreover, the polynucleotide sequences can be modified to enhance stability or transfection/delivery into a cell, such as 2'-O-methyl, phosphorothioate, and polyethylene glycol (PEG) backbone modifications, or to enhance detection or purification, *e.g.*, radiolabels or biotinylation. Additionally, the present invention provides further modification of the polynucleotide sequences of SEQ ID NOs:2-236 and 247-283 via the addition of antisense sequences that aid in annealing the aptamers to an appropriate location on a pre-mRNA to facilitate *cis*- or *trans*-splicing. The present invention further encompasses compositions comprising one or more nucleotide sequences of SEQ ID NOs:2-236 and 247-283 and a carrier therefor. In one embodiment, the composition can be a pharmaceutical composition, and the carrier can be a pharmaceutically acceptable carrier or excipient.

Aptamers identified by the cell-based SELEX methods described herein (including SEQ ID NOs:2-236 and 247-283) can be used to identify tissue-specific or cellular-specific factors that interact with these aptamers, which can be therapeutic targets. Specific cellular extracts can be panned to identify these aptamer-interacting

factors. Thus, the present invention provides new methods of identifying tissue-specific or cell-specific RNA splicing-associated factors, by contacting a tissue-specific or cellular-specific extract with one or more aptamers identified by cell-based SELEX and isolating or purifying tissue-specific or cellular-specific RNA splicing-associated factors that bind to one or more aptamers. The aptamers can bind directly to RNA splicing-associated factors or indirectly through an intermediary. In some embodiments, neural-specific extracts are used, *e.g.*, motor-neuron-specific extracts. In other embodiments, the particular cell extract used can depend upon the tissue implicated in particular splicing diseases, *e.g.*, breast tissue or leukocytes. In various embodiments, the one or more aptamers are one or more of SEQ ID NOs:2-236, particularly SEQ ID NOs:44-236, or 247-283. Once tissue-specific or cellular-specific interacting factors have been identified, microarrays with these tissue-specific factors can be constructed. These microarrays can be used to determine the expression levels of the factors and can be used in diagnosing the presence and/or severity of RNA splicing-associated diseases.

Aptamers identified by cell-based SELEX methods described herein can also be used as experimental reagents to purify tissue-specific or cellular-specific RNA splicing associated factors. Thus, the present invention includes new methods of isolating or purifying tissue-specific or cellular-specific RNA splicing associated factors comprising contacting tissue-specific or cellular-specific cellular extracts with one or more aptamers identified by cell-based SELEX (*e.g.*, SEQ ID NOs:2-236, particularly SEQ ID NOs:44-236, and 247-283) and isolating or purifying tissue-specific RNA splicing associated factors that bind to the one or more aptamers. The aptamers can bind directly to RNA splicing-associated factors or indirectly through an intermediary. In some embodiments, the "lure" aptamers are modified (*e.g.*, biotinylation or phosphorylation at the 5' or 3' end, radiolabeling) to enhance detection or purification. Other modifications will be apparent to one skilled in the art. Fragments of aptamers can also be used for this method.

The present invention further includes methods of treating SMA by administering an aptamer having a polynucleotide sequence of one of SEQ ID NOs:2-236 and/or 247-283 to a subject with SMA. In one embodiment, the aptamer is modified to enhance stability or transfer into a cell. In some embodiments, the nucleotide sequence of SEQ ID NOs:2-236 and/or 247-283 has a 2'-O-methyl, phosphorothioate, or 5'-PEG backbone modification. Other modifications will be apparent to one skilled in the art. In one embodiment, the aptamer further comprises an antisense sequence of the SMN2 pre-

mRNA, or a fragment thereof, to facilitate annealing to the SMN2 pre-mRNA.

Different approaches can be taken to utilize the aptamers isolated by the new methods described herein. For example, small RNA molecules containing aptamer sequences can be targeted to exon 7 of the SMN2 pre-mRNAs through antisense oligos. Once the aptamer is anchored to exon 7, factors interacting with the aptamer will recruit other components of the splicing machinery. Alternatively, aptamer sequences can be used as an engineered exon 7 for spliceosome-mediated RNA trans-splicing (SMaRT; Puttaraju et al., *Nat. Biotechnol.* 17:246-252 (1999), *Mol. Ther.* 4:105-114 (2001)). Here, the mature transcript is derived from two pre-mRNAs. This approach allows the insertion of aptamers in pre-mRNA (also called the pre-*trans*-splicing molecule, PTM) that will be directed to splice with the natural pre-mRNA derived from the SMN2 gene. The success of *trans*-splicing will depend upon annealing of a PTM to the target pre-mRNA through appropriate guiding sequences (Rusconi, *Mol. Ther.* 4:162-163 (2001)). Based on RNA structural analysis, regions within SMN2 pre-mRNA that could serve as guide sequences have been identified. The SMaRT technique has been successful in correcting the cancer-associated pre-mRNA encoding β subunit of human chorionic gonadotropin gene 6 (Puttaraju et al., *Nat. Biotechnol.* 17:246-252 (1999)) and the endogenous DF508 cystic fibrosis transmembrane conductance regulator (CFTR) transcripts (Liu et al., *Nat Biotechnol.* 20:47-52 (2002)). The correction of SMN2 pre-mRNA splicing by the above two methods is attractive because such methods do not require excessive expression of a transacting factor(s), which is mostly accompanied by unwanted non-specific interactions. Further, the repair will be done on the resident SMN2 pre-mRNA and therefore the specificity of therapy is mostly driven by the natural expression pattern of SMN2 gene. A schematic of these two approaches is shown in Figure 7.

The new aptamers described herein can be bifunctional, i.e., modified by inclusion of antisense sequence, to target the aptamer to a specific region of pre-mRNA (see, e.g., Cartegni and Krainer, *Nat. Struct. Biol.* 10:120-125 (2003); Skordis et al., *Proc. Natl. Acad. Sci. USA* 100:4114-4119 (2003)). For example, a bifunctional aptamer can include antisense sequences directed to the SMN1/SMN2 genes, either exonic or intronic sequences, including polypyrimidine tracts. These antisense sequences are useful for targeting the aptamer to the particular region of the pre-mRNA, e.g., to facilitate *trans*-splicing by bringing a splicing-enhancing protein into proximity with the sequence to be spliced.

It is noted, however, that such antisense sequences are not necessary for *trans*-splicing to be successful, as the Examples below demonstrate. Moreover, numerous modifications to the nucleotide sequences, such as backbone modifications (e.g., 2'-O-methyl, phosphorothioate, and 5'-PEG), can be used to enhance stability and/or delivery into the cell.

Methods of Screening

Also included herein are methods of screening for test compounds that modulate (e.g., enhance or inhibit) splicing of a selected exon. For example, once a region or sequence has been identified by a method described herein as a region or sequence that modulates splicing, oligonucleotides can be designed to include that sequence and can be used to screen test compounds. For example, a simple binding assay can be used as an initial screen to detect activity of the test compound. Once binding is detected, the test compound can be considered a candidate compound and subject to additional assays, e.g., in a splicing assay as described herein. Additionally, where a protein that binds to the selected region is known or identified, the test compound can be evaluated for its ability to inhibit binding of the protein to the region. When binding of the protein results in inhibition of splicing, a test compound that inhibits binding of the protein would enhance splicing, and vice versa.

In some embodiments, the region or sequence that is used is the 3' cluster, the conserved region, or the Exinct region of SMN1 or SMN2.

The test compounds can be, *inter alia*, aptamers, peptides, peptoids, small molecules, ribozymes, antisense (including morpholino oligos), or small interfering RNAs (siRNAs). A skilled artisan can readily recognize that there is no limit to the structural nature of the compounds.

Peptide compounds can be prepared using standard solid phase (or solution phase) peptide synthesis methods, as is known in the art. In addition, the DNA encoding these peptides can be synthesized using commercially available oligonucleotide synthesis instrumentation and/or produced recombinantly using standard recombinant production systems. The production of polypeptides using solid phase peptide synthesis can be used where it is desirable to include non-nucleic acid-encoded amino acids.

Small molecules can be natural products or members of a combinatorial chemistry library. A set of diverse molecules should be used to cover a variety of functions such as

charge, aromaticity, hydrogen bonding, flexibility, size, length of side chain, hydrophobicity, and rigidity. Combinatorial techniques suitable for synthesizing small molecules are known in the art, e.g., as exemplified by Obrecht, D. and Villalgrodo, J.M., Solid-Supported Combinatorial and Parallel Synthesis of Small-Molecular-Weight
5 Compound Libraries, Pergamon-Elsevier Science Limited (1998), and include those such as the “split and pool” or “parallel” synthesis techniques, solid-phase and solution-phase techniques, and encoding techniques (see, for example, Czarnik, A.W., Curr. Opin. Chem. Bio., (1997) 1, 60). In addition, a number of small molecule libraries are commercially available. A number of suitable small molecule test compounds are listed
10 in U.S. Patent No. 6,503,713, which is incorporated herein by reference in its entirety

Methods of Treatment

The present invention provides pharmaceutical compositions and methods of treatment of SMA using an aptamer as described herein and a pharmaceutically
15 acceptable carrier or excipient. A pharmaceutically acceptable carrier or excipient is intended to mean any compound used in forming part of the formulation that is intended to act merely as a carrier, *i.e.*, not intended to have biological activity itself. The pharmaceutically acceptable carrier is generally safe, non-toxic and neither biologically nor otherwise undesirable. More than one pharmaceutically acceptable carrier or
20 excipient can be used in a formulation. By “treating” SMA is meant obtaining a desired pharmacological and physiological, e.g., clinical, effect. The effect can be prophylactic in terms of inhibiting or preventing the disease, symptom, or condition thereof and/or can be therapeutic in terms of a partial or complete cure of the disease, condition, symptom, or adverse effect attributed to the disease. An “amount effective to treat SMA” or a
25 “therapeutically effective amount” is an amount that brings about one or more of the effects of treating SMA discussed above.

In addition to the active therapeutic agent, the therapeutic compositions described herein can be formulated to contain suitable pharmaceutically acceptable carriers, e.g., carriers comprising excipients and auxiliaries that facilitate processing of the active
30 compounds into preparations that can be used pharmaceutically for delivery to the site of action. Suitable formulations for parenteral administration include aqueous solutions of the active compounds in water-soluble form, for example, water-soluble salts. In addition, suspensions of the active compounds as appropriate injection suspensions may

be administered. In some embodiments, the injection suspension is an oily suspension; suitable lipophilic solvents or vehicles include fatty oils, for example, sesame oil, or synthetic fatty acid esters, (e.g., ethyl oleate or triglycerides). Aqueous injection suspensions can contain substances that increase the viscosity of the suspension include, for example, sodium carboxymethyl cellulose, sorbitol and/or dextran. Optionally, the suspension can also contain stabilizers. Liposomes and other viral and non-viral vectors can also be used to encapsulate or otherwise prepare the compound for delivery into the cell.

Methods described herein, and methods known in the art, can be used to select a compound with a favorable therapeutic profile, i.e., a compound that has the greatest desirable effect with the least unwanted side effects. For example, therapeutic efficacy and toxicity of such compounds and compositions can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., for determining the LD₅₀ (the dose lethal to 50% of the population) and the ED₅₀ (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and can be expressed as the ratio LD₅₀/ED₅₀. Compounds and compositions that exhibit large therapeutic indices are preferred.

The nucleic acid aptamers described herein can be administered in a therapeutically effective amount by any of the accepted modes of administration for agents that serve similar utilities. The compounds can be formulated for administration by a variety of routes, including, but not limited to, oral, parenteral (e.g., subcutaneous, subdural, intravenous, intramuscular, intrathecal, intraperitoneal, intracerebral, intraarterial, or intralesional routes of administration), topical, localized (e.g., surgical application or surgical suppository), rectal, and pulmonary (e.g., aerosols, inhalation, or powder). The compounds can be infused continuously or by bolus injection. The actual amount of the compound of the subject invention, i.e., the active ingredient, can depend on a number of factors, such as the severity of the disease, the age and relative health of the subject, the potency, the route and form of administration, and other factors.

Having successfully established the cell-based selection of an entire exon, the methods described herein have potential for analysis of other alternatively spliced exons. The results described herein demonstrate 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. The methods described herein offer an attractive alternative

to 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 cell-based selection of terminal exonic sequences, the methods described herein offer a significant advantage.

5 Once a compound (or agent) of interest has been identified, standard principles of medicinal chemistry can be used to produce derivatives of the compound. Derivatives can be screened for improved pharmacological properties, for example, efficacy, pharmaco-kinetics, stability, solubility, and clearance. The moieties responsible for a compound's activity in the assays described above can be delineated by examination of
10 structure-activity relationships (SAR) as is commonly practiced in the art. A person of ordinary skill in pharmaceutical chemistry could modify moieties on a candidate compound or agent and measure the effects of the modification on the efficacy of the compound or agent to thereby produce derivatives with increased potency. For an example, see Nagarajan et al., J. Antibiot. 41:1430-8 (1988). Furthermore, if the
15 biochemical target of the compound (or agent) is known or determined, the structure of the target and the compound can inform the design and optimization of derivatives. Molecular modeling software is commercially available (e.g., Molecular Simulations, Inc.) for this purpose. The derivatives can be evaluated for inhibitory activity, therapeutic activity, and therapeutic efficacy *in vivo* and/or *in vitro*, e.g., using a method described
20 herein.

Methods of Diagnosis and Prognosis

Beyond helping to understand the exon-specific mechanism of alternative splicing, the methods and results described herein can be used in predicting disease
25 severity. As evident from the available expressed sequence tags (ESTs), there are variations in exon 7 sequences among human population. Most SMA patients carry copies of *SMN2*. Understanding alternative splicing of *SMN* genes is critical for considering novel treatment strategies.

The results described herein suggest that people with *SMN1* deletions will have
30 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. Thus, improving exon 7 inclusion in *SMN2* holds the promise for a cure. Our results provide additional avenues in

which potential drugs could be designed to target the 5' ss, leading to the increased inclusion of exon 7 in *SMN2*.

Thus the invention includes methods of diagnosing a subject with SMA by obtaining a sample of genomic DNA from the patient, and determining the presence of one or more sequences that negatively affect splicing of exon 7 of SMN, wherein the presence of a sequence that negatively affects splicing of exon 7 of SMN indicates that the subject has SMA. In some embodiments, the methods further include determining the presence of one or more sequences that positively affect splicing, wherein the presence of a sequence that positively affects (or does effect, e.g., allows wild-type) splicing of exon 7 of SMN indicates that the subject does not have SMA, or has only a mild form of SMA. In some embodiments, the sequences that negatively affect splicing include one or more of Exinct and/or the 3'-Cluster.

Without further description, it is believed that one of ordinary skill in the art can, using the preceding description and the following illustrative examples, make and utilize the compounds and methods of the present invention. The following working examples therefore, specifically point out particular embodiments of the present invention, and are not to be construed as limiting in any way the remainder of the disclosure.

EXAMPLES

The invention is further described in the following examples, which do not limit the scope of the invention described in the claims. First, the materials and methods are described.

Materials and Methods

Plasmid construction: Unless otherwise stated, all minigenes used in this study were constructed in the pCI mammalian expression vector (Promega, Madison, WI), using a strategy described earlier (Singh et al., *Biochem Biophys Res Commun* 315:381-388 (2004)). The minigene splicing cassettes pSMN1 and pSMN2 were the same as described in Lorson et al., *Proc. Natl. Acad. Sci. USA* 96:6307-6311, (1999). Minigene splicing cassettes pNT1 and pNC4 were constructed by deleting a major portion of intron 6 from pSMN1 and pSMN2 respectively. The deletion was made through PCR using oligos 5'Xba (GGTACCCGGGTCTAGACGCGTGTCTTGTGAAACAAAATGC) (SEQ ID NO:237) and 3'Xba (ACGCGTCTAGACCCGGGTACC GAGAGAAGCAAGTAGTAT) (SEQ ID NO:238) as shown in Figure 1A. The amplified product was digested with endonuclease XbaI and ligated. This gave rise to a

truncated intron 6 splicing cassette. The truncated intron 6 retained a total of 227 nucleotides, of which 91 and 105 nucleotides belong to wild type sequences from the 5' and 3' end of intron 6 respectively. A 21-nucleotide spacer was introduced between 91 and 105 nucleotides for amplification and cloning purposes.

Minigene splicing cassettes pSMN1 Δ I6 and pSMN2 Δ I6 were constructed by deleting ~6 kb within intron 6 from pSMN1 and pSMN2, respectively (Singh et al., Biochem Biophys Res Commun 315:381-388 (2004)), allowing detection of partially spliced intermediates along with fully-spliced products. For all experiments performed in this report, pSMN1 Δ I6 and pSMN2 Δ I6 refer to *SMN1* and *SMN2*, respectively.

The splicing pattern of pNT1 and pNC4 was similar to pSMN1 and pSMN2, respectively (Figure 1B). The splicing cassette pBxT2 was derived from pNT1 by two modifications: (1) the *Bsa*X1 site was deleted from the vector backbone; (2) the entire exon 7 was replaced with 27 nucleotides (GGCGCCAGAACTAGTCctccATCCGGA) (SEQ ID NO:239). The lower-case sequences represent the engineered *Bsa*X1 site, which upon digestion with *Bsa*X1 restriction endonuclease (New England Biolabs, MA), removes the entire 27-nucleotide insert. The *Bsa*X1 linearized plasmid (pBxT2) was routinely used for inserting amplified exon 7 from the selection experiment. All mutations were inserted by high fidelity PCR and confirmed by sequencing.

Partial randomization of exon 7 of the *SMN1* gene: The initial pool of oligonucleotides of partially random exon 7 was generated using a 90-mer oligonucleotide (E7Rand) (5'CCTTTATTTTCCTTACAGggttcagacaaaatcaaaaagaa ggaaggtgctcacattccttaaattaag-gaGTAAGTCTGCCAGCATTA3'; SEQ ID NO:240). Lower case letters represent partially random 54-nucleotide-long exon 7, upper case letters represent two stretches of flanking intronic sequences (18 nucleotides each). The randomization (doping) was performed with 70% wild-type and 10% of each of the three non-wild-type nucleotides (Singh et al., J Mol Biol 318:287-303 (2002)). Using high fidelity PCR with Pfx, E7Rand as an upstream primer, PCI-DN (AGCATCACAAATTTACAAATAAA; SEQ ID NO:241) as a downstream primer, and pSMN1 Δ I6 as a template, a 1.1 kb DNA fragment was generated. The 1.1. kb product was isolated by gel purification. This fragment contained a partially random exon 7, the entire intron 7 and the entire exon 8. Approximately 4 μ g ($\sim 4 \times 10^{12}$ unique molecules) of the gel-purified fragment was used as a mega-primer in the second PCR with PCI-UP (TGACATCCACTTTGCCTTT CTCTC; SEQ ID NO:242) 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 the 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 cell-based selection experiments. Sequences of the clones from the initial pool confirmed partial randomization (not shown).

A portion of the ligated mixture was used for transfection of *Escherichia coli* (*E. coli*), whereas the major portion was stored for cell-based selection experiments in the human cervical carcinoma C33a cells. The *E. coli* transfection was used to check the characteristics of the initial pool (pool-0). The sequencing of 42 clones confirmed the partial randomization of pool-0 as shown in Figure 3. Based on sequences of 42 clones from pool-0, the average substitution per molecule was determined as 18.9. That would give a randomization of 35%, which was close to the intended value of 30%. Based on the algorithmic calculations, the 35% randomization of exon 7 (which is 54 nucleotides long) would produce most molecules with substitutions between 16 and 22. However, molecules with lower as well as higher substitutions would also be present in the pool. Confirming this diversity, some clones from the initial pool had substitutions as high as 41 and as low as 9. Certain positions were randomized more often than others. The actual substitution at each position has been indicated in Figure 3.

Cell culture, transfection, and RT-PCR analysis: Human cervical carcinoma (C33a) cells were grown in minimal essential media (BioWhittaker) supplemented with 2 mM glutamine and 10% fetal bovine serum, or grown in Dulbecco's modified Eagle medium (DMEM) supplemented with 10% fetal bovine serum (GIBCO-BRL), 100 units/ml penicillin and 100 mg/ml streptomycin (Mediatech). Transfections were performed using the LipofectAMINE™ 2000 transfection kit (Life Technologies, Inc.). Approximately 1 μg of DNA was used to transfect $\sim 3 \times 10^5$ cells plated 24 hours before transfection. At 24 hours post-transfection, cells were harvested, and total RNA was isolated using TRIzol reagent (Life Technologies, Inc.). Total RNA was reverse-transcribed with Moloney murine leukemia virus reverse transcriptase for 1 hour at 37°C using poly-dT priming.

Radioactive PCR analysis was then performed with different primer sets to identify spliced products. PCR was done for 20 cycles in the presence of 0.2 mM dNTP supplemented with trace amounts of [³²P]dATP. Each cycle consisted of a 45s

denaturation step at 94°C, a 45s annealing step at 56°C, and a 90s extension step at 72°C. Primers used to amplify all spliced and unspliced products were pCI-UP (5'-TGACATCCACTTTGCCTTTCTCT-3' (SEQ ID NO:242) and pCI-DN (5'-AGCATCACAAATTTACAAATAAA-3' (SEQ ID NO:241). As a normalization control, the hypoxanthine phosphoribosyltransferase gene (*HPRT*) was simultaneously amplified by using the primers HPRT-Fw (5'-AAGGAGATGGGAGGCCAT-3') (SEQ ID NO:243) and HPRT-Rev (5'-GTTGAGAGATCATCTCCACCAAT) (SEQ ID NO:244).

Amplification products representing the cell-based 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.

Cell-based SELEX: The cell-based selection experiment started with transfection of C33a cells with the initial pool of splicing cassettes (~2X10¹¹ unique molecules per 4X10⁶ cells) and continued with the selection and amplification steps outlined in Fig. 4: (i) amplification of the spliced products by RT-PCR, (ii) gel isolation of the product with included exon 7, (iii) amplification of exon 7 in a second PCR reaction, (iv) ligation of amplified exon 7 back into a splicing cassette, (v) transfection of C33a cells with the ethanol precipitated ligated products. Steps (i) to (v) comprise 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 in order to maintain pool diversity (Singh et al., 2004, supra; Coulter et al., Mol. Cell. Biol. 17:2143-2150 (1997)). Transfection was done with approximately 10¹² molecules using highly transfectable C33a cells. After transfection, the cells were incubated for 24 hours, followed by RNA isolation and RT-PCR. Exon 7 was further amplified using another set of primers, B1 (TTCATGGTACATGAGTGGCACTCATACTCCCTATTATCAG) (SEQ ID NO:245) and B2 (CATTACGACCGTTGCTCGTGAGGCTTTACTGTGGTGATTT) (SEQ ID NO:246) as shown in Figure 4. Upon digestion with restriction endonuclease *Bsa*X1, gel purified fragments were inserted back into *Bsa*X1 linearized plasmid pBxT2. Again, ligation mixture was used for transfection. Four cycles of selection were

performed until exon 7 inclusion efficiency increased more than 100-fold as seen in Table 2 and Figure 5.

Ligation was performed in a 100- μ l reaction containing ~ 1 μ g of *Bsa*XI-digested pBxT2 and ~ 50 ng of enriched exon 7 fragment. ~ 1 μ g of ligated plasmid ($\sim 2 \times 10^{11}$ molecules) was used to transfect two 60 mm plates of C33a cells ($\sim 2 \times 10^6$ cells per plate) using standard calcium phosphate coprecipitation procedure. About 20 hours post-transfection, cells were harvested, total RNA was isolated and the spliced intermediates amplified by RT-PCR essentially as described earlier except that ~ 3 μ g of total RNA was used per 20 μ l 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 *Bsa*XI and inserted back into the *Bsa*XI-linearized pBxT2. At each round, a fraction of ligation mixture was transformed into *E.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.

Transfection was done with approximately 10^{12} molecules using highly transfectable C33a cells. After transfection, the cells were incubated for 24 hours, followed by RNA isolation and RT-PCR. Exon 7 was further amplified using another set of primers, B1 (TTCATGGTACATGAGTGGCACTCATACTCCCTATTATCAG) (SEQ ID NO:245) and B2 (CATTACGACCGTTGCTCGTGAGGCTTTACTGTGGTGATTT) (SEQ ID NO:246) as shown in Figure 4. Upon digestion with restriction endonuclease *Bsa*XI, gel purified fragments were inserted back into *Bsa*XI linearized plasmid pBxT2. Again, ligation mixture was used for transfection. Four cycles of selection were performed until exon 7 inclusion efficiency increased more than 100-fold as seen in Table 2 and Figure 5.

TABLE 2

	% Full	% Skipped	Inclusion Potential	Exclusion Potential	Inclusion Efficiency
	[A]	[B]	[A/B]	[B/A]	[(A/B)/(B/A)]
Pool-0	48	52	0.92	1.08	0.85 ± 0.02
Pool-1	86	14	6.14	0.16	37.7 ± 11.2
Pool-2	91	9	10.1	0.09	102.2 ± 21.5
Pool-3	96	7	13.2	0.08	176.5 ± 26.7
Pool-4	95	5	19.0	0.05	361.0 ± 55.4

NT1	83	17	4.88	0.20	23.8 ± 9.2
NC4	22	78	0.28	3.54	0.08 ± 0.02

Pool-0 represents the population of sequences containing the initial random pool before selection. Numbers 1-4 represent selection cycles that the population of molecules has undergone. Inclusion efficiency was defined as the number of transcripts including exon 7 for every transcript skipping exon 7. The pool strength was determined simultaneously using similar amounts of *E. coli* purified plasmids derived from each pool. The values of the control experiments using similar amounts of NT1 and NC4 plasmids are also given. Standard deviation was calculated for triplicate experiments. The exon 7 inclusion efficiency of pool-4 is about 15 times higher than NT1. Supporting this observation, none of the clones from pool-4 had a detectable amount of exon 7 exclusion product as shown in Figure 5.

Analysis of Sequences: A statistically significant number of clones was analyzed from the final round of cell-based selection. To determine the position-specific significance of nucleotides, sequences were forced aligned from one end to the other (Bartel et al., Cell 67:529-536 (1991); Singh et al., J Mol Biol 318:287-303 (2002)).

These sequences are depicted in Figure 6A-C. These sequences contain most, if not all, information necessary for exon 7 inclusion. Representing the characteristics of the initial pool, most sequences maintained the length of exon 7 as 54 nucleotides. Generally, the presence of cryptic splice sites would have reduced the length of the exon. Lack or rarity of such incidence supports the utility and efficacy of the design of the cell-based selection analysis of the entire exon. Complete randomization can not have resulted in maintaining the size of the exon.

The “mutability values” shown in Fig. 6D were calculated by comparing the ratios (R) of mutant (mut) to wild-type (wt) nucleotides of exon 7 in selected (pool-4) and initial pool (pool-0), using the equation $[(R_{(mut/wt)pool-4})/(R_{(mut/wt)pool-0})]-1$ (Deminoff et al., Genetics 141:1263-1274 (1995)). Motif analyses were done using MacVector software. Covariation analysis was done manually.

Example 1: Determination of the Critical Features of Exon 7 of the *Snn1* Gene Responsible for Alternative Splicing Using Cell-based Selection

In order to perform a comprehensive and context-specific analysis of exonic nucleotides, an approach that is independent of a more generalized approach based on

score matrices (Liu et al., *Nat. Genet.* 27:55 (2001)) and computational analysis (Fairbrother et al., *Science* 297:1007-1013 (2002)) was adopted. A powerful technique of functional cell-based SELEX was developed, which permitted the examination of the mutability of every nucleotide within exon 7. The results support the hypothesis that the single nucleotide substitution, which improves the 5'-splice site (5'-ss), eliminates the requirement of ASF/SF2 as well as hTra2- β 1-associated-ESEs for exon 7 inclusion. Similarly, the presence of a unique ESS that obviates the requirement of ASF/SF2-associated ESE was demonstrated.

Functional SELEX is a powerful method to analyze exonic sequences that modulate alternative exon inclusion during pre-mRNA splicing in living cells (Cooper, *Meth. Mol. Biol.* 118:405-417 (1999)). However, due to technical difficulty in separation and amplification, the cell-based SELEX of an entire exon has not been reported. A carefully designed and optimized protocol, described herein, enabled the examination of the entire 54 nucleotide length of exon 7 of the *SMN1* gene.

In order to minimize the selection of cryptic splice sites and non wild type enhancer elements, partially random ("doped") sequences were used containing 70% wild-type and 10% each of the three non-wild-type nucleotides at each position of exon 7. This approach generates an initial pool of sequences that harbors a range of substitutions with the maximum being close to the 30%. Sequencing of 42 clones from the initial pool showed that the average extent of randomization was 35%, close to the desired value, with the extent of randomization at each position varying from 15% to 55% (data not shown). Assuming 35% randomization, most (96%) of the RNAs in the pool would have 9 to 29% substitutions, and variants with 19 substitutions would be maximally represented ($\sim 10^{11}$ molecules; about 10% of the initial pool). This degree of complexity allows definition of conserved sequences and structures in the context of the natural binding sites for splicing factors.

The pool was subjected to four rounds of selection using 10^{12} molecules in a cell-based splicing assay, as shown in Figure 4A. At each round, the exon 7-included products were amplified such that the adjoining exonic sequences were replaced by intronic sequences (Figure 4B). This allowed reconstruction of the splicing cassette with enriched exon 7 sequences for the next round of selection. For each round, the exon inclusion efficiency was determined as shown in Table 2. The final pool had more than 300-fold enriched splicing efficiency compared to the initial pool. At each cycle, the pool

diversity and splice junctions were determined by analyzing sequences of randomly selected 30 or more clones generated from exon 7-included products.

Results: The sequences of unique clones selected from pool 4 are shown in Figure 6A-C. All of these clones had maintained the correct splice sites preserving the size of exon 7 as 54 nucleotides. Reflecting the characteristics of the original pool, most selected sequences had substitutions between 9 and 28. The splicing pattern of all clones derived from pool 4 showed exon 7 inclusion efficiency better than the *SMN1* gene; 3 representative clones have been shown in Figure 5.

Comparison of 113 selected sequences from primary SELEX showed that exon 7 variants contain highly to moderately conserved residues. For example, positions 1, 2, and 53 were mostly conserved for the wild-type G residues. Other moderately conserved positions that preferred wild-type residues were 8, 20, 38, 39, 42, 43, 44, and 52. Supporting the principles of unigenic evolution (Deminoff et al., *Genetics* 141:1263-1274 (1995)), the inhibitory nature of wild-type residues at positions 3, 4, 5, and 54 was apparent by frequent substitution. Most noticeable among them was position 54 that overwhelmingly selected a non-wild-type guanosine residue (A54G substitution). The frequency of G residues at terminal positions exceeded the statistical probability (Burge et al., pp. 525-560 in *The RNA World*, Cold Spring Laboratory Press, Cold Spring Harbor, NY (1999)) of such occurrence suggesting a dominating role of context in overall selection. The prominent role of context is further strengthened by retention of wild-type G residues at positions 2 and 53. Statistically, all residues represent the 2nd position of internal exons (Burge et al., pp. 525-560 in *The RNA World*, Cold Spring Laboratory Press, Cold Spring Harbor, NY (1999)), although uridine has been shown to be the preferred residue for interaction with U2AF65^{65/35} heterodimer (Wu et al., *Nature* 402:832-835 (1999)). Similarly, the penultimate position of internal exons (position 53 in case of exon 7) of human U2-type introns mostly contain adenosine residues (Burge et al., pp. 525-560 in *The RNA World*, Cold Spring Laboratory Press, Cold Spring Harbor, NY (1999)), however, guanosine residues at this position will also base pair with a uridine residue of U1 snRNA, the RNA part of the U1 snRNP.

The pattern of selected bases clearly shows the conserved nature of hTra2- β 1-associated ESE as purine-rich region extending from position 19 to 27. Compared to an expected 3 substitutions per molecule in hTra2- β 1-associated ESE, more than 60% clones showed less than 2 substitutions. Positions 25 and 26 were more conserved than the

others. Though all residues were tolerated at each position in this region, certain residues were preferred. For example, A24U was more common than A27U substitution.

Pyrimidines were less preferred at position 26 representing less than 5%. Contiguous uridine substitutions G25U and G26U that have been previously shown to abrogate hTra2- β 1-associated-ESE (Hofmann et al., *Proc. Natl. Acad. Sci. USA* 97:9618-23 (2000)) were not selected. The other relatively conserved positions in the vicinity of hTra2- β 1-associated ESE were 29 and 30, where G residues were mostly selected. The significance of positions 29 and 30 in exon 7 is unknown, though these positions can be critical for recruitment of other transacting factors that interact with hTra2- β 1.

The ASF/SF2-associated ESE that extends from positions 6 to 12 (Cartegni and Krainer, *Nat. Genet.* 30:377-384 (2002)) was found to be only partially conserved. For example, we observed A11G substitution restoring the inhibitory effect of C6T substitution, however position 11 did not select G as a preferred residue that would have created a high score motif (Cartegni and Krainer, *Nat. Genet.* 30:377-384 (2002)) for ASF/SF2-associated ESE. Similarly, G8A (8%), C10A (12%), C10U (10%), and A12U (10%) substitutions that are inhibitory for ASF/SF2-associated ESE were selected. Overall, the nature of the selected bases did not reflect the characteristics of the typical ESE associated with ASF/SF2. This can suggest that motifs associated with recognition of ASF/SF2 are different in different exonic contexts. Alternatively, it is possible that the frequently disposable nature of ASF/SF2 is easily compensated by exonic substitutions outside the ASF/SF2-associated ESE. Supporting this observation, positions 3, 4, and 5 that immediately precede ASF/SF2-associated ESE contain the highest density of triple substitutions as more than 20% selected variants contain these substitutions. This level of substitution is approximately five times higher than the expected value obtainable from the initial pool. This suggests that uridine residues at positions 3, 4, and 5 are highly inhibitory, which can constitute an ESS that becomes active when ASF/SF2-associated ESE is abrogated.

Example 2: Identification of Mutations Supporting the Results of SELEX

To evaluate the complex outcome of cell-based selection of an entire exon, several single, double, and multiple substitutions within exon 7 were made.

First, the conserved nature of the G residue at position 1 was confirmed by G1A, G1C and G1U substitutions. All of these substitutions led to exon 7 exclusion, suggesting

the inhibitory nature of non-wild-type residues at position 1. The inhibitory effect of A1C mutation selected in one clone (in clone P04-72) seems to be compensated for by other mutations. Significantly, a single nucleotide substitution at position 16 compensated for the inhibitory effects of G1C substitution. Such substitution was also selected in clone P04-72. The inhibitory effects of substitutions at position 1 was also compensated for by an A54G substitution. Supporting the results of cell-based selection, the uridine was not tolerated at position 20. Despite the overwhelming selection of G residues at position 2 and 53, all residues were tolerated at position 2 (not shown) and A and C were tolerated at position 53. However, this was not surprising given the fact that the partial randomization tends to preserve the wild-type nature of the nucleotides that is located next to a predominantly selected residue. The preference for G residues at the 5'-splice site is well-documented, however, in the case of exon 7, A54G substitution seems to play a dominant role. Confirming this observation, the A54G substitution in the *SMN2* gene was found to favor exon 7 inclusion. This clearly suggests that the role of ASF/SF2-ESE is not critical for exon 7 inclusion in variants with an improved 5'-splice site. This argument is supported by the poorly conserved nature of ASF/SF2-ESE. To further evaluate the impact of the A54G substitution vis-a-vis the role of hTra2- β 1-ESE, a G25U-G26U substitution in the *SMN2* gene was made. This substitution is known to disrupt hTra2- β 1-ESE, which is absolutely required for exon 7 inclusion (Hofmann et al., *Proc. Natl. Acad. Sci. USA* 97:9618-9623 (2000)). Surprisingly, the A54G substitution in the *SMN2* gene was able to override the inhibitory effects of G25U-G26U substitution, suggesting that the A54G substitution is dominant enough to obviate the requirement of well-established ASF/SF2-ESE and hTra2- β 1-ESE. These results also validate the selection of a large proportion of clones with more than expected substitutions, because most of these clones had A54G substitutions.

To verify the nature of an ESS at the 3'-ss, substitutions were made at positions 3, 4, and 5 in the *SMN2* construct. Single and double substitutions had no effects, however many triple substitutions restored exon 7 inclusion. Supporting these results, deletion of three uridines also restored exon 7 inclusion. To further characterize the nature of the ESS, we performed deletion and addition mutations within *SMN1* construct. Deletion of one nucleotide from the 5' end of exon 7 was tolerated, though double and triple deletions led to exon 7 exclusion. These results support the results of cell-based SELEX, which did not select the uridine residue at the first position of the exon. Surprisingly, addition of a

single G in the *SMN1* construct led to exon 7 exclusion, despite the presence of the wild-type ASF/SF2-associated ESE. It is possible that addition of one nucleotide creates a high affinity ESS motif that dominates over existing ESE motif associated with ASF/SF2.

The single most conserved position in the middle of exon 7 identified by cell-based SELEX was 20, which did not tolerate any uridine residue substitution. Supporting this observation, A20U substitution in the *SMN1* construct led to exon 7 exclusion in about 50% transcripts. Interestingly, a previous study that incorporated triple uridine substitutions including positions A19U, A20U, and A21U did not observe this effect (Lorson and Androphy, *Hum. Mol. Genet.* 9:259-265 (2000)). This can suggest that the balanced effect of multiple contiguous substitutions has a combinatorial effect. Supporting this argument, the cell-based SELEX did select uridine residues as the preferred substitution at position 21. The positive effect of A21U substitution can have compensated for the negative effect of A20U giving rise to the null effect in the triple substitution at SE2a, an ESE splicing sequence (Lorson and Androphy, *Hum. Mol. Genet.* 9:259-265 (2000)).

The selection of residues based on combinatorial principles using cell-based SELEX is guided by many factors. Though it was possible to extract the role of many positions, the overwhelming selection of A54G substitution can have undermined the role of other positions. Also, certain residues can have been selected due to preference for a particular RNA structure. Certain substitutions that created nonsense mutations can have been differently affected through an altered mechanism of RNA splicing and/or stability. Finally, the selection of the final pool is dependent upon the nature of the initial pool, as there was no saturation of sequence space. The significance of all of these factors is not easily discernible. Supporting the above argument, single or multiple substitutions at positions 43, 44, and 45 had no effect on exon 7 inclusion, although those positions mostly selected the wild-type residues.

Mutability of Exonic Nucleotides Indicates Position-Specific Roles

The results of the cell-based selection experiments described herein can be explained by the mutability of exonic sequences. A highly mutable position is considered as inhibitory for exon 7 inclusion, whereas the least mutable position is considered as stimulatory (Deminoff et al., *Genetics* 141:1263-1274 (1995); Guo et al., *Science* 289:452-457 (2000)). Mutability of wild-type residues of exon 7 were calculated using an equation that corrects selection results for the bias of the initial pool (Fig. 6D;

Deminoff et al., 1995, supra). 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 due to 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 mutable. 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 cell-based 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. 6D). 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 towards 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.

Example 3: Secondary Cell-based SELEX of the First 52 Nucleotides of Exon 7

In order to evaluate the conserved nature of residues in the presence of a weak wild-type 5'-ss, secondary cell-based SELEX of the first 52 nucleotides of exon 7 was performed. The strategy of randomization and cloning was the same as that used for the cell-based SELEX of entire exon 7. 70% doped wild-type nucleotides were used at each position. Three rounds of selection were performed that gave rise to an exon 7 inclusion

efficiency of more than 100-fold.

The sequences of numerous randomly selected clones are described in Figure 6A-C. The exon 7 inclusion efficiency of all clones was better than the wild-type *SMN1* gene. Noticeably, the over all pattern of selected sequences were only moderately affected by the absence of a strong 5'-ss. For example, the poly-purine nature of hTra2- β 1-associated ESE was mostly conserved, while a strong ASF/SF2 associated ESE was not selected. Surprisingly, more clones with C6U or C6A substitutions that are known to abrogate ASF/SF2-associated ESE were observed. As a matter of fact, there was no preference for C residue at position 6 because all other residues were equally tolerated. In addition, the C6U or C6A substitutions were mostly not accompanied by compensatory A11G mutations that would have restored the ASF/SF2-associated ESE. Interestingly, more substitutions immediately preceding and following the ASF/SF2-associated ESE were observed suggesting the presence of a 3' ESS, inhibitory secondary structure, or both.

Example 4: Role of RNA Secondary Structure as Supported by Cell-Based SELEX

The SELEX of larger sequences using partially random oligonucleotides has been extremely successful in predicting the critical information related to RNA secondary and higher order structures. In order to extract such information from the SELEX studies, the secondary RNA structure of exon 7 in the context of adjoining introns was partially probed. The use of ribonuclease T1, chemical agent DMS, and ribonuclease V1 identified the unpaired guanosines, unpaired adenosines, and all paired bases respectively. The probed secondary structure as shown in Figure 8 was also supported by highly favorable secondary structure predicted by mfold analysis (Mathews et al., *J. Mol. Biol.* 288:911-940 (1999)). The frequently mutable residues as determined by cell-based SELEX have been indicated on the secondary structure.

The local RNA secondary structure of exon 7 consists of two terminal stem loops (TSL1 and TSL2), three internal loops (IL1, IL2, and IL3) and one internal stem (IS1). The TSL1 is exclusively formed by exonic sequences and harbors the poorly conserved ASF/SF2-ESE that is mostly located in the terminal loop (TL1). The TSL1 also overlaps the HMT1, the longest tract of sequence prone to substitution. The types of substitutions within TSL1 clearly indicate the inhibitory nature of the terminal stem (ST1). The internal loop (IL1) connecting TSL1 with TSL2 is formed by intron 6 sequences. The

intron 6 and exon 7 boundary forms the internal stem IS1 by extended base pairing of 10 nucleotides. The IS1 harbors the relatively conserved part of hTra2- β 1. The substitutions in the IS1 seem to maintain the stem through canonical or non-canonical base pairing. The TSL2 is mostly formed by exon 7 sequences through the first two bases of intron 7, which are also buried into TSL2. The TSL2 also contains the sequences of HMT2, though most of the mutable residues are located in the loop. The mutation at position 45 was frequently compensated for by mutations at position 49, restoring the base-pairing. Other mutations in the upper part of TSL2 tend to maintain the stem, however, the mutations in the lower part of the stem can form alternative RNA-structures. The mfold analysis predicted the formation of a coaxial stack following A54G substitution. The role of coaxial stacking in the splicing of pre-mRNA is not known, although it is predominantly present as higher order RNA structures in tRNAs (Moore, *Annu. Rev. Biochem.* 68:287-300 (1999)). Coaxial stacking in equilibrium with the protracted stem has been shown to regulate the self-splicing of ribosomal Group I intron (Cao and Woodson, *RNA* 4:901-914 (1998)). In the case of exon 7, the coaxial stacking will make the 5' splice site available by exposing the terminal G residue for interaction with U1C protein.

To further elucidate the role of TSL1 in exon 7 inclusion, a number of mutations were made that altered the characteristics of TS1 or TL1. All mutations in 3'TS1 disrupting the upper part of TS1 restored the inhibitory effects of C6U substitution. Several double substitutions in 5'TS1 disrupting the upper part of TS1 also restored the inhibitory effects of C6U substitution albeit with variable degrees. However, triple G substitutions in 5'TS1 disrupting the TS1 had adverse effects. This could be due to an alternative secondary structure formed between G-rich exonic sequences and the adjacent polypyrimidine tract. Such structure will bury the 3'ss into a protracted stem (not shown). Mutations towards the lower part of TS1 had variable consequences. Interestingly, several critical positions were identified in TSL1 that play an important role in exon 7 inclusion. For example, A11C and C16G substitutions were found to restore the inhibitory effect of C6U substitution. These results are significant as they represent substitutions not associated with ASF/SF2-ESE. The role of position 16 merits special mention as C16A and C16U substitutions had adverse effects on exon 7 inclusion despite the presence of ASF/SF2-ESE. The compensatory role of C16G substitution that overcomes the inhibitory effects of G1C, G1A, and C6U substitutions suggests the tight

cooperation among *cis*-elements spanning beyond the usual size of ESE elements that are usually 6 nucleotides long. (Fairbrother et al., *Science* 297:1007-1013 (2002)). These results also support a hypothesis that ESE containing sequences depend upon structural context for effectiveness.

5

Example 5: Tertiary Cell-Based SELEX of the First 6 Nucleotides of Exon 7

To further evaluate the role of the first six nucleotides of exon 7, cell-based SELEX was performed using a completely random pool containing all 4096 possible combinations of 6 nucleotides. The method to introduce random sequences at the first 6
10 positions of exon 7 and the selection protocols were similar to the one used for primary cell-based SELEX. The randomization at every position was confirmed by sequencing 25 clones from pool-0, the initial pool (not shown). Three rounds of selection were performed that gave rise to an enriched pool with exon 7 inclusion efficiency better than the *SMN1* gene (data not shown). A compilation of sequences of 49 randomly selected
15 clones and splicing pattern of 12 representative clones from the final pool has been given in Figure 6. In agreement with the results of primary and secondary cell-based SELEX, only G residues were selected at position 1. This also confirms that 3/4 of the total population of the initial pool (3070 variants per 4096 variants) containing substitutions at position 1 were not selected. This is a remarkable observation in light of the C16G
20 substitution that compensates for the inhibitory effects of substitutions at position 1. The critical role of the G residue at position 1 in exon 7 of the *SMN1* gene in particular and possibly many internal human exons in general seems to be associated with the lack of a strong *cis*-element within the exon.

The nature of selected residues showed a preference for C and G residues at
25 position 6, supporting the role of ASF/SF2. However, in about 1/4 of cases, A and U residues were selected, disrupting the ASF/SF2-ESE and suggesting the inhibitory role of other residues within the first six nucleotides. Two or more substitutions that removed inhibitory residues tolerated C6U and C6A substitutions. Most of these substitutions seem to break ST1, however, not all such substitutions were preferred. For example,
30 simultaneous G substitutions at positions 3, 4, and 5 were not selected due to possible formation of an alternative secondary structure as discussed earlier.

Example 6: Deletion and Addition Mutations Support the Finding of Cell-based Selex

For aptamer-mediated splicing of exon 7, it is important to know the relative positioning of ESE elements within exon 7 and also with respect to 5'-ss and 3'-ss. To this end, deletion and addition mutations from the 5' and 3' end of exon 7 were performed.

5 Deletion of one nucleotide from the 5' end, which will maintain at least one guanosine residue at the 5' end of exon 7, had no adverse effect on splicing. However, deletions of 2 and 3 nucleotides from the 5' end of exon 7 had dire consequences. A G1T substitution, which is not a preferred nucleotide at position 1, can be created by deletion of 2 and 3 nucleotides. In contrast to deletions from the 5' end of exon 7, deletions of up to 3

10 nucleotides from the 3' end of exon 7 had positive effects on exon 7 inclusion. Supporting the results of SELEX, these deletions would create substitutions, such as either A-1G and/or A-4T (negative numberings are from the 3' end of exon 7). It is also possible that deletion of 3 nucleotides from the 3' end of exon 7 will bring the 5'-ss (of intron 7) closer to the 3'-ss (of intron 6), which will optimize the distance for U1 snRNA-mediated splicing (Hwang and Cohen, *Mol. Cell. Biol.* 17:7099-7107 (1997)) of upstream

15 intron 6. Serial additions of nucleotides (such as G, GG, GGT, GGTT, GGTTT, GGTTTC, etc.) at the intron 6/exon 7 junction, duplicated a portion of the exon 7 sequences at the 5' end of exon 7. Insertion of 1, 2, and 3 nucleotides eliminated exon 7 inclusion. The

20 element was fully restored by the addition of six nucleotides (GGTTTC). These additions will not abrogate any known ESEs, though the additions move them away from the 3'-ss (of intron 6). Some of these additions will also create the inhibitory C6T substitution.

Based on the above results, the 5' and 3' region of exon 7 should be accessible to spliceosomal machinery. Thus, the design of annealing sequences of the aptamers should

25 avoid these regions. Results of SELEX and other mutations suggest that the best place for aptamer annealing is in the middle of exon 7 (Fig. 8). This can also suppress the inhibitory portions of exon 7, such as that between positions 19 and 24 and 31 to 36. Additionally, the region between 29 to 38 nucleotides forms a 10-nucleotide loop (LII) that should support the annealing of antisense sequences that are added to the aptamers.

Example 7: RNA Secondary Structure of Exon 7

For aptamer-mediated splicing of *SMN2* gene, it is important to find the accessible bases (generally in the loops) that could bind with aptamer-associated antisense

sequences. RNA-secondary structure is considered to play a vital role in spliceosome mediated pre-mRNA splicing (Balvay et al, *Bioessays* 15:165-169 (1993)). In order to determine the secondary structure of exon 7 RNA, chemical and enzymatic methods were used. Unpaired adenosines and guanosines were probed by DMS and ribonuclease T1 respectively. The double-stranded region was probed by ribonuclease V1. Probing was done by treatment of the gel-purified transcripts (T7 transcripts derived from NotI-linearized pNT1 and pNC4 cassettes) with the appropriate reagents followed by primer extension using standard protocols (Ehresmann et al., *Nucl. Acids Res.* 15:9109-9128 (1987)).

Structure probing predicts three stem-loops (SL) within exon 7, designated as SL1, SLII, and SLIII as shown in Figure 8. The probed secondary structure was also supported by the results of SELEX. For example, G29T, A50T, and A51T substitutions will strengthen the stems. Similarly, a G1A substitution, which is inhibitory, will probably disrupt SLI and make the guanosine residue in the CAG motif (at the 3'-ss of intron 6) paired. Upon superimposing the conserved bases over the secondary structure of exon 7, compensatory mutations that maintained the base pairing of stems were found. However, several mismatch base pairs (Leontis et al., *Nucl. Acids. Res.* 30:3497-3531 (2002)) were also found. A few mismatch base pairs that will produce a bulge in the stem will likely be tolerated.

Role of secondary contacts and RNA structure:

Position 8 (8G) is 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 binds 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/or hnRNP G make such contacts to 8G in order to recruit U2AF³⁵ at the 3' ss. The role of a 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 (Fig. 8 and Singh et al., 2004, supra). 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., EMBO J 20:7259-

7270 (2001)). Recently, it has been implicated for SR-proteins that show specificity to RNA conformations (Buratti et al., Mol Cell Biol 24:1387-400 (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, Nature 419:86-90 (2002)). The above observations support a common theme in which distance and distribution of sequences that flank critical motifs are equally important.

While many of the structural elements can not have any significance, few can have been eliminated owing to their inhibitory nature. Supporting this argument, mutations that disrupted a stem-loop near the 3'-ss of exon 7 resulted in a substantial increase in exon 7 inclusion. Though yet to be confirmed, the pronounced effect of A54G substitution could be partly due to an alteration of a secondary structure at the 5'-ss of exon 7. Zuker's *mfold* analysis predicted a complete transformation of RNA-secondary structure as a consequence of this substitution. Single nucleotide substitutions, which are predicted to disrupt a stem-loop structure at the 5'-ss of exon 10 of the human *tau* gene, have been shown to increase the exon 10 inclusion. Probably due to similar reasons, when exon 7 of the *SMN1* gene was substituted with exon 10 of the *tau* gene, exon 10 skipping was not observed (data not shown). However, this observation underscores the fact that the role of terminal G in U1 snRNA base pairing can not be as critical for inclusion of *tau* exon 10 as in the case of exon 7 of *SMN* genes. Thus, it is presumed that the impact of A54G substitution in exon 7 is unique and guided by many factors including size, sequence and structure of exon 7. The present results support the idea that restoration of U1 snRNA base-pairing with the 3'-terminal exon-7 sequences promotes exon-7 inclusion. This result has therapeutic implications because engineering U1 snRNA can be used to improve exon 7 inclusion by the *SMN2* gene in SMA patients. While the role of RNA secondary structure on exon 7 inclusion is a matter of future study, the lack of a strong 5'-ss in exon 7 in both *SMN* genes seems to be more due to deliberate design of nature than merely an accident. This is particularly true for position 54, which will have no consequence on the translated protein.

Example 8: Aptamer-Driven Inclusion of Exon 7 Through *Cis*-Splicing

Small RNA molecules corresponding to selected exon 7 aptamers are generated. Various post-selection modifications can be introduced for *in vivo* stability and delivery,

such as 2'-O-methyl, phosphorothioate, and 5'-polyethylene glycol backbone modifications. Other modifications will be readily apparent to one skilled in the art. The diagrammatic representation of typical aptamer-driven inclusion of exon 7 is shown in Figure 7A.

5 The success of aptamer-mediated *cis*-splicing will depend on delivery of aptamers to the intron-exon boundary. One way to ensure such delivery is to attach binding sequences to the aptamers. Binding sequences in this case will correspond to the RNA sequences complementary to a portion of exon 7. Traditionally, the size of annealing sequences used for antisense targeting has been in the range of 15 to 25 nucleotides
10 (Agrawal and Kandimala, *Mol. Med. Today* 6:72-81 (2000)). The information derived from the structure probing of exon 7 will assist in locating optimal binding sites within exon 7. The preferred binding sequence should be located in the extended loop so that sequences are easily accessible for pairing. Such structure has already been located within exon 7. It extends from bases 29 to 38 of exon 7 as shown in Figure 8.
15 Additionally, the choice of this site is also better for suppressing an inhibitory element predicted by cell-based SELEX.

 The chosen aptamers (barring antisense part of the molecule) should be able to form a rigid secondary structure so that cellular nucleic acids should have minimum or no chance of pairing. A number of such aptamers have been identified such as the one
20 shown in Figure 10. The use of 2'-O-methyl, phosphorothioate, and 5'-PEG backbone modifications will increase the *in vivo* stability of RNA and transfer into the cell. Such oligos can be synthesized commercially. Because of the limits on the length of synthetic RNA, the final size of RNA molecules used for exon 7 targeting will vary from about 40 to about 60 nucleotides.

25 The aptamers containing antisense binding sequences can be transfected into C33a cells. Stable cell lines expressing either *SMN1* or *SMN2* minigene can be generated. Virtually any eukaryotic cell line can be used for transfection, however, human cell lines are preferred. Neuronal cell lines, such as NSC34, can also be used. Spliced products are verified by RT-PCR followed by sequencing. For quantification, real time PCR can be
30 used. The *in vivo* stability of aptamers can be verified by RNase protection assays. Positive results can be verified in neuronal cell lines and in cells derived from SMA patients. If SMA patient cells are used, increased levels of SMN protein can be monitored with anti-SMN antibodies. If SMN protein levels are increased, transformed

patient cells can be used to monitor the localization and reorganization of nuclear bodies associated with SMN. Changes in GEMS (nuclear bodies which are markers for SMN protein), Cajal bodies, and other nuclear structures (Wang et al., *Proc. Natl. Acad. Sci. USA* 99(21):13585-13588 (2002)) can be monitored with immunofluorescence techniques. *In vitro* splicing assays can also be used to confirm the effect of the aptamers on splicing. Splicing competent nuclear extract (Dignam et al., *Nucl. Acids. Res.* 11:1475-1489 (1983)) and fractionated S100 extract (Mayeda and Krainer, *Methods in Mol. Biol.* 118:309-314 (1999)) can be used for *in vitro* splicing assays.

In addition to using synthetic ribo-oligos, which can be cost-prohibitive, cytoplasmic expression of aptamers, or "intramers," can be used. Intramers have recently been acknowledged for their striking potential as rapidly generated intracellular therapeutic ligands (Famulok et al., *Chem. Biol.* 8:931-939 (2001)). In the present case, expression cassettes based on the human tRNA^{met} and U6 snRNA promoters (Good et al., *Gene Ther.* 4:45-54 (1997)), in which transcripts encoding small DNA inserts are protected against RNase attack from the 3'-end, can be generated to establish a vector system that guarantees a high concentration of aptamers in human cell nuclei.

Example 9: Inclusion of Exon 7 through Trans-Splicing

The *trans*-splicing method of SMaRT can be used to increase production of functional SMN protein from the SMN2 gene. The diagrammatic representation of the typical mechanism for SMaRT is shown in Figure 7B. SMaRT represents a general approach for reprogramming the sequence of targeted transcripts. In this approach, the pre-*trans*-splicing molecule (PTM) includes a 5' region that matches, in antisense orientation, the 3'-splice acceptor region of the target gene. The *trans*-splicing reaction is thought to occur after hybridization of the guide sequence of the PTM with the target precursor RNA, by the natural machinery (the spliceosome). In the hybridized region, the 3' splice acceptor of the target is usually masked and it is therefore assumed that the spliceosome will connect the 5' donor of the target with the alternative acceptor, thus generating a chimeric mRNA whose region beyond the *trans*-splice point can be designed at will for therapeutic or experimental purposes. One of the advantages of SMaRT over other molecular medicine strategies is the tight regulation of the spliced product, which is dependent upon endogenous levels of pre-mRNA.

PTMs are designed to target intron 6 of the *SMN2* gene. Such targeting will allow

ligation of exon 6 of the endogenous SMN2 pre-mRNA to the engineered exon 7 of the PTM. Masking of the natural 3'-ss seems to be important for effective *trans*-splicing. (Puttaraju et al., *Mol. Ther.* 4:101-114 (2001)). However, the defective 3'-ss of intron 6 in the present case does not require such masking. Accordingly, *trans*-splicing was achieved without any artificial guide sequences (Figure 9). Thus, there is flexibility in designing the guide sequences for annealing to a very small portion (20-30 nucleotides) of the 6 kb long endogenous intron 6 of SMN2 pre-mRNA. Figure 9 depicts the *trans*-splicing cassettes pEx6Int6 (5'-leader or 5'-L) and pInt6Ex7 (PTM). The 5'-L contains the entire exon 6 sequence followed by the first 100 nucleotides of intron 6. The PTM contains the last 100 nucleotides of intron 6 following by the entire exon 7. The pre-mRNA derived from these two constructs should anneal through natural base pairing of 10 contiguous bases.

When C33a cells were co-transfected with the above two constructs, about 5% of products was found to be *trans*-spliced compared to *cis*-spliced. Similarly, when PTM was transfected with pNC4 constructs, the *trans*-spliced products were obtained (Figure 9). Here again, the *trans*-splicing efficiency was about 5%. This level of *trans*-splicing was achieved through natural base pairing of about ten contiguous nucleotides. Careful planning will likely produce a higher level of *trans*-splicing. Such planning includes reinforcement of extended base pairing (Puttaraju et al., *Nat. Biotechnol.* 17:246-252 (1999); Puttaraju et al., *Mol. Ther.* 4:105-114 (2001)), use of secondary structure that mimics the 3'-ss (Puttaraju et al., *Mol. Ther.* 4:105-114 (2001)), and the use of aptamer sequences such as those depicted in Figures 3 and 6.

The data from cell-based SELEX experiments are extremely useful in designing PTMs. Wild-type exon 7 sequences can be substituted with aptamer sequences. The amino acid sequence of the SMN protein should be maintained will making such substitutions. The substituted sequences have an optimal combination of ESE elements. In other systems, ESE sequences are known to improve the chances of *trans*-splicing. The success of *trans*-splicing will depend upon annealing of PTM to a target binding site in intron 6 of the SMN2 transcript. A preferred binding site within intron 6 should not possess strong RNA secondary structure. Such binding site(s) can be identified by Zuker's mfold analysis (Zuker and Jacobson, *RNA* 4:669-679 (1998)). Where appropriate, RNA secondary structure can be inserted near the 3'-ss in PTM. Such structure mimics the wild-type *cis*-element at the 3'-ss. (Puttaraju et al., *Nat. Biotechnol.* 17:246-252

(1999)).

The construction of PTMs can be done in pCI vector (Promega, Madison, WI). Highly transfectable C33a cells, neuronal NSC34 cells, and fibroblast cells derived from SMA patients (Andreassi et al., *Hum. Mol. Genet.* 10:2841-2849 (2001); DiDonato et al.,
5 *Hum. Gene Ther.* 14:179-188 (2003); Skordis et al., *Proc. Natl. Acad. Sci. USA* 100:4114-4119 (2003)), *inter alia*, can be used. RT-PCR is used to measure the efficiency of *trans*-splicing. Protein expression levels can be verified with SMN-specific antibodies. As a consequence of high levels of full-length SMN protein (due to *trans*-splicing), a major macro-molecular reorganization in cells derived from SMA patients
10 should occur. Such reorganization will include redistribution of components associated with biogenesis of spliceosomal snRNPs (Liu et al., *Cell* 90:1013-1021 (1997)) and Cajal bodies (Hebert et al., *Dev. Cell* 3:329-337 (2002)). This macromolecular reorganization of spliceosomal components and telomerase can be probed by immunofluorescence using fibroblast cells from SMA patients.

Where the *trans*-spliced products are low, PTMs can be identified from negative
15 cell-based selection. A combinatorial library of PTMs can be generated, where target binding sequences are randomized. Unspliced transcripts can be amplified and recycled for further *trans*-splicing. This process is repeated several (5-8) times similarly to cell-based SELEX (Cooper, *Meth. Mol. Biol.* 118:405-417 (1999)). The unspliced pool will
20 be left with inhibitory sequences that are incompetent for binding to a target site.

While the invention has been described and illustrated herein by reference to various specific materials, procedures and examples, it is understood that the invention is not restricted to the particular materials, combinations of materials, and procedures selected for that purpose. Numerous variations of such details can be implied and will be
25 appreciated by those skilled in the art. All U.S. Patents cited herein are hereby incorporated by reference in their entirety for all purposes.

Example 10: Selection of Terminal Guanosine Residues Exceeds Statistical Probability

Terminal exonic positions interact with a number of protein factors at different
30 steps of splicing. However, due to 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 (Burge et al., in *The RNA World*, R.F. Gesteland and J.F. Atkins, eds. (Cold

Spring Harbor Laboratory Press, Cold Spring Harbor, NY), pp. 525–560 (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% instances for a constitutive exon (Burge et al., 1999, *supra*). In this case 1A, 1C and 1U should represent the
 5 remaining 50% of the molecules. Since our initial pool contained A, C and U residues in ~28% cases, we should have selected these residues in ~14% 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, cell-based selection
 10 of the first six nucleotides of exon 7 retained 1G in all the isolates (Singh et al., 2004, *supra*).

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, *supra*). 2G was isolated in ~81% instances, which is ~14%
 15 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 U2AF³⁵, which interacts with the 3' ss, had shown a strong preference for 2U (Wu et al., Nature 402:832-835 (1999)). This may suggest interaction of other proteins that are specific to exon 7. The nature of residues selected at position 2 matched
 20 with the cell-based selection of the first six nucleotides of exon 7 (Singh et al., 2004, *supra*).

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% cases in humans (Burge et al., 1999,
 25 *supra*). Since our initial pool contained 54G in ~11% 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*.

Example 11: Cell-based Selection Confirms the Presence of Exinct

Consistent with the presence of Exinct (Singh et al., 2004, *supra*), cell-based 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. 6D). 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, supra). 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, Hum Mol Genet 9:259-265 (2000); Cartegni and Krainer, Nat Struct Biol 10:120-125 (2002)) and cell-based selection of a stretch of the first six nucleotides of exon 7 (Singh et al., 2004, supra), 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, supra). 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, supra). 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.

Example 12: The 3'-Cluster Represents a Novel Inhibitory *Cis*-Element

The cell-based selection identified a number of mutable residues that form a 7-nucleotide-long cluster towards the 3' end of exon 7, which we call the "3'-Cluster" (Fig. 11B). 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. 11A). Deletion of UUA promoted exon 7 inclusion in *SMN2*, whereas deletion of the preceding codon had no effects (Fig. 11C, 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. 11C). Interestingly, a

single A51C mutation restored exon 7 inclusion to the level of *SMN1*. Due to 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, a
5 cell-based selection of a small stretch of sequence spanning this region preferred non-wild-type residues.

Example 13: The Tra2- β 1-ESE is Located Within a Long Conserved Tract

The double substitution G25U/G26U abrogates the Tra2- β 1-ESE (between
10 positions 19 and 27, Fig. 14) and has been shown to promote exon 7 skipping (Hofmann et al., 2000, supra). In agreement with the role of Tra2- β 1-ESE, none of the selected variants had G25U/G26U substitution (Fig. 6E). In isolates with less than 33% substitutions, a purine-rich sequence in the middle of exon 7 was (mostly) conserved (Fig. 6E). In some instances, pyrimidine residues within the Tra2-ESE were also selected,
15 suggesting that pyrimidines are not inhibitory for Tra2 binding. Consistent with this observation, a recent report demonstrates that a Tra2- β 1 binding site can contain a substantial number of pyrimidines (Stoilov et al., Hum Mol Genet 13:509-524 (2004)). Validating the stimulatory role of the long conserved tract, many mutations within this tract led to exon 7 exclusion in *SMN1* (Fig. 12A-B). Mutations used in our study are
20 different than an earlier study, which also demonstrated the stimulatory role of sequences that overlap the long-conserved-tract (Lorson and Androphy, 2000, supra). 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
25 wild-type residues. Since A36/C37 fall outside the earlier described binding site of Tra2- β 1 (Hofmann et al., 2000, supra), other protein(s) 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.

Despite the predominant selection of G residues at the terminal positions, the
30 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. While analyzing 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. 12A-B). These results suggest that multiple proteins may contact the long-conserved-tract in the middle of exon 7.

Example 14: Validation of the Relative Significance of the Terminal Positions

To validate the results of cell-based selection, we made substitution mutations of terminal residues that are generally ignored in computation programs (Fairbrother et al., Science 297:1007-1013 (2002)). The functional significance of G at position 1 was confirmed by G1H substitutions (H connotes for A or C or U), which decreased exon 7 inclusion (Fig. 13B, 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 inhibitory 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. 13B, lane 12, 13). These results demonstrate that the 54G-mediated strong 5' splice site 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. While 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. 13B, 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. 5B). Interestingly, all of these

clones had multiple mutations within the 3'-Cluster (Fig. 6E), supporting the inhibitory nature of other residues that contributed to the creation of the weak 5' ss of exon 7.

Cell-based 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. 13A-B) confirmed its critical role in the definition of this exon. The dominant role of 1G can not be replaced by any combination of substitutions within the first six positions of exon 7 (Singh et al., 2004, supra). In addition to the interaction with U2AF³⁵ (Wu et al., 1999, supra), functions of 1G may include the interactions with proteins that participate in the second transesterification step (Moore, Nat Struct Biol 7:14-16 (2000); Collins & Guthrie, Nat Struct Biol 7:850-854 (2000); Chua & Reed, Mol Cell Biol 21:1509-1514 (2001); Lallena et al., Cell 109:285-296 (2002)). Control at the second step of splicing is consistent with the proofreading function of the spliceosomal machinery. Insertion of a G residue at intron 6/exon 7 boundary obliterates exon 7 inclusion (Singh et al., 2004, supra). This clearly demonstrates that 1G executes a stimulatory role only in the context of a 54-nucleotide-long exon 7, not the 55-nucleotide-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. The results described herein provide evidence that the 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, supra). 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. 6E).

One of the important findings of cell-based selection was the role of the non-coding 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, supra), 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 homologue of the U1C protein showed its preference for a G residue at the last position of

exon (Du & Rosbash, 2002, supra). Interaction of U1C protein precedes and possibly assists U1 snRNA base pairing at the 5' ss. Further, 54G may break a 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. 13B, 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 & Hertel, J Biol Chem 276:45476-83 (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. 13B, lane 16). In several other instances such as exon 11 of NF1 (Fang et al., J Mol Biol 307:1261-1270 (2001)), exon 9 of human ATP synthase γ -subunit (Hayakawa et al., J Biol Chem 277:6974-6984 (2002)) and exon 12 of CFTR (Pagani et al., Hum Mol Genet 12:1111-1120 (2003)), the exons are alternatively spliced despite G at the last position. In addition, 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., Biochem Biophys Res Commun 283:762-9 (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. 6E). This also demonstrates that the selected 54H variants were equally good if not better than the 54G containing isolates. It is logical to believe 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 cell-based 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. 11B-C).

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 43-45 (Fig. 11A). 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. 11B). This is 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 ensues at either end 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 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.

OTHER EMBODIMENTS

It is to be understood that while the invention has been described in conjunction with the detailed description thereof, the foregoing description is intended to illustrate and not limit the scope of the invention, which is defined by the scope of the appended claims. Other aspects, advantages, and modifications are within the scope of the following claims.

WHAT IS CLAIMED IS:

1. A method of identifying nucleic acid sequences that modulate RNA splicing, the method comprising:
 - providing a pool of nucleic acids, wherein each nucleic acid includes a partially randomized exonic sequence;
 - subjecting the pool of nucleic acids to one or more rounds of selection in a living cell;
 - selecting those nucleic acids with pre-selected splicing characteristics;
 - determining a distribution of mutations that are associated with the pre-selected splicing characteristics,wherein the distribution of the mutations indicates which sequences modulate RNA splicing.
2. The method of claim 1, wherein each of the partially randomized exonic sequences comprises an entire exon.
3. The method of claim 1, wherein each of the partially randomized exonic sequences comprise at least about 50 nucleotides.
4. The method of claim 1, wherein each of the partially randomized exonic sequences comprise at least about 75 nucleotides.
5. The method of claim 1, wherein each of the partially randomized exonic sequences comprise at least about 100 nucleotides.
6. The method of claim 1, wherein each of the partially randomized exonic sequences comprise about 70% wild type nucleotides.
7. The method of claim 1, wherein the pre-selected splicing characteristic is inclusion or exclusion of an exon.
8. The method of claim 1, wherein subjecting the pool of nucleic acids to one or more rounds of selection in a living cell comprises
 - (a) expressing the pool of nucleic acids in a eukaryotic cell,
 - (b) isolating spliced mRNA from the transfected eukaryotic cell,
 - (c) amplifying an exon-included product from the isolated mRNA,

- (d) isolating the exon-included product,
 - (e) inserting double cutter or quadruple cutter restriction enzyme sites outside the boundary of the included exon to create a double cutter- or quadruple cutter-engineered oligonucleotide,
 - (f) digesting the double cutter- or quadruple cutter-engineered oligonucleotide with a double cutter or quadruple cutter restriction enzyme and purifying the exon, and
 - (g) ligating the purified exon into a double cutter- or quadruple cutter-linearized plasmid.
9. The method of claim 8, further comprising
- (h) transfecting a eukaryotic cell with the ligated plasmid of step (g) containing the exon.
10. The method of claim 8, further comprising
- (j) transfecting a bacterial cell with the ligated plasmid of step (g) including the included exon and analyzing the sequence of the exon to determine where wild-type nucleotides are retained or where non-wild-type nucleotides are substituted, and
 - (k) designating a position where wild-type nucleotides are retained as a portion of a potential positive splicing element and positions where non-wild-type nucleotides are substituted as a portion of a potential negative splicing element.
11. A method of screening for candidate compounds that enhance splicing of a Survival of Motor Neuron (SMN) RNA, the method comprising:
- providing a ribonucleic acid comprising SEQ ID NO:247 or an active fragment thereof;
 - contacting the ribonucleic acid with a test compound; and
 - measuring binding of the test compound to the ribonucleic acid,
- wherein binding indicates that the test compound is a candidate compound to enhance splicing of an SMN RNA.
12. The method of claim 11, wherein the ribonucleic acid comprises nucleotides 1-22, 23-40, or 35-54 of SEQ ID NO:247.

13. The method of claim 11, wherein the ribonucleic acid comprises nucleotides 3-15, nucleotides 16-44, or nucleotides 45-51 of SEQ ID NO:247.
14. The method of claim 11, wherein the ribonucleic acid comprises the Exinct region, the binding site of SP2/ASF, the conserved central region, the Tra2 binding site, or the 3'-Cluster of SEQ ID NO:247.
15. A method of screening for candidate therapeutic compounds for the treatment of spinal muscular atrophy (SMA), the method comprising
expressing a deoxyribonucleic acid comprising a sequence encoding a Survival of Motor Neuron 2 (SMN2) (SEQ ID NO:247) ribonucleic acid in a cell;
contacting the cell with a candidate compound identified in claim 3; and
detecting the presence of a properly spliced ribonucleic acid transcribed from the deoxyribonucleic acid in the cell;
wherein the presence of the ribonucleic acid indicates that the compound is a candidate therapeutic compound for the treatment of SMA.
16. A method of identifying a therapeutic compound for the treatment of spinal muscular atrophy (SMA), the method comprising administering a candidate therapeutic compound identified in claim 12 to an animal model of SMA, and determining the effect of the candidate therapeutic compound on a symptom of SMA in the animal model, wherein an improvement in a symptom of SMA indicates that the compound is a therapeutic compound for the treatment of SMA.
17. The method of any one of claims 11-16, wherein the test compound, candidate compound, or candidate therapeutic compound is an aptamer
18. A method of treating a patient having spinal muscular atrophy (SMA), the method comprising administering to the subject a therapeutically effective amount of an aptamer that enhances inclusion of exon 7 of Survival of Motor Neuron 2 (SMN2), wherein the aptamers bind to regulatory elements that interact with cis-elements.
19. The method of claim 17 or 18, wherein the aptamer comprises the sequence UUUUUGAUUUUGAUCGUAUGAUGAUGA (SEQ ID NO:253).

20. A method of diagnosing a subject with SMA, the method comprising obtaining a sample of genomic DNA from the patient, and determining the presence of one or more sequences that negatively affect splicing of exon 7 of SMN, wherein the presence of a sequence that negatively affects splicing of exon 7 of SMN indicates that the subject has SMA.
21. The method of claim 20, wherein the sequences that negatively affect splicing include one or more of Exinct and/or the 3'-Cluster.

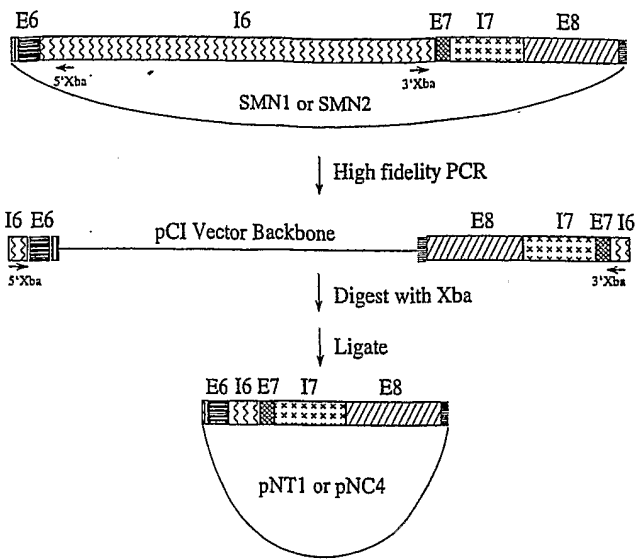


Fig. 1A

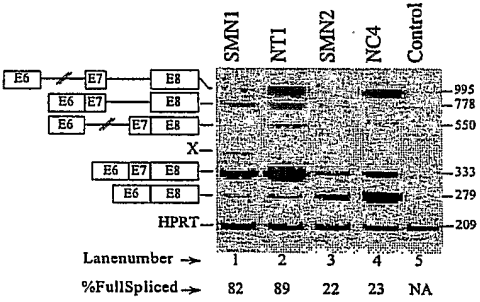


Fig. 1B

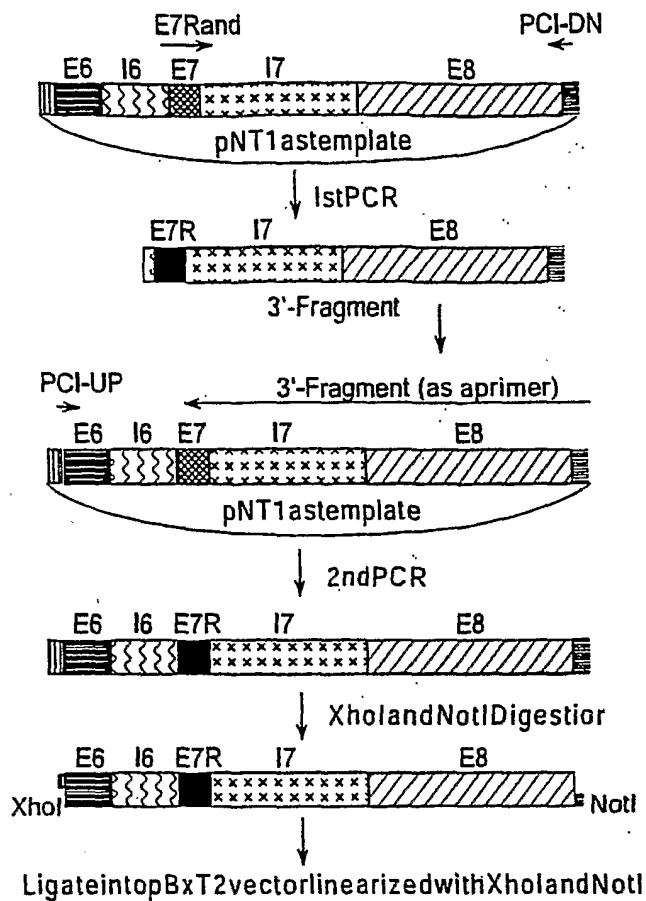


Fig. 2

Sequence ID. No.

No.	Name	Subst.
1	pNT1	NA
2	P0-01x	41
3	P0-02x	16
4	P0-03x	38
5	P0-05x	24
6	P0-06x	24
7	P0-02	17
8	P0-03	21
9	P0-04	15
10	P0-06	14
11	P0-10	23
12	P0-11	17
13	P0-16	29
14	P0-18	13
15	P0-19	20
16	P0-20	18
17	P0-21	16
18	P0-24	11
19	P0-25	18
20	P0-26	13
21	P0-28	24
22	P0-29	17
23	P0-32	15
24	P0-33	18
25	P0-35	21
26	P0-38	15
27	P0-43	16
28	P0-44	17
29	P0-45	14
30	P0-46	12
31	P0-48	15
32	P0-49	16
33	P0-50	20
34	P0-51	18
35	P0-52	16
36	P0-53	9
37	P0-54	24
38	P0-55	17
39	P0-56	28
40	P0-57	12
41	P0-65	19
42	P0-67	25
43	P0-68	18

Fig. 3

Fig. 4A

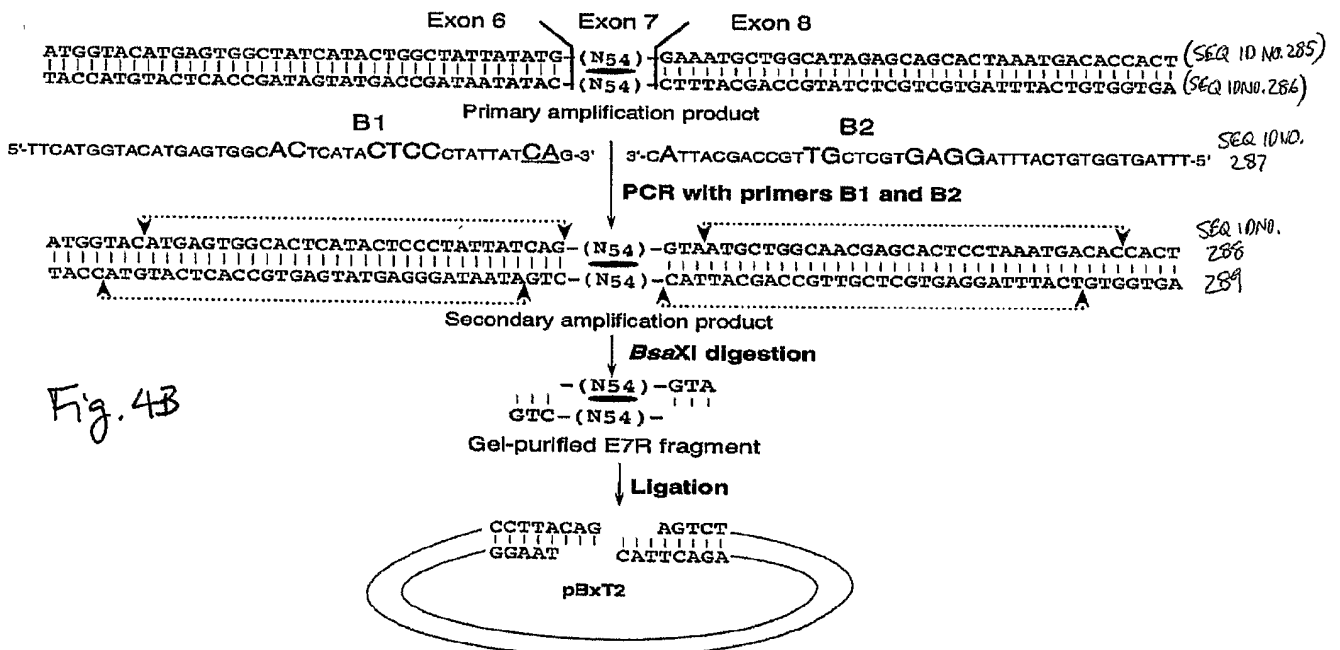
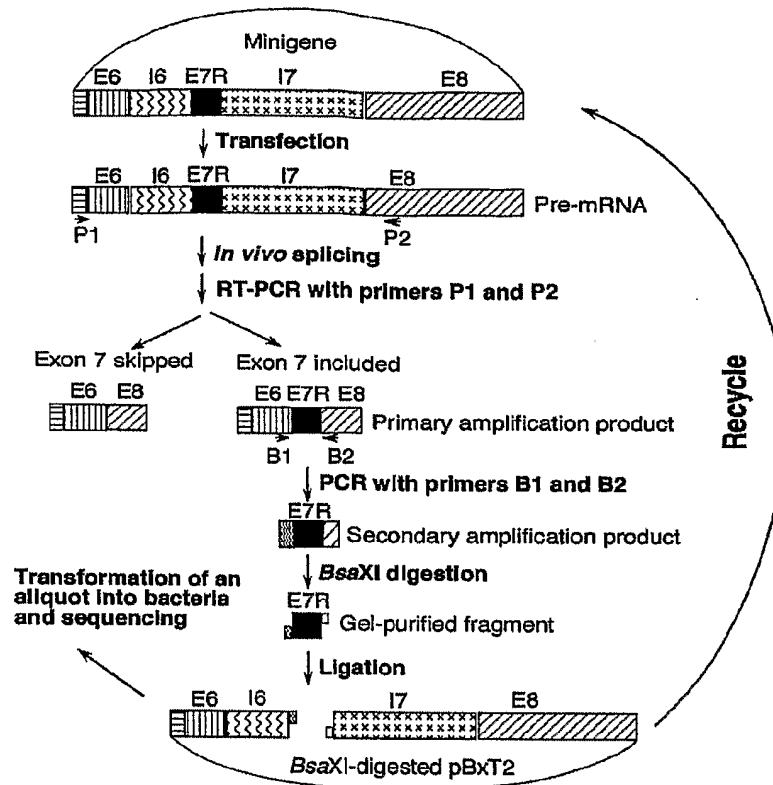


Fig. 4B

SEQ ID
NO:

132 P04-61 GGCTGCGCAA - T T A T T T C A A G A C G C A - G G G A G G T G C T C A G A T T C C G T A C G T T A T G G G
133 P04-62 G T T G C G A T A - C A A A A T T A A A A A G - A T G G A A G C C G - T T A C A T T G C C T T T A G T C T T G G A
134 P04-63 G G C T C C A T A - C A A G A T C A A A A T G C G - T G A A G G G G A C G T C G T G G C T T T A A A T T A G G G
135 P04-65 G G A T A A A G - T A C G A A C A A A A A G T T - G G A A G G T G C T C A C A A T T C T T A C A T T A G G
136 P04-67 G G T A T C A G A - C T A G T T C T A A A C G A A - G G G A G G T G A T C A C A T T C T T A T A T T A G G A
137 P04-68 G C C T T C A G A - C A A A A T G A A A A G G C - G G C A G G T G C C C A C A G C C G T T A A A T T A T G G G
138 P04-69 G G T A T C C T A - T G A A C T C A T A A A G G A - G G A A G G T G C T C G C A T T G A T T A A A T T A T G G G
139 P04-70 G G G G C C A G A - C A A G A T T A A G A A G T A - G G A A G G T G C T C A G A T G C C T G A A A C T A A G G G
140 P04-71 G G G C T C A G G - C A C T T A C T A A A A G A A - G G A A G G G G C T A A T A T T T C T T A C T T A A G G G
141 P04-72 C G G G G C A T T - G A G A A C A A A A G A A A - C G A A G A T G G T C A C A G C C C T T A A A T C A A G T G
142 P04-73 G G T T T G A G A - C A A C G T C C A T A A G A A - G G A T G G T G A T T C A T T T C C T T A A C T A A A G G G G
143 P04-74 G G C G C G G C G - G C A A T C C A G G A A G A G - G G G A G C T G T T C A G A T T C C T T A A A T T A - G G G
144 P04-75 G G T T G A G G T - G C T A T G A A A A A G G C A - G A G G G T G T T C C C G T T C C T T A A A G T A A G G G
145 P04-76 G G C G C C T C A G A C A T A A G C T G A A A G A A - G G G A G G T T C T C A G T T G C C T T T A G T C A G G G
146 P04-77 G G T C C G C G G - G A A A C T C T G A A G T C A - G G A G C G T G T T G A C A T T C A T T T A A T T A A G G G
147 P04-78 G G C B T A A G A - C C A A A T C G A A A T G A A - G G A C G G T G C T C C C C T T - G T T A G A T T A T G G G
148 P04-79 G G G T T C T G A - C A A C A T G A A A A A G A T - G G A A G G T G G G C G C A T G C C T T A A G T T A T G G
149 P04-80 G G T T T G A G C - C A A A T G T G G T A C G A T - G A A A G G T C C T C A C T T G C C T T T G T A A G G G
150 P04-81 G C G T G C A G A - C T T A A T C A A G G A G T G - G T A A G G T G A T C T C A T T C T T C A A A A T A C G G A
151 P04-83 G G T C A C A G A - C A C G A T C A C T A A G A A - G G A A G G A G A G T A A A G T C G A T A A T A T G C G C A
152 P04-85 G G T T A A C G A - C A A G A A C A A G G T G G A - G G A A C G T G C A C A C A T T C C T G A A T C T G A G G G
153 P04-86 G G T G T G A A A - G A A G A C A A G A A A A - G G A A G G T T C T C A C A A T C C A T G A T G A A G G G
154 P04-88 G G A A T C T G A - C A A T C G A A A C A A T A G - G G A T A G G A C T C C C G T T A C A T G G A C T A A G T G
155 P04-89 G A A C T C A G C - C C A A C A C A A A G A G A A - G G A C T G T G T T G A C A T T C A T T T A A A T T A A G G A

Selected clones from secondary *in vivo* SELEX of first 52 nucleotides of exon 7 of *SMN1* gene

151 A01-01 G A T T T C G A G A - C G A A A T C A C A A C T G A - G G A A A G T G T T C A C A T T C T C G - T A A T T A A G
152 A01-03 C G G T T A A T G - C G A A A T G A A A A C G A T - G C A T G G T G C T T C C A T T A C T C - A T C T G A A G
153 A01-05 G T G T T C A G A - C A A A C T C A A A A A T A A - G C A A G G G C G C T C A C A T T G C A T T - A A A T T A T G
154 A01-07 G G G A G C A T C - C A C A A T C A A A A A G A A - T G G A G G G G G G C G C G T G C A T T - A A A T T C A G
155 A01-08 G G T A T C C T G - C T G A A T C A A A A T G A C - G G C A G G T G T T C G C A A T C C G T - C C A T G A A G
156 A01-09 G C T T T C A G A - C A A T A T C A A A A A G A - T G A A G T G C A C C C G G T C G T C - A A A T T C A G
157 A01-11 G G T A G C A G G - T T A A G T C A A A A T G A A - T G A A T C T G C T C A C A C T A C G T - A T A T T C C G
158 A01-12 G G T A G C A G G - T T A A G T C A A A A T G A A - T G A A T C T G C T C A C A C T A C G T - A T A T T C C G
159 A01-13 G A T T T C T T G - C A C A G C C G A A G A A - T G A A G G T G C A C A C A G A T C C G T - A C A T T A A G
160 A01-14 G G T T T C A G A - C T C A T T C A A A G C A A - G G A C G G T G C A C A C A G A T C C G T - G G A T T A A G
161 A01-15 G C T A C A G A - T C A A A T C A C A A A G A A - C G A A G A T G G T C A C A G A T C C G T - A C A T T A A G
162 A01-16 G G C T A C G T A - C A T T A T C A C A A G C G T A - G G A C G G T G G G C A C C T T C C T - G A C T T A A G
163 A01-17 G G T C T T A G A - C A G A C T G G A A A G A T - G G T G G G G G T T C A T C T T C C T - T A A T T A A G
164 A01-18 G G C T A C A A G T - C A A G T G A C A A A G T A G - C A A C C T G C T T C C A T T G C T G - A T A T A T A G
165 A01-19 G G T T T G T G G - C A C A A T G C A A A A G A A - T A A G G T G C G G A C T T C C C G T - T A C T T A A G
166 A01-21 G T A A A C A G A - C A T T T C T A A A A G A A - G G A T G G T C C A C A C A T T C C G T - A A C G T A A G
167 A01-23 G G T T T C T G A - C A A C A T G T C A A A G A C - G A T C G G T A C G C A C A T T C C G T - A A A C T A A T
168 A01-24 G G T T T C C G A - G C C A A C C A T T T A G A A - G G A G G G T G C T G A C A T T C C G T - C T C T T G A G
169 A01-25 G G T T C C G G A - C A A G T T C A A A A A G A C - G A A G T C T G C T C A C A G G C G A T - G A T A T A A A
170 A01-26 G G T T A C C G T - G A T A G T C G A A A A G A A - G G A A G G A C C T C A T C C G G T - A A G T T A A G
171 A01-27 G G T T T A A G A - C A A A A T C A A A A A T T C - G G A A G C G G G T C A C A T T C C T T - A C A T T T C G
172 A01-28 G C T G T C C T A - C A T A A T C G C G A T G A T - T G A A A A T G T A T G C C G A A C T T - A A G T T A A G
173 A01-29 G G T T T C A G T - C A A T A T C A A A A A A A T G G C - T A C T C A C A T T C C T C - A T A T T A T A
174 A02-02 G G T T T C A G T - C G C T A T T A A A A A T A T - G G C A G G A G C C T A T A C C C C T C - G A T T T T A A
175 A02-07 G G A T T C C A G T - C G A C A T C G A T T A G A A - G G C C G G T G C T A A C A T C C C T C - A A G T C A A G
176 A02-08 G A T G T C A G T - C A C A A C C A A A A G G G A - G G A A G G C A A T C A C G T T C A T T - A T T T C A A G
177 A02-09 G A G G T C C T A - C T A G A T C A A A A A G A A - G T A A A G T G C G C A A G A G C C T T - A A C T T A A G
178 A02-10 G G T T T G A G G - C A A A G T C A A G A A T A T - G G A C G G T G G T T G G A T T C A T T - T A T T T A T G
179 A02-11 G G C A G T A G A - C G A G A T C G A A A G A A - G T A C G G T G C G C A C A T T C C T T - A A C G T A A G
180 A02-13 G G A T T C T G A - G A C A G T C G A A A T G A A - G G A T G G T G C T A T C A T T G C T T - A A A T T T T G
181 A02-14 G G C T T C A T A - C A T C A G C A A A A T G C G - T G A A G G G G A C G T C G T G G C T T - T A A T T A A G
182 A02-15 G G T T T C G G T - C A A A - C C A C T A A A G A C - G T T C G G T G G A C A T T T C C C T A - C A A T T A C G
183 A02-16 G C T T T A A G G - A A G A A T C A G A A A G G C - C G A C G G T G C T C A C A T A G A T - T A T T T A A G
184 A02-18 G G T G A C T G A - G A A A C T C A G A G A G C - G C C A G G T G C C C C A A T A T C G C - A A C T G A A G
185 A02-19 G G C T T C A T A - C A T C A G C A A A A T G C G - T G A A G G G G A C G T C G T G G C T T - T A A T T A A G
186 A02-23 G G T T T G A G G - C A A A A G T C A A G A A T A T - G G A C G G T G G T T G G A T T C A T T - A A A T T A A G
187 A02-25 G G T T T C G G A - G T A A A T C A A T A A G A A - G G A A G G T G C T C T T C T T C C T T - T G T A T A A G
188 A02-26 G G G A G A A G A - C A A A A T C A A A A A G A A - G G A A G - T G C T C A C A T T C C T T - A A A T T A A G
189 A02-27 G G G A C A G A C - G A A G T C T A A G A A A A - G G C A G G C G G G T A C A T T T G G - C A A G G A T A
190 A02-28 G G A G T C A C A - C A G A A T C A G A T G C C - G G A C C G T G C T G A C A T T C C T T - A G T T G A A G
191 A02-29 G A T G A C A G A - C A A C A T C A T A C A G A A - G G A T G G T G T T T G T T G C C G T - T T A T T A A G
192 A02-32 G G T G T T A G T - C A A A G G C A A A A G G A A - G A A T G A T T C C T A C A G T C C T T - G A A T T A A G
193 A02-35 G G G T A C A G T - C A G A A T C A G A G G G A - C G G A A G G T G C T T G T T T C T G A - A A A T A A T G
194 A02-36 G G G C T A C G G - A C G A C A A C A G G A C G A - T G G A G G T G C T T A C A G T C A T T - A C A T T A A G
200 A02-38 G G T T T C A G A - C A A C G T T A A G A C A A G - G G A C T G A G C C C A C A T T C C T G - G A A T C A G G
201 A02-39 G G T T T C A G G - C A T G A G C A A C A G G G A - G G T C G T T G A A T A C C T G T C T T - A A T T T A A C
202 A02-42 C G T T T C A G A - C A A T A T C A A A A A G A - T G A A G T T G C A C C C G G T C G T - A A A T T C A G
203 A02-43 G G G T T C C T A - C A A A A T C A T G C A G A T - G G A A G G T G C T A A C A T T C C C T - C A A T T G A C
204 A02-44 G G T T T C G C - C A A C A G C A T A G C G A A - G G A C G G T G C T G A - A C T A C A A - A A A T T C C G
205 A02-46 G G T T T T A G G - C A A A A T G A A G A A G C G - G G A A G T T A C T C C C A T T C C C T - A G A T G A C G
206 A02-47 G T T T T T A C A - A A A A A T C A A A T G G A T - G G A A G G T G G T C A C G T G C A T T - G A T T C T C G
207 A02-48 G G T T T T A G A - C G G A A - C A G T C C C A A - G G T A G C T G C T G A C A T T A C T T - T A A T C A T C
208 A02-49 G G T C G C A G A - C A A A A T C A A A C T G G A - C C T A A T T G T T C A T A G T G T T - T A T T T G A A
209 A02-50 G G T T T C A G A - C A A C G T C A A G A C A A G - G G A C T G A G C C C A C A T C C C T G - G A A T C A G G
210 A02-51 G G A T A A A T T - T A A A A T C T A A G A G A A - G G A A G G C G C T C A C A G T T C C C - C A A C T T A A C
211 A02-52 G G C A G G G T C C T G G A C A A G T T C A G C A C - G G A G T A C G G T G C T T C C A A T G T G C A A A T T A A G
212 A02-53 G G A C T C G G G - C T A A A T C A A G A A G A A - G C A A T G T G C T G A C A T T A C G T - G A A T T A A A
213 A02-54 G G T T C C A G A - A A A A C C A G A C G G G G C - T T A T G G T G C A A A C A T T C C A G - A A A T A T A A
214 A02-55 G G T T C C G A - C A T C T G A G A T A C G C - G G C A G G G C C G C A C A T T C T T - A A C T T T T G
215 A02-56 G G T T C C A G A - C A A G A T C A T T A C G A A - G C C C G C T G A C A C A G T C C A A - A T G T G A A A
216 A02-57 G G T T C C G G A - G A A G A T C A T T A C G A A - G C C C G C T G A C A C A G T C C A A - A T G T A T G
217 A02-58 G G T T A T T G G - C A A A G T C A A A A A G A A - G G A A G G T G C T C A C G T T C C T T - G A A T T C G G

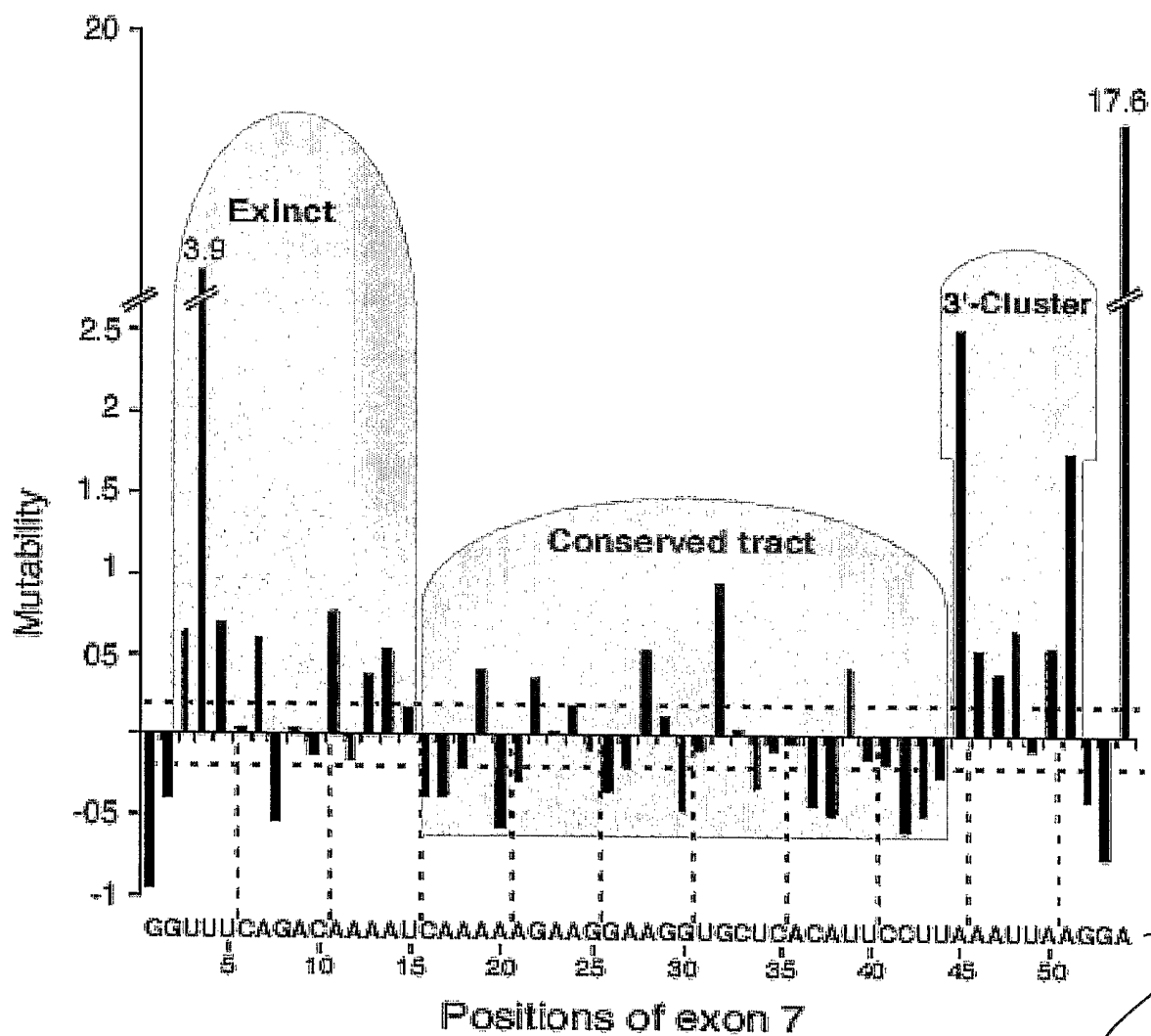
Fig. 6B

SRQ ID NO!

219	A02-59	GGTTTTCGGG	-	GAATATCAAATAGAT	-	ATCAGTGGCCCGCTTTCCTC	-	ATCTTGAA
220	A02-60	GGTGTTCAGA	-	AGTGGCCAAATAGAT	-	GGAGGAGGGTACGCTTCGTT	-	AAATTAAG
221	A02-61	GTGTTCCGA	-	CAAAATCAAATGGA	-	AATTGGTGATCACTTTGCGA	-	GAATCAAG
222	A02-62	GGTTGCTGG	-	CACGAACACAATGAA	-	GGAGGGTGGGACATTTGTG	-	TATTGAAG
223	A02-63	GGTTTGAGG	-	CAAAGTCAAGAATAT	-	GGACGGTGGTTGGATTCAAT	-	TATTTATG
224	A02-64	CGTTTCAGA	-	CACAATCAATTAGAA	-	GGACGCTGCTGACATGGCAT	-	TATTTAAG
225	A02-65	GGGTCTCAGA	-	GTAAATCAGAAAGAA	-	TGGAAGCGATCAAACCTCCTT	-	ATATATG
226	A02-68	GGAGACAGA	-	CACAATCAAGAGCAA	-	GCAAGGTGTTCAATTACCGT	-	AATATAAC
227	A02-69	GCATTCAGA	-	AAGTATCAAAAAGCA	-	GGAAGGTGCCGACATTCCTT	-	ATCTACG
228	A02-70	GGTTTGTGA	-	CGCGATGTGAAAGAA	-	GGAAGGTTCCTCAAAATTTCTT	-	AAATCAAG
229	A02-71	GGTTTGGGA	-	CGAACCTAAGACGAT	-	GGAAGGTGCCGACGATCCTT	-	AAATGCG
230	A02-72	AGAATGAGG	-	CAGAATCAAACAGGA	-	GGCGGGTGATCACATTCCTT	-	CAAGTATG
231	A02-73	GTTTCCAGA	-	ACAAACGAAACGGTA	-	GGAAGGTGCACACATTGATC	-	AAAATAAG
232	A02-74	GGTTTTGGA	-	AAAGTTGAAAAGAA	-	GATAGGCGGAGATAATCCTT	-	GATTTATG
233	A02-75	GGTA-TAGA	-	CCCGATCCAGTAGAA	-	GGAAGGTTCGACATTTCATC	-	AAGTTGAG
234	A02-76	GATTACTGA	-	GCAAATGGAATCGTA	-	CGCAGGTGCTCACTTTTTGT	-	ACATTAAG
235	A02-77	GGTTGGAGA	-	CCGAATCCAGAGGGA	-	AGGCGCA-CTCAGTCGTTCC	-	ATTCTG
236	A02-78	GGTCTTAGA	-	CGAAATCGAAGGGGC	-	CGTGAAGCGGACATTCGCT	-	GAATTAAG

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52

Fig. 6C



SEQ ID NO: 6.

Fig. 6D

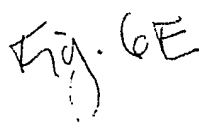


Fig. 7A. Aptamer-mediated splicing of exon-7

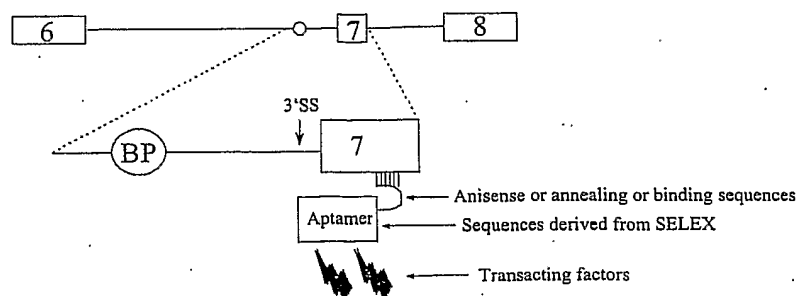
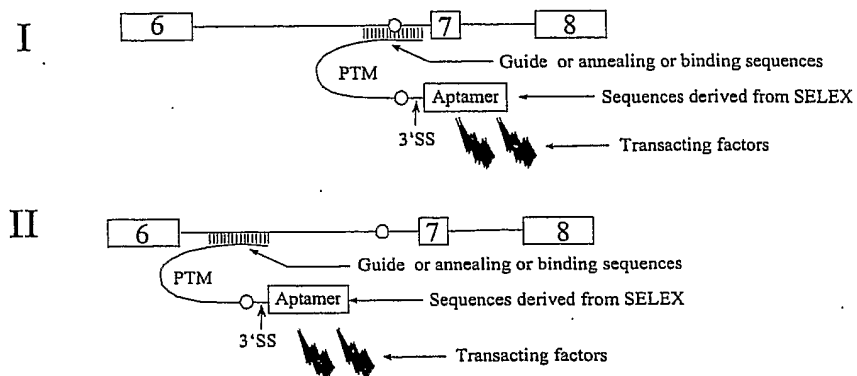


Fig. 7B. Spliceosome-mediated RNA *trans*-splicing (SMaRT)



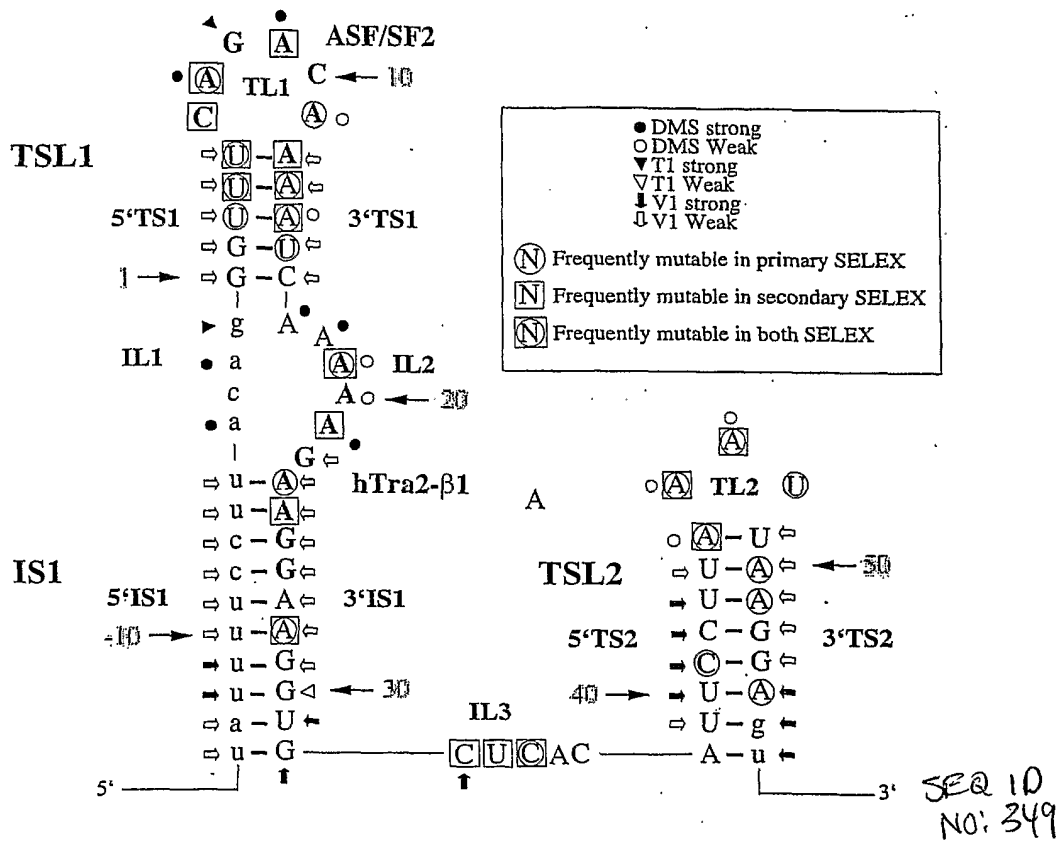


Fig. 8

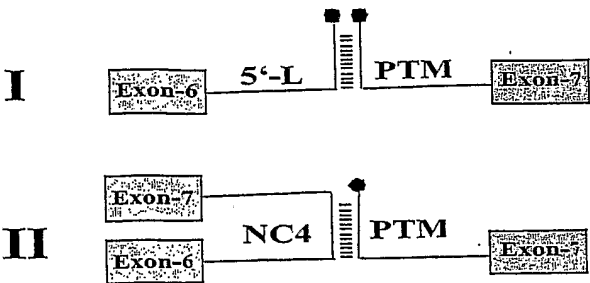


Fig. 9A

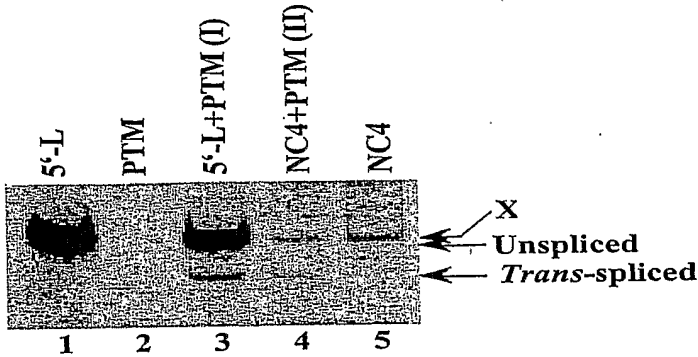
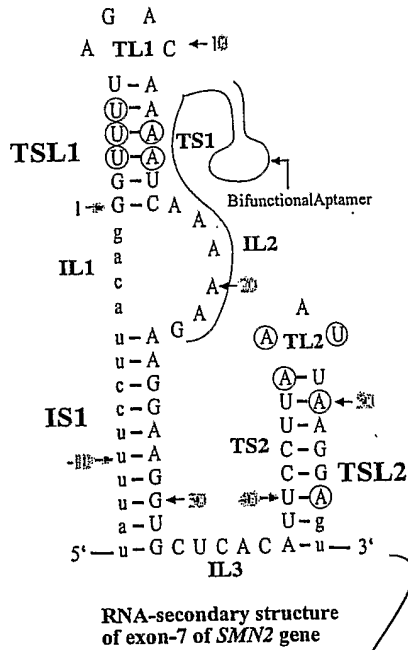


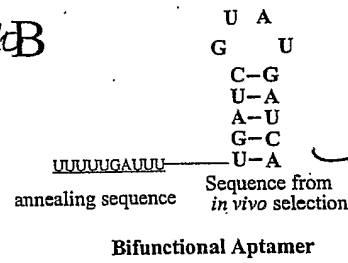
Fig. 9B

Fig. 10A



SEQ ID NO. 252

Fig. 10B



SEQ ID NO. 253

Fig. 11A

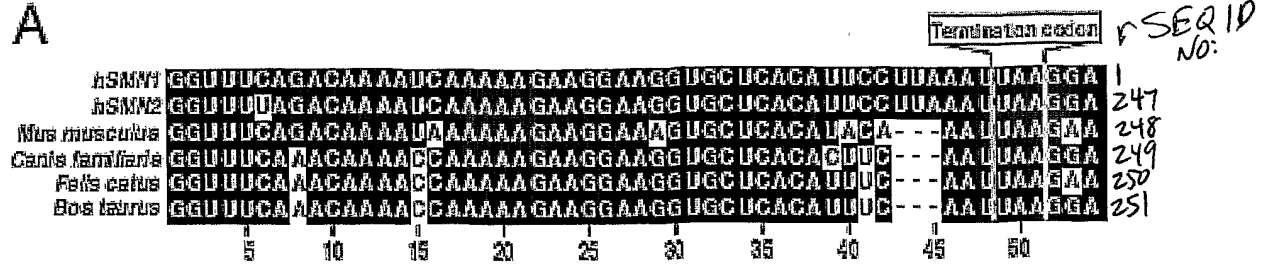


Fig. 11B

(i)

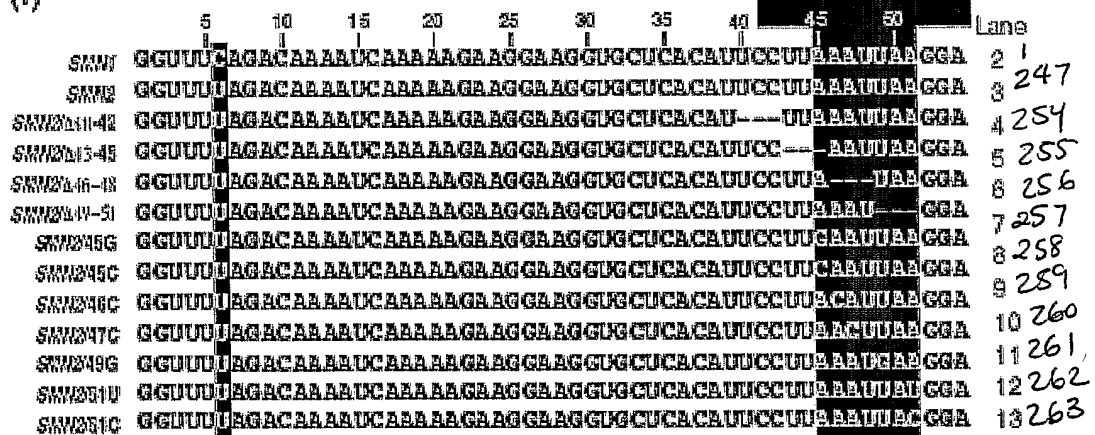


Fig. 11C

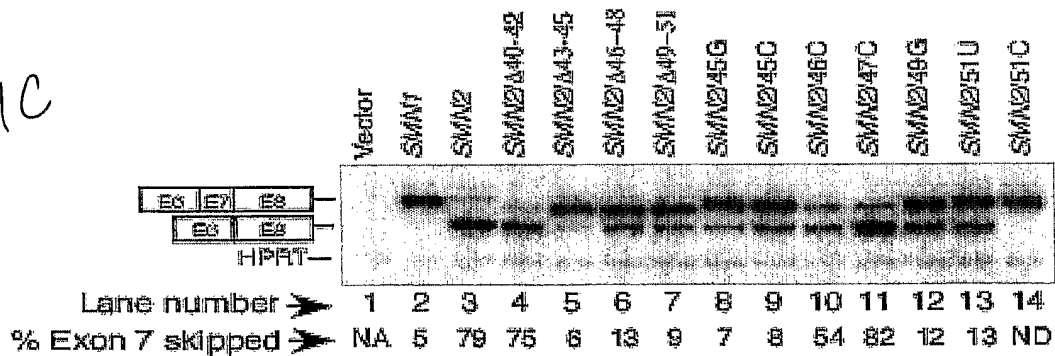


Fig. 12A

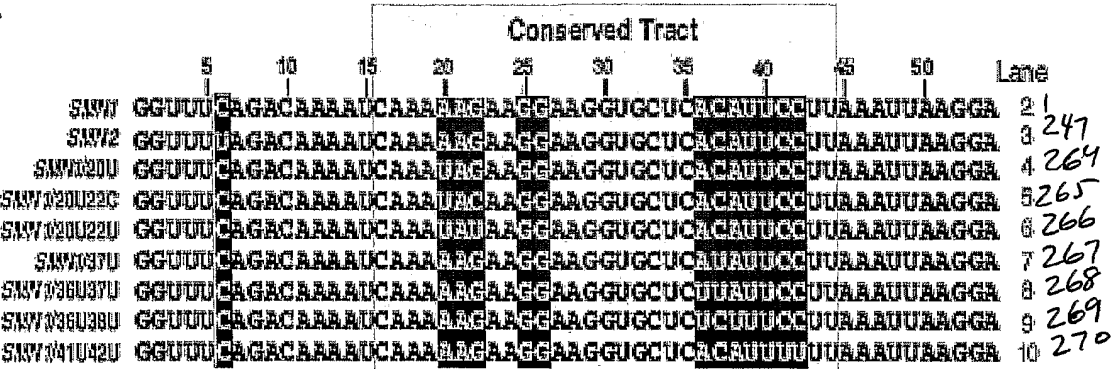


Fig. 12B

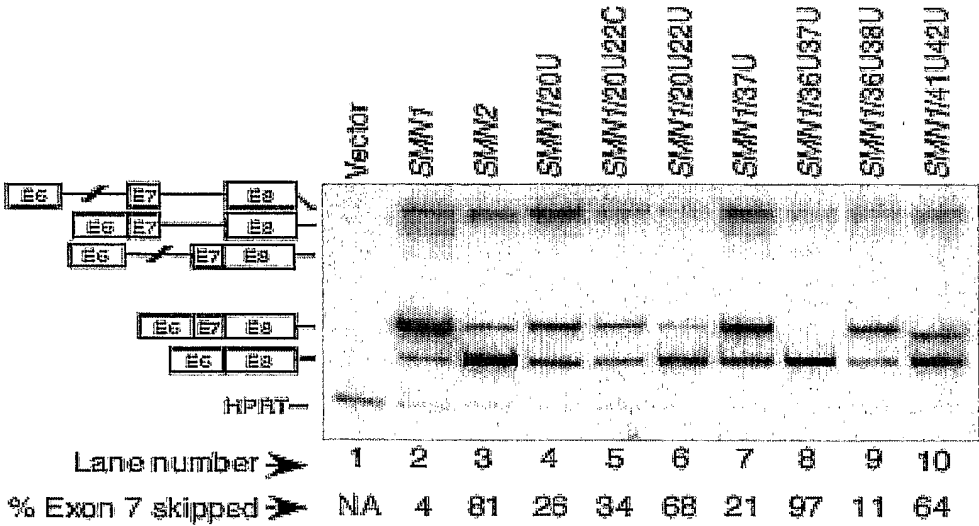
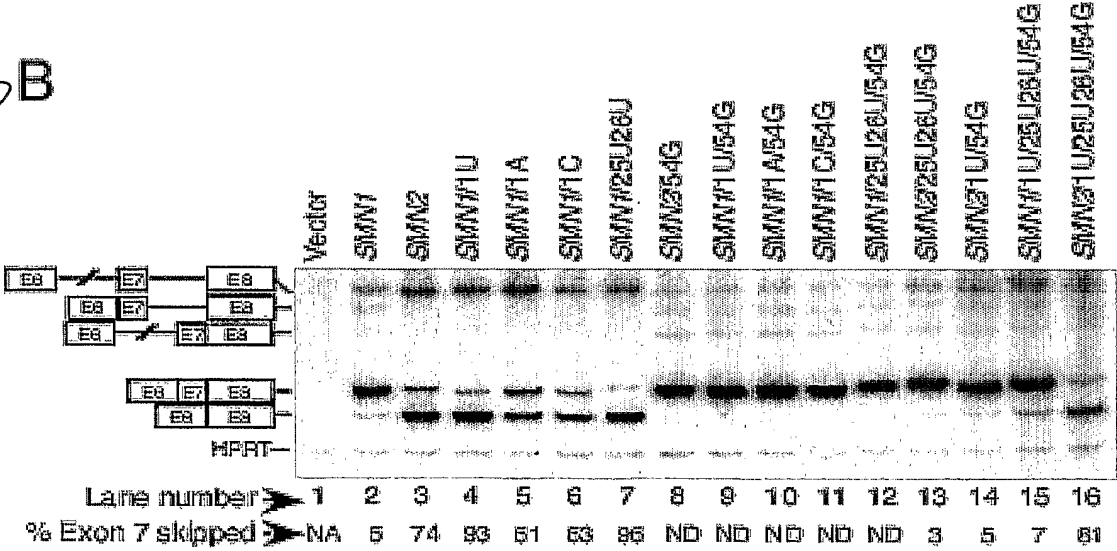


Fig. 13A



Fig. 13B



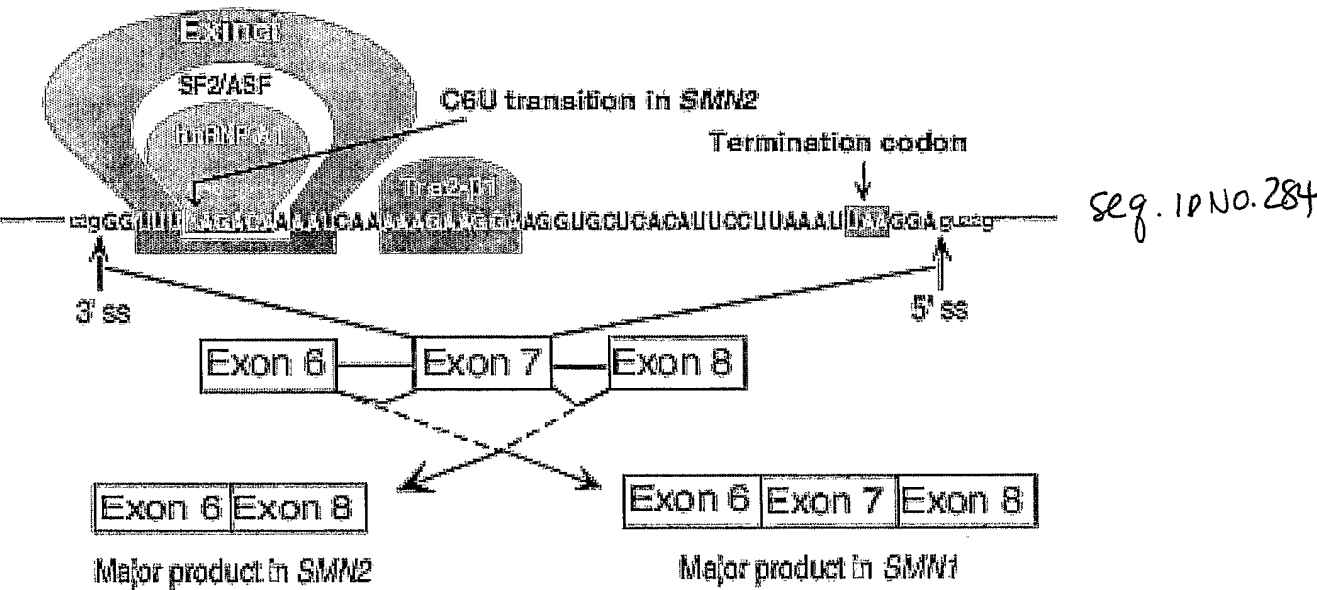


Fig. 14