[Frontiers in Bioscience 12, 3263-3272, May 1, 2007]

Unfolding the mystery of alternative splicing through a unique method of *in vivo* selection

Ravindra N. Singh

Department of Medicine, University of Massachusetts Medical School, Worcester, MA 01605-2324, USA

TABLE OF CONTENTS

1. ABSTRACT

Alternative splicing of pre-messenger RNA (pre-mRNA) is a fundamental mechanism of gene regulation in higher eukaryotes. In addition to creating protein diversity, alternative splicing provides the safest mode of gene evolution. Of late, more and more forms of alternatively spliced transcripts (mRNAs) are being discovered for key genes. Some of the alternatively spliced transcripts are also associated with major human diseases. This has created a sense of urgency to find the methods by which regulation of alternative splicing of specific exons could be best understood. Here I review a powerful *in vivo* selection method that uses a combinatorial library of partially random sequences. Several advantages of this method include *in vivo* analysis of large sequences, identification of unique sequence motifs, determination of relative strength of splice sites and identification of long-distance interactions including role of RNA structures. This unique method could be applied to identify tissue-specific cis-elements. Similarly, the method is suitable to find cis-elements that become active in response to specific treatments of cells. Considering this unbiased method uses *in vivo* conditions, it has potential to identify critical regulatory elements as therapeutic targets for a growing number of splicing-associated diseases.

2. INTRODUCTION

Nature has embraced a well-thought-out plan in which organisms have continued to evolve by mingling relatively less-conserved non-coding sequences (introns) with the increasingly conserved coding sequences (exons). Thanks to this arrangement, alternative splicing creates more than 200,000 proteins from about 30,000 genes in humans. A similar number of genes in mouse produce equally high number of proteins. However, several differences between human and mouse genome is responsible for producing species-specific alternative splicing. Spliceosome, a macromolecular complex, catalyzes the dynamic process of pre-mRNA splicing (1). However, non-spliceosomal factors are key to the process of alternative splicing (2-7). Some of these non-spliceosomal factors are expressed only in certain tissue-types providing the tissue-specific splicing (8-10).

Unique arrangement of cis-elements (or regulatory sequences) on pre-messenger RNA (pre-mRNA) assists the splicing proteins in defining the exon-intron boundaries or the 5′ and 3′ splice site (5′ss and 3′ ss) located towards the 5′ and 3′ ends of intron, respectively. Traditionally, cis-elements have been referred to as exonic or intronic splicing enhancers (ESEs or ISEs) and silencers

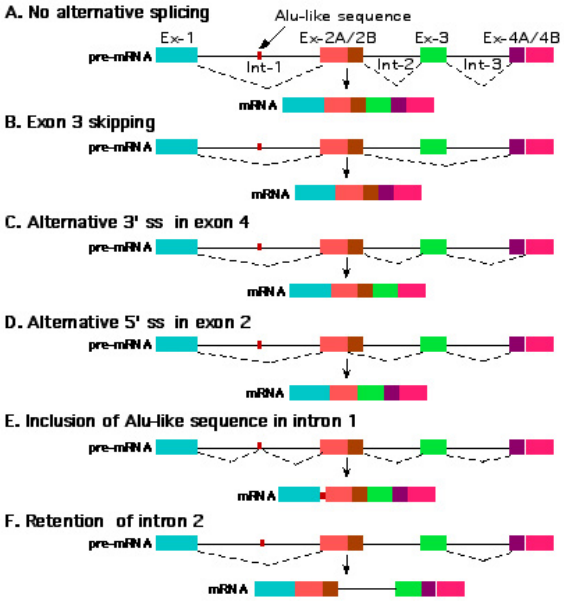

Figure 1. Types of alternative splicing of a hypothetical gene containing four exons. Exons are shown as colored boxes, whereas introns are shown as solid lines. Dotted lines indicate the joining of exonic sequences by removal of intronic sequences. A. In the absence of alternative splicing all exons are joined by removal of intronic sequences. B. Exon 3 is skipped due to absence of recognition of both, the 3′ and 5′ ss of exon 3. C. Only a portion of exon 4 is included due to activation of an alternative 3′ ss within exon 4. D. Only a portion of exon 2 is included due to activation of an alternative 5′ ss within exon 2. E. An Alu-like sequence is included due to activation of the additional 5′ and 3′ ss within intron 1. F. Intron 2 is retained due to absence of recognition of both, the 5′ ss of exon 2 and the 3′ ss of exon 3. Abbreviations: Int stands for intron, whereas ss stands for splice site.

(ESSs or ISSs). Enhancers and silencers promote or suppress splice-site selection, respectively (4). Most of our knowledge of cis-elements is limited to small sequence motifs that could be identified using various computational programs (11-13). Currently, there is no method to identify large regulatory motifs including RNA structures formed by local and long-distance intra-molecular RNA-RNA interactions. An increasing number of reports have unequivocally confirmed the very significant role of RNA structures in alternative splicing (14-16). Further adding to complexity, alternative splicing has been linked to transcription (17-19) and polyadenylation (20-22). But the most intriguing part of alternative splicing is the production of several unknown intermediates that are either degraded or not translated due to lack of initiation codon or presence of a termination codon (23-30). Considering there is also a proofreading at the second step of splicing (31-33), some of the transcripts may be eliminated after the very first step of splicing (34). Thus, it is hard to predict the fraction of pre-mRNAs that eventually make up to a translatable mRNA.

Most commonly reported alternatively-spliced transcripts include exon skipping, intron retention, insertion of Alu-like sequences, and alternate usage of the 3′ and/or 5′ ss (35-37; Figure 1). In some cases, majority of the above-mentioned variants are present in mRNAs derived from a single gene (38, 39). Alternative splicing can be induced by inherited or acquired mutations. In fact, it is hard to find a single major human disease (genetic or acquired) that is not associated with alternative splicing. These include anxiety (40), Alzheimer's disease (41), cancer (42), cardiovascular disease (43), diabetes (44), neuromuscular disease (45, 46), obesity (47-49), and Parkinson's disease (41, 50). As the number of patients affected by alternative splicing rises, there will be pressing need to find novel therapies for alternative splicing-associated human diseases.

As for now, we are confronted with the challenge of understanding the very mechanism by which exon-intron boundaries are accurately defined. One thing is clear; every single exon is regulated by a distinct rule. The only common principle is the ability of the cellular machinery to adopt a rule of combinatorial control in which decision to include or exclude an exon is made by multiple cis-elements and their cognate transacting factors. In this process, some cis-elements may play a dominant role because of their location vis-à-vis location of other cis-elements. In a recent review, Baralle and coworkers have wisely termed the dominant cis-elements as "master checkpoints" (51). Such checkpoints may serve as the effective targets for correcting aberrant splicing in human diseases. Here I describe the lessons learned from the very first *in vivo* selection experiment performed on the entire exon 7 of human *Survival Motor Neuron 1* (*SMN1*) gene (52, 53). This method offers one of best possible alternatives to identify novel cis-elements that are currently beyond the scope of being predicted by available computational programs or otherwise.

3. *IN VIVO* SELECTION OF THE ENTIRE *SMN1* EXON 7

Humans have two copies of *SMN*: *SMN1* and *SMN2*. Because of a critical C to U mutation at position 6 (designated as C6U) in exon 7 of *SMN2*, majority of mRNAs derived from *SMN2* lack exon 7 (reviewed in 54). Hence, inability of *SMN2* to compensate for the loss of *SMN1* results into spinal muscular atrophy (SMA), a leading genetic cause of infant mortality. SMA has become a model disease in which a defective copy of the gene (*SMN2*) is almost universally present and holds the potential to be corrected by merely altering the aberrant splicing of one exon (i.e. exon 7). There have been several studies that have implicated a number of cis-elements and their cognate proteins in regulation of alternative splicing of *SMN* exon 7 (reviewed in 55). Here I describe the results of a unique experimental design of *in vivo* selection that revealed novel cis-elements without their cognate protein factors being identified (52, 53).

3.1. Experimental design

In vivo selection in this review refers to iterative selection or cycled selection of exonic sequences that promote their own inclusion during the process of pre-

mRNA splicing in an intact cell. In contrast to *in vitro* selection, *in vivo* selection recapitulates a true process of molecular evolution in which selected sequences undergo through several quality controls in addition to their ability to be included in the processed transcript. For example, sequences that undergo through nonsense-mediated decay are not selected. Similarly, sequences that interact with the insoluble aggregates of the cell are eliminated.

In vivo selection method used for *SMN1* exon 7 analysis is an advanced version of iterative *in vitro* selection method "SELEX" (Systematic Evolution of Ligands through EXponential Enrichment) that was originally developed to isolate tight binding nucleic acid ligands called "aptamers" (56-58). Most *in vitro* SELEX experiments have been performed using ~10^{15} molecules generated by complete randomization of small sequences. Because complete randomization creates artificial binding motifs, aptamer-protein interactions show no resemblance with the RNA-protein interactions prevalent in nature. Hence, not surprisingly aptamers have been generated against proteins lacking nucleic acid binding domains (56). To suppress the creation of artificial binding motifs, many modified SELEX experiments have been performed using partially random sequences. These experiments have revealed valuable and rather critical features of RNA-protein interactions prevalent in nature (59-63). In contrast to complete randomization, partial randomization allows examination of larger sequences, while maintaining the wild-type characteristics of the molecule. Except for the limitations of oligonucleotide synthesis, there is no limit on the size of molecules that could be analyzed using partial randomization.

Since *SMN* exon 7 is only 54-nucleotide long, it presented an ideal size to test the very first *in vivo* selection of an entire exon. Partial randomization was used to create the initial pool of ~10^{11} unique molecules (52). Limitations of *in vivo* experiment do not allow analysis of a larger pool. However, a pool size of ~10^{10} unique molecules is enough to test the position-specific role of every nucleotide within the entire exon 7. We used 30% randomization by doping 10% each of the non-wild-type residues at every exonic position. Traditionally, this degree of randomization has been successfully used to obtain valuable information (59-63). The commercially synthesized partially random exon 7 sequences were inserted in *SMN1* minigene and the percentage of randomization was confirmed by sequencing of 51 individual clones from the initial pool. It should be mentioned that the sequencing of the initial pool is necessary to normalize the skewed randomization during commercial synthesis (52). Of note that the commercial synthesis of the partially random oligonucleotides may be the only expensive step of the selection protocol, however cost of oligonucleotide synthesis is continuously going down.

For iterative selection, cells were first transfected with exon-7-randomized *SMN1* minigene and specific primers were used to amplify exon-7-included transcripts. The amplified exon-7-included products were purified and reamplified with specific primers that allowed restoration of the minigene (Figure 2). The success of *in vivo* selection experiment was apparent from the very first cycle. For example, there was no smear during the amplification of the spliced products (52). A complete randomization would have created cryptic-splice sites resulting into non-amplifiable bands. After only four cycles substantial enrichment was achieved. To further confirm the enrichment, a total of 59 clones from the 4th cycle were analyzed. As expected, none of the selected clones showed exon-7-excluded products in the highly sensitive radioactive RT-PCR (52).

3.2. Analysis of results and validation

Having obtained the required pool of sequences that promote exon 7 inclusion, a very strict criteria was applied to determine the position-specific role of residues. For this purpose, all sequences were aligned end-to-end. This approach is distinct from those in which sequences are aligned to obtain small motifs (64, 65). Upon end-to-end alignment of sequences, exonic positions that were found to retain the wild-type residues were considered conserved. Similarly, exonic positions that were found to discard the wild-type residues were considered mutable. A conserved position is supposed to have positive impact on exon 7 inclusion, whereas a mutable position is likely to have the negative impact on exon 7 inclusion. The concept of mutability was originally applied in 1995 to find the important residues within a protein (66). Since then, this approach has been successfully tested in other systems (62, 67-69).

3.2.1. Mutability profile reveals three novel cis-elements

Based on the mutability of residues, exon 7 was found to contain two negative and one positive element (Figure 3). Negative elements were termed as **Exinct** (stands for "**Ex**tended **in**hibitory **c**onte**xt**") and **3´-Cluster** located towards the 5´ and 3´ ends of exon 7, respectively. The lone positive element was termed as **Conserved tract** located in the middle of exon 7. Nature of above elements was validated by site-specific mutations (52, 70). Both Exinct and Conserved tract are large sequence motifs. Such motifs are currently not predicted by computational programs. It is possible that a large sequence motif provides a higher degree of specificity due to interaction with a larger protein complex. In fact most splicing proteins (particularly SR and SR-like proteins) are capable of forming large complexes by interacting among each other before interaction with their cognate cis-elements on a given pre-mRNA. It is well known that a multi-component protein complex may have different specificity to a RNA sequence than the individual components of the complex. Consistently, Gold and colleagues have demonstrated that even a slight change in stoichiometry of two proteins in a complex is sufficient to produce different sequence motifs in an *in vitro* SELEX experiment (71). Extending this principle to pre-mRNA splicing, Singh and Vacarcel have recently reasoned that alternative splicing is regulated by specific arrangement of proteins that are generally non-specific on their own (6). Hence, it is rather reassuring that results of *in vivo* selection revealed larger sequence motifs that may be unique to a particular exon.

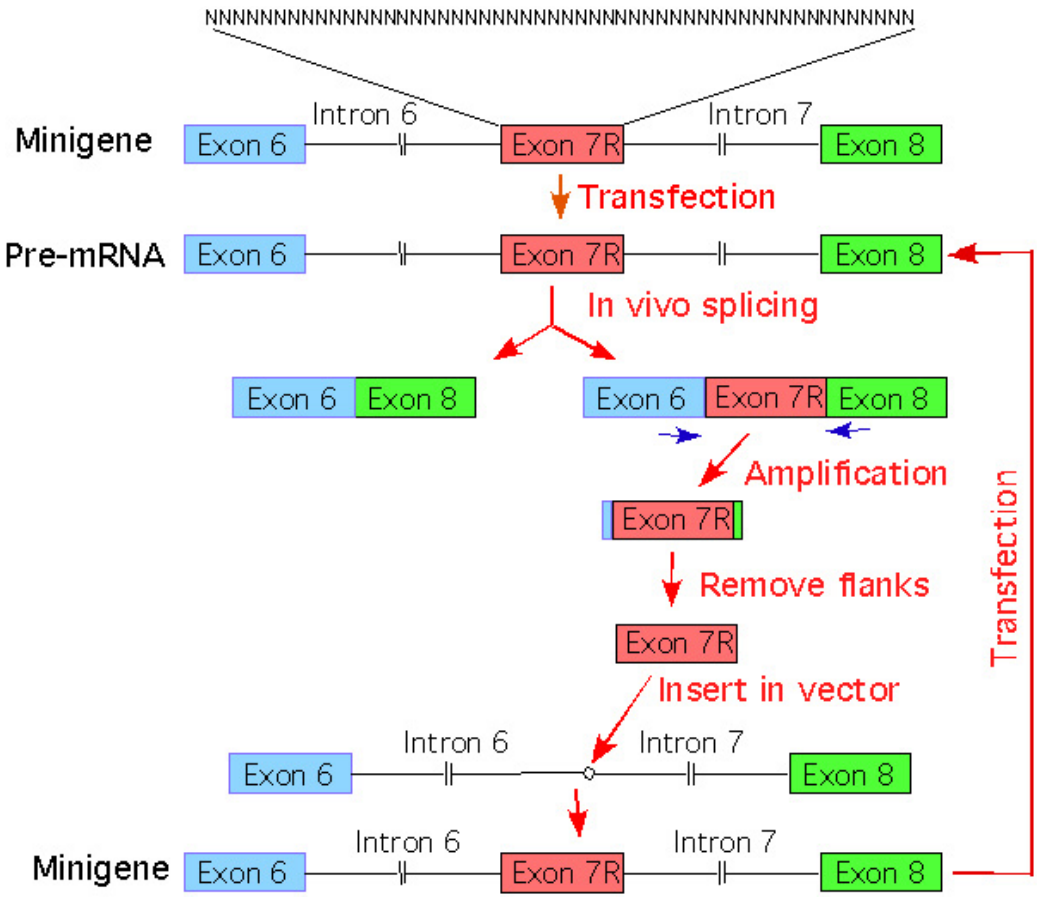

Figure 2. Strategy for the iterative *in vivo* selection of the entire *SMN1* exon 7. Exons are shown as colored boxes, whereas introns are shown as solid lines (diagram not to the scale). *SMN1* minigene containing exon 6, short intron 6, partially-random exon 7 (Exon 7R), intron 7 and exon 8. Ns represent all positions of exon 7 that were partially randomized. Details of minigene construction are described elsewhere (52). Cells were first transfected with minigene containing Exon 7R. After *in vivo* splicing, exon-7 included products were amplified by RT-PCR, separated on gel and purified. A second amplification step was used to amplify exon 7 sequences with flanks containing special restriction sites. Upon digestion at the special restriction site, intact E7R was released and inserted in the original vector to restore the minigene (52). Repeated steps of selection procedure amplified exon 7 sequences that promote their own inclusion (52).

Another valuable outcome of *in vivo* selection was the identification of the variable impact of residues within a given cis-element. For example, not all residues within (Exinct) were inhibitory, whereas, not all residues within the conserved tract were stimulatory. Such residues are referred to as incompatible residues. Due to several compelling reasons, presence of an incompatible residue within a cis-element could have been preferred. For example, incompatible residues may have avoided creation of tight binding sites for protein complexes. It is not in the interest of a dynamic and efficient process of cellular regulation to have a tight-bound complex in which protein components are trapped. On the other hand a loose complex would allow protein components to be recycled multiple times. Further, role of an incompatible residue could be associated with providing secondary contacts for proteins that interact with other motifs. It is also possible that incompatible residues are important for providing the stability to the transcripts, as well as may help in transport of transcripts across the nuclear membrane.

3.2.2. Role of terminal exonic positions on exon skipping

Terminal gaunosine residues play an important role in exon definition and are predominantly represented in internal exons (72). Yet, plethora of exons without terminal gaunosines is efficiently included in the processed transcripts. Hence it is difficult to predict the exact contribution of terminal residues in defining an exon. *In vivo* selection of entire *SMN1* exon 7 provided the first opportunity in which relative significance of both terminal positions were tested. Underscoring the significance of terminal guanosines, the wild-type guanosine at the first position (1G) was conserved, and a non-wild-type guanosine at the last position (54G) was overwhelming selected. Additional experiments confirmed that 1G is absolutely required to include exon 7 in *SMN1*. However, effect of 54G was found to be more dominating than 1G. For instance, inhibitory effect of G1H mutation (H stands for A or C or U) was easily compensated by 54G. Further experiments showed that 54G also compensates for the loss

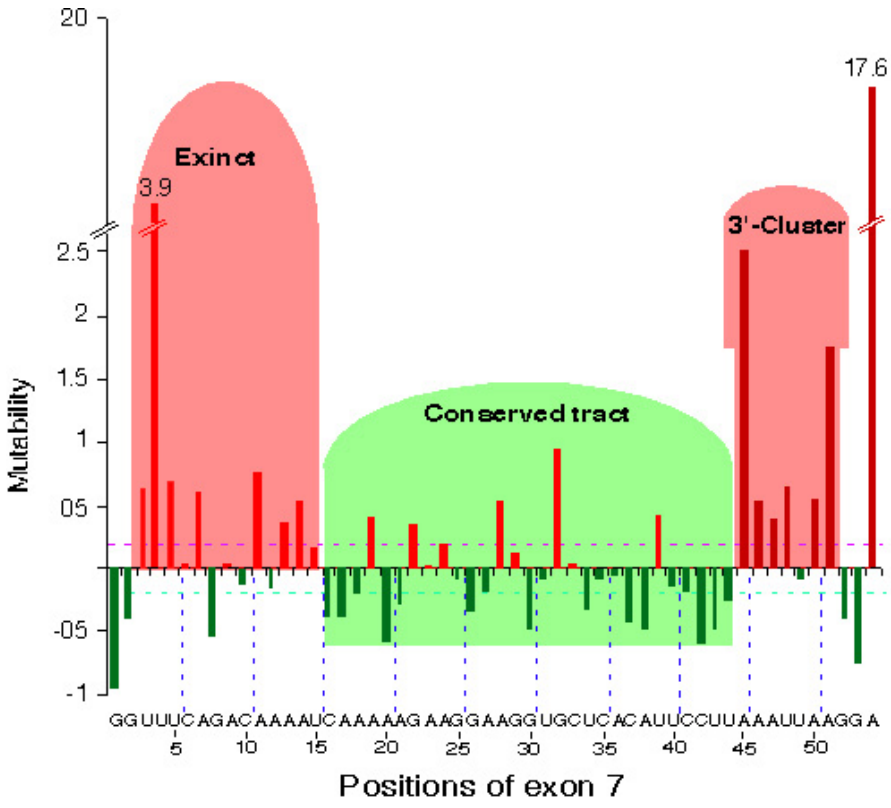

Figure 3. Mutability of residues based on the results of *in vivo* selection. The negative bars (in green) represent conserved positions, whereas the positive bars (in red) represent the mutable position. Mutable and conserved positions play negative and positive roles in exon inclusion, respectively. 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, two negative elements (Exinct and 3′-Cluster) and one positive element (Conserved tract) have been shown. Nature of these elements was confirmed by site-specific mutagenesis (52, 70). The exceptionally high mutability of position 54 is consistent with the dominant effect of A54G substitution on exon 7 inclusion (52). Interestingly, 3′-Cluster overlaps with TSL2, a terminal stem-loop structure that was recently found to play an inhibitory role (16; Figure 4).

of an ESE associated with Tra2-β1. In fact 54G was able to compensate for the simultaneous loss of two cis-elements, Tra2-β1 ESE and 1G (52). These results are in full agreement with the results of *in vivo* selection that showed that the last position of exon 7 is the-most-inhibitory position (or highly mutable position) within the entire exon 7. It must be mentioned that finding of 54G as the most powerful position within exon 7 is a major discovery thanks to *in vivo* selection of entire exon 7. This very fact has been confirmed in a recent report, which compared the results of *in vivo* selection with available computational programs (51). Among various possible effects, it is likely that 54G breaks an inhibitory context created by a RNA structure at the 5′ ss of exon 7 (16). However, the exact mechanism by which 54G has such a dominant impact is beyond the scope of this review.

3.2.3. Relative impacts of weak splice sites on exon skipping

As mentioned above, C6U is the critical difference between exon 7 of *SMN1* and *SMN2* (also see references in 54). Prior to *in vivo* selection of the entire exon 7, all published reports focused exclusively on the weak 3′ ss as the major cause of exon 7 skipping in human *SMN2* (cited in 16, 54). Only after realizing the fact that 54G compensates for the loss of a number of positive elements, it became obvious that the 5′ ss of exon 7 is extremely weak. Surprisingly, existing computational programs show their strong limitation as they fail to detect any significant difference between 54G and 54A on the strength of splice sites (51). Thus *in vivo* selection offers a strong alternative to the existing computational programs for predicting the relative strength of splice sites of a skipping exon. In addition to the presence 54A, other cis-elements also contribute towards the weak 5′ ss of exon 7. The first reported example is the 3′-Cluster, a major cis-element revealed by *in vivo* selection of entire exon 7. Most recently, an intronic element, ISS-N1, was also found to contribute towards the weak 5′ ss of exon 7 (73). Noticeably, 3′-Cluster and ISS-N1 are not conserved between human and mouse, suggesting an evolutionary significance. It will not be surprising if factors interacting with 3′-Cluster also communicate with ISS-N1 or vice versa. Of note that discovery of ISS-N1 is a major step

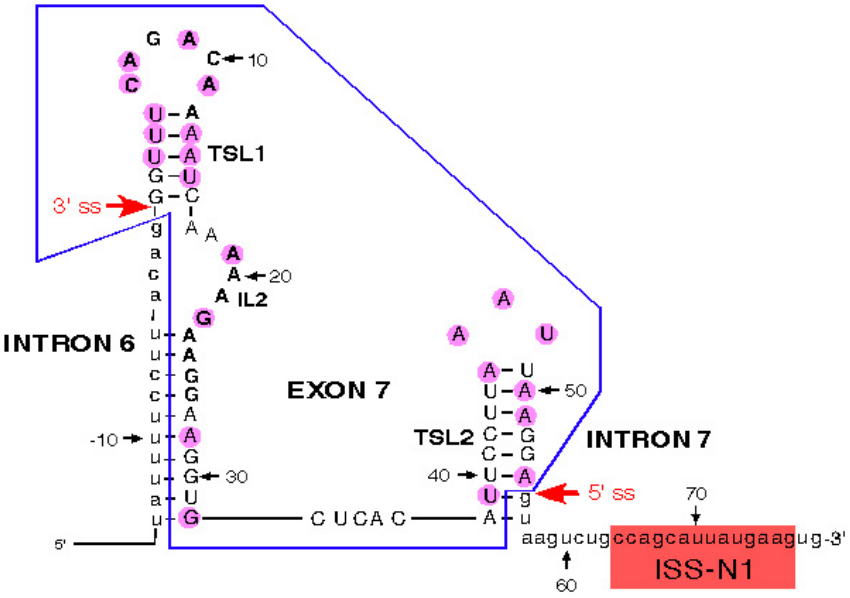

Figure 4. RNA secondary structure of human *SMN1* exon 7. The same structure holds true for human *SMN2* exon 7 (except for C6U transition, a C to U mutation at position 6 in exon 7 of *SMN2*; also see reference 54). Numbering starts from exon 7 (the entire exon 7 is boxed). Intron 7 starts from position 55. Negative numbers represent intron 6. The exonic sequences are shown in large-case letters, whereas intronic sequences are shown in small-case letters. The local secondary structure of exon 7 RNA consists of two terminal-stem loops (TSL1 and TSL2) and one internal-loop (IL2). TSL1 is exclusively formed by exonic sequences. TSL2 is mostly formed by exon 7 sequences, although the last two base pairs of the TSL2 are formed by sequences of exon 7 and intron 7. Shaded residues represent mutable positions determined by *in vivo* selection of entire exon 7. A recent report validated the above structure by enzymatic probing (16). This report also confirmed the inhibitory nature of TSL2. Positions of the 3´ ss, 5´ ss and ISS-N1 (a negative element) are shown. Details of ISS-N1 are described elsewhere (73).

forward for our understanding of *SMN* exon 7 splicing. Antisense oligonucleotides that block ISS-N1 were found to restore exon 7 inclusion in *SMN2* mRNA (73). Hence, ISS-N1 provides the very first gene-specific target for correction of *SMN2* exon 7 splicing. ISS-N1 also serves as one of the classic examples of correcting exonic defects through use of unique intronic sequences. It should be mentioned that *in vivo* selection of entire exon 7 was one of the crucial steps that led to an early discovery of ISS-N1.

3.2.4. Role of RNA structure on exon skipping

The very first application of the partially random large sequence in an iterative selection experiment gained significance due to identification of a unique structure involving non-canonical interaction (59). Further use of the partially random large sequences in the iterative selection experiments confirmed the role of other critical structural motifs (60, 61, 63). RNA structure remains one of the most challenging aspects of alternative splicing. Single mutations that create/abrogate a cis-element could also bring a drastic change in RNA structures. But, delineating the role of RNA structure using site-specific mutations is a daunting task. Remarkably, mutability values obtained after *in vivo* selection provided a strong clue about the inhibitory roles played by local RNA structures. For example, two terminal stem-loop structures, TSL1 and TSL2, formed by exon 7 were found to harbor the majority of highly mutable residues (Figure 4). Most of these mutations are capable of breaking TSL1 and TSL2. Indeed, a recent structural analysis by enzymatic probing confirmed the existence of TSL1 and TSL2 (16). Further, a systematic analysis of residues within TSL2 confirmed the inhibitory nature of TSL2 (16). Most convincing among them were compensatory mutations that restored exon 7 exclusion by reinstating the inhibitory structure, TSL2. Having proved its use in revealing critical role of RNA structure, *in vivo* selection approach could be applied to large exonic sequences. With proper modifications, method could also be applied to structures that involve base pairing between exonic and intronic sequences.

4. CONCLUSIONS AND PERSPECTIVES

Alternative splicing is an essential process of gene regulation in higher eukaryotes. This process is also associated with a growing number of human and animal diseases. Methods are being developed to identify critical cis-elements that regulate alternative splicing of a given exon. A proper understanding of splicing cis-elements has multiple ramifications. Most important among them is the use of cis-elements for correction of the aberrantly spliced genes. *In vivo* selection of the entire exon 7 of *SMN1* is the first successful *in vivo* application of a very promising and unbiased methodology. Not only this method was capable of identifying novel regulatory elements, but was able to reveal critical structural motifs, identify prominent role of certain key positions within the entire exon and disclose the relative strength of splice sites. There is no report of any *in vivo* selection study in which such a wealth of information was generated in a single experiment.

Success of *in vivo* selection method could be largely attributed to the partial randomization. This approach maintained the wild-type character of the molecule, as well as tested the relative significance of every exonic position. Hence, identification of cis-elements was made in the context of the entire exon. Method provides flexibility in choosing the degree of randomization for examination of specific region of an exon. For example, a higher degree of randomization could be used to check the more conserved regions, whereas a lesser degree of randomization could be used to check the highly mutable regions. This way, significance of every single position within an exon could be tested. In any case, information generated from *in vivo* selection of large sequences would be of great help for future generations of computational programs.

Now that the concept of *in vivo* selection of an entire exon has been validated, experiments could be designed to analyze complex exons. Such exons are generated by incorporation of different portions of the adjoining intronic sequences. Actually, number of complex exons that contain multiple splice sites within a single exon is continuing to grow. Alternative splicing of mu opioid receptors is one of the classic examples in which addition of adjoining intronic sequences could produce up to five different variants of exon 5 (38). Incorporation of different portions of the last intron creates alpha and beta isoforms of glucocorticoid receptors (GRs) (74). Since more than 10% genes in humans are affected by GRs (74), it becomes rather essential to understand the critical role of cis-elements that determine the very ratio of alpha and beta isoforms of GRs. A recent study confirmed that different antidepressants have variable effects on the expression of alpha and beta isoforms of GRs (75). It is likely that certain cis-elements become more responsive to treatment of a particular drug. *In vivo* selection of large sequences has a great potential to address these issues. Another example of alternative splicing involving incorporation of different portions of the last intron is myelin oligodendrocyte glycoprotein (MOG) (39). Interestingly, one of the MOG isoforms is specific to primates (39). Conditions under which these isoforms are generated are not known. Here again, *in vivo* selection of large exonic sequences will be able to reveal critical cis-elements that promote inclusion of the last exon.

Several genes contain more than one alternatively spliced exon. Examples of these include but not limited to, B cell receptors, cystic fibrosis transmembrane receptor (CFTR), dopamine receptors, iodothyronine deiodinase (DIO1), LDL receptors, metabotropic glutamate receptors, MOG, mu opioid receptor, neurofibromatosis protein (NF1), phosphodiesterase (PDE), SECIS binding protein 2 (SBP2), Tau, Toll-like receptors, Tra2 and vascular endothelial growth factor (VEGF). There is a strong likelihood that alternative splicing of one exon affects the alternative splicing of another exon. However, there is no systematic study to this effect. *In vivo* selection method coupled with partial randomization of exonic sequences would undoubtedly provide an unparallel approach to analyze alternative splicing of multiple exons in a single experiment. In fact such experiments would revolutionize the process of target discovery for correction of a number of splicing-associated diseases.

Pre-mRNA splicing is coupled with other cellular processes *in vivo*. These include transcription, 5′-capping, polyadenylation, exon-junction complex (EJC) formation, mRNA transport, nonsense-medicated decay (NMD), nonsense-altered splicing (NAS) and translation (see references in 52). Additionally, it is likely that RNA structure affects most of the above processes. Unfortunately, there is no method that could address the interrelationship between pre-mRNA splicing and these processes. *In vivo* selection method could be tailored to ask specific questions. Since method using partial randomization allows analysis of large sequences without creating artificial motifs, it is highly feasible to test the role of multiple overlapping wild type motifs (e.g. associated with alternative splicing, NMD, NAS, EJC formation and transport) in the context of a particular gene.

In summary, there is a paucity of methods that reveal complex biological motifs of fundamental significance. *In vivo* selection method described here offers several advantages over existing methodologies and has a huge potential to advance our understanding of posttranscriptional regulation. The method will help find key regulatory elements that have profound impact on alternative splicing of genes associated with a growing number of human diseases. Once the key regulatory elements are known, different strategies could be adopted to target these elements for producing the needed protein isoforms. For example, the negative cis-elements could be blocked by antisense oligonucleotides to promote exon inclusion in a given mRNA. Similarly, to promote exon exclusion, positive cis-elements could be targeted. Structure-specific aptamers could be used to target critical structural motifs. Aptamers are emerging as the powerful tool of target validation and therapy (57, 58). To an equal significance, *in vivo* selection method described here will address several fascinating questions of biological evolution. After all, *in vivo* selection is an unbiased process of biological evolution with the advantage of receiving very specific answers in a time and manner of our own choice.

5. ACKNOWLEDGEMENTS

Author would like to acknowledge the generous support of Families of SMA-USA (www.fsma.org), without which the very first *in vivo* selection of the entire *SMN* exon 7 would not have been possible. Author is currently supported by grants from National Institutes of Health-USA (R21 NS055149 and R01 NS055925) and Muscular Dystrophy Association-USA (MDA3969).

6. REFERENCES

1. Nilsen T. W.: The spliceosome: the most complex macromolecular machine in the cell? *Bioessays* 25, 1147-1149 (2003)

2. Graveley B. R.: Sorting out the complexity of SR protein

functions. RNA 6, 1197-1211 (2000)

3. Hastings M. L. & A. R. Krainer: Pre-mRNA splicing in the new millennium. *Curr Opin Cell Biol* 13, 302-309 (2001)

4. Maniatis T. & B. Tasic: Alternative pre-mRNA splicing and proteome expansion in metazoans. Nature 418, 236-243 (2002)

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

6. Singh R. & J. Valcarcel: Building specificity with nonspecific RNA-binding proteins. *Nat Struct Mol Biol* 12(8), 645-653 (2005)

7. Soret J., M. Gabut & J. Tazi: SR proteins as potential targets for therapy. *Prog Mol Subcell Biol* 44, 65-87 (2006)

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

9. Cooper T. A.: Alternative splicing regulation impacts heart development. *Cell* 120(1), 1-2 (2005)

10. Blencowe B. J.: Alternative splicing: new insights from global analyses. *Cell* 126(1), 37-47 (2006)

11. Fairbrother W. G., Yeo G. W., Yeh R., Goldstein P., Mawson M., P.A. Sharp & C.B. Burge: RESCUE-ESE identifies candidate exonic splicing enhancers in vertebrate exons. *Nucleic Acids Res* 32, W187-90 (2004)

12. Wang Z., Rolish M. E., Yeo G., Tung V., M. Mawson & C. B. Burge: Systematic identification and analysis of exonic splicing silencers. *Cell* 119(6), 831-845 (2004)

13. Smith P. J., Zhang C., Wang J., Chew S. L., M.Q. Zhang & A. R. Krainer: An increased specificity score matrix for the prediction of SF2/ASF-specific exonic splicing enhancers. *Hum Mol Genet* 15(16), 2490-2508 (2006)

14. Buratti E. & F. E. Baralle: (2004) Influence of RNA secondary structure on the pre-mRNA splicing process. *Mo. Cell Biol* 24,10505-10514 (2004)

15. Graveley B. R.: (2005) Mutually exclusive splicing of the insect *Dscam* pre-mRNA directed by competing intronic RNA secondary structures. *Cell* 123, 65-73 (2005)

16. Singh N. N., R. N. Singh & E. J. Androphy: Modulating role of RNA structure in alternative splicing of a critical exon in the spinal muscular atrophy genes. *Nucleic Acids Res* In Press. (2007) doi:10.1093/nar/gkl1050

17. Bentley, D.L.: Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors. *Curr Opin Cell Biol* 17, 251-256 (2005)

18. Masuda S., Das R., Cheng H., Hurt E., N. Dorman & R. Reed: Recruitment of the human TREX complex to mRNA during splicing. *Genes Dev* 19(13), 1512-1517 (2005)

19. Fededa J.P., Petrillo E., Gelfand M.S., Neverov A.D., Kadener S., Nogues G., Pelisch F., Baralle F.E., A.F. Muro & A.R. Kornblihtt: A polar mechanism coordinates different regions of alternative splicing within a single gene. *Mol Cell* 19(3), 393-404 (2005)

20. Castelo-Branco P., Furger A., Wollerton M., Smith C., A. Moreira & N. Proudfoot: Polypyrimidine tract binding protein modulates efficiency of polyadenylation. *Mol Cell Biol* 24(10), 4174-4183 (2004)

21. Gawande B., Robida M.D., A. Rahn, & R. Singh: Drosophila Sex-lethal protein mediates polyadenylation switching in the female germline. *EMBO J* 25(6), 1263-1272 (2006)

22. Kyburz A., Friedlein A., H. Langen & W. Keller: Direct interactions between subunits of CPSF and the U2 snRNP contribute to the coupling of pre-mRNA 3' end processing and splicing. *Mol Cell* 23(2), 195-205 (2006)

23. Xing Y. & C. J. Lee: Negative selection pressure against premature protein truncation is reduced by alternative splicing and diploidy. *Trends Genet* 20(10), 472-475 (2004)

24. Buhler M. & O. Muhlemann: Alternative splicing induced by nonsense mutations in the immunoglobulin mu VDJ exon is independent of truncation of the open reading frame. *RNA* 11(2), 139-146 (2005)

25. Stamm S., Ben-Ari S., Rafalska I., Tang Y., Zhang Z., Toiber D., T. A. Thanaraj & H. Soreq: Function of alternative splicing. *Gene* 344, 1-20 (2005)

26. Lejeune F. & L. E. Maquat: Mechanistic links between nonsense-mediated mRNA decay and pre-mRNA splicing in mammalian cells. *Curr Opin Cell Biol* 17(3), 309-315 (2005)

27. Matlin A. J., F. Clark & C. W. Smith: Understanding alternative splicing: towards a cellular code. Nat *Rev Mol Cell Biol* 6(5), 386-398 (2005)

28. Magen A. & G. Ast: The importance of being divisible by three in alternative splicing. *Nucleic Acids Res* 33(17), 5574-5582 (2005)

29. Hyvonen M. T., Uimari A., Keinanen T. A., Heikkinen S., Pellinen R., Wahlfors T., Korhonen A., Narvanen A., Wahlfors J., L. Alhonen & J. Janne: Polyamine-regulated unproductive splicing and translation of spermidine/spermine N1-acetyltransferase. *RNA* 12(8), 1569-1582 (2006)

30. Grellscheid S. N. & C. W. Smith: An apparent pseudo-exon acts both as an alternative exon that leads to nonsense-mediated decay and as a zero-length exon. *Mol Cell Biol* 26(6), 2237-2246 (2006)

31. Chua K. & R. Reed: The RNA splicing factor hSlu7 is required for correct 3' splice-site choice. *Nature* 402(6758), 207-210 (1999)

32. Lallena M. J., Chalmers K.J., Llamazares S., A.I. Lamond & Valcarcel J. Splicing regulation at the second catalytic step by Sex-lethal involves 3' splice site recognition by SPF45. *Cell* 109(3), 285-296 (2002)

33. Soares L. M., Zanier K., Mackereth C., M. Sattler & J. Valcarcel: Intron removal requires proofreading of U2AF/3' splice site recognition by DEK. *Science* 312(5782), 1961-1965 (2006)

34. Hilleren P. J. & R. Parker: Cytoplasmic degradation of splice-defective pre-mRNAs and intermediates. *Mol Cell* 12(6), 1453-1465 (2003)

35. Pagani F. & F. E. Baralle: Genomic variants in exons and introns: identifying the splicing spoilers. *Nat Rev Genet* 5(5), 389-396 (2004)

36. Lei H. & I. Vorechovsky: Identification of splicing silencers and enhancers in sense Alus: a role for pseudoacceptors in splice site repression. *Mol Cell Biol* 25(16), 6912-6920 (2005)

37. Ranum L. P. & T.A. Cooper: RNA-mediated neuromuscular disorders. *Annu Rev Neurosci* 29, 259-277 (2006)

38. Pan Y. X.: Diversity and complexity of the mu opioid receptor gene: alternative pre-mRNA splicing and promoters. *DNA Cell Biol* 24(11), 736-750 (2005)

39. Delarasse C., Della Gaspera B., Lu C. W., Lachapelle F., Gelot A., Rodriguez D., Dautigny A., C. Genain & D. Pham-Dinh: Complex alternative splicing of the myelin oligodendrocyte glycoprotein gene is unique to human and non-human primates. *J Neurochem* 98(6), 1707-1717 (2006)

40. Battaglia M. & A. Ogliari: Anxiety and panic: from human studies to animal research and back. *Neurosci Biobehav Rev* 29(1), 169-179 (2005)

41. Licatalosi D. D. & R. B. Darnell: Splicing regulation in neurologic disease. *Neuron* 52(1), 93-101 (2006)

42. Srebrow A. & A. R. Kornblihtt: The connection between splicing and cancer. *J Cell Sci* 119(13), 2635-2641 (2006)

43. Liao P., Yong T. F., Liang M. C., D. T. Yue & T. W. Soong: Splicing for alternative structures of Cav1.2 Ca2+ channels in cardiac and smooth muscles. *Cardiovasc Res* 68(2), 197-203 (2005)

44. Yonekura H., Yamamoto Y., Sakurai S., T. Watanabe & H. Yamamoto: Roles of the receptor for advanced glycation endproducts in diabetes-induced vascular injury. *J Pharmacol Sci* 97(3), 305-311 (2005)

45. Osborne R. J. & C. A. Thornton: RNA-dominant diseases. *Hum Mol Genet* 15 (2) R162-169 (2006)

46. Meshorer E. & H. Soreq: Virtues and woes of AChE alternative splicing in stress-related neuropathologies. *Trends Neurosci* 29(4), 216-224 (2006)

47. Tansey J. T., Sztalryd C., Hlavin E. M., A. R. Kimmel & C. Londos: The central role of perilipin a in lipid metabolism and adipocyte lipolysis. *IUBMB Life* 56(7), 379-385 (2004)

48. Minn A. H., Lan H., Rabaglia M. E., Harlan D. M., Peculis B. A., A. D. Attie & A. Shalev: Increased insulin translation from an insulin splice-variant overexpressed in diabetes, obesity, and insulin resistance. *Mol Endocrinol* 19(3), 794-803 (2005)

49. Chen M., Gavrilova O., Liu J., Xie T., Deng C., Nguyen A. T., Nackers L. M., Lorenzo J., L. Shen & L.S. Weinstein: Alternative Gnas gene products have opposite effects on glucose and lipid metabolism. *Proc Natl Acad Sci USA* 102(20), 7386-7391 (2005)

50. Ravindranath V., R. P. Kommaddi & H. V. Pai: Unique cytochromes P450 in human brain: implication in disease pathogenesis. *J Neural Transm Suppl* (70), 167-171 (2006)

51. Buratti E., M. Baralle & F. E. Baralle: Defective splicing, disease and therapy: searching for master checkpoints in exon definition. *Nucleic Acids Res* 34(12), 3494-3510 (2006)

52. Singh N. N., Androphy E. J. & R. N. Singh: *In vivo* selection reveals combinatorial controls that define a critical exon in the spinal muscular atrophy genes. *RNA* 10(8), 1291-1305.

53. Buratti E. & F. E. Baralle: Another step forward for SELEXive splicing. *Trends Mol Med* 11(1), 5-9 (2005)

54. Singh N. N., Androphy E. J. & R. N. Singh: The regulation and regulatory activities of alternative splicing of the SMN gene. *Crit Rev Eukaryot Gene Expr* 14(4), 271-285 (2004)

55. Wirth B., L. Brichta & E. Hahnen: Spinal muscular atrophy and therapeutic prospects. *Prog Mol Subcell Biol* 44, 109-132 (2006)

56. Daniels D.A., Chen H., Hicke B. J., K. M. Swiderek & L. Gold: A tenascin-C aptamer identified by tumor cell SELEX: systematic evolution of ligands by exponential enrichment. *Proc Natl Acad Sci USA* 100(26), 15416-15421 (2003)

57. Blank M. & M. Blind: Aptamers as tools for target validation. *Curr Opin Chem Biol* 9(4), 336-342 (2005)

58. Lee J. F., G. M. Stovall & A. D. Ellington: Aptamer therapeutics advance. *Curr Opin Chem Biol* 10(3), 282-289 (2006)

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

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

61. Saldanha R., Ellington A., A.M. Lambowitz: Analysis of the CYT-18 protein binding site at the junction of stacked helices in a group I intron RNA by quantitative binding assays and *in vitro* selection. *J Mol Biol* 261, 23-42 (1996)

62. Guo H., Karberg M., Long M., Jones J. P. 3rd, B. Sullenger & A. M. Lambowitz: Group II introns designed to insert into therapeutically relevant DNA target sites in human cells. *Science* 289, 452-457 (2000)

63. Singh R. N., Saldanha R. J., L. M. D'Souza & A. M. Lambowitz: Binding of a group II intron-encoded reverse transcriptase/maturase to its high affinity intron RNA binding site involves sequence-specific recognition and autoregulates translation. *J Mol Biol 318*:287-303 (2002)

64. Coulter L. R., M. A. Landree & T. A. Cooper: Identification of a new class of exonic splicing enhancers by *in vivo* selection. *Mol Cell Biol* 17(4), 2143-2150 (1997)

65. Wu S., Romfo C. M., T. W. Nilsen & M. R. Green: Functional recognition of the 3'-splice site AG by the splicing factor $U2AF^{35}$. *Nature* 402, 832-835 (1999)

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

67. Cui X., Matsuura M., Wang Q., H. Ma & A. M. Lambowitz: A group II intron-encoded maturase functions preferentially in cis and requires both the reverse transcriptase and X domains to promote RNA splicing. *J Mol Biol* 340(2), 211-231 (2004)

68. Behrsin C. D., Brandl C. J., Litchfield D. W., B. H., Shilton & L. M. Wahl: Development of an unbiased statistical method for the analysis of unigenic evolution. *BMC Bioinformatics* 7, 150 (2006)

69. Behrsin C. D., Bailey M. L., Bateman K. S., Hamilton K. S., Wahl L. M., Brandl C. J., B. H. Shilton & D. W. Litchfield: Functionally Important Residues in the Peptidyl-prolyl Isomerase Pin1 Revealed by Unigenic Evolution. *J Mol Biol* 2006

70. Singh N. N., E. J. Androphy & R. N. Singh: An extended inhibitory context causes skipping of exon 7 of *SMN2* in spinal muscular atrophy. *Biochem Biophys Res Commun* 315, 381-388 (2004)

71. Vant-Hull B., Payano-Baez A., R. H. Davis & L. Gold: The mathematics of SELEX against complex targets. *J Mol Biol* 278(3), 579-597 (1998)

72. Burge C. B., T. Tuschl & P. A. Sharp: Splicing of precursors to mRNAs by the spliceosomes. *The RNA World*, R.F. Gesteland and J.F. Atkins, eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY), pp. 525–560 (1999)

73. Singh N. K., Singh N. N., E. J. Androphy & R. N. Singh: Splicing of a critical exon of human Survival Motor Neuron is regulated by a unique silencer element located in the last intron. *Mol Cell Biol* 26(4), 1333-1346 (2006)

74. Lu N. Z. & J. A. Cidlowski: Glucocorticoid receptor isoforms generate transcription specificity. *Trends Cell Biol* 16(6), 301-307 (2006)

75. Funato H., A. Kobayashi & Y. Watanabe: Differential effects of antidepressants on dexamethasone-induced nuclear translocation and expression of glucocorticoid receptor. *Brain Res* 1117(1), 125-134 (2006)

Key Words: Alternative splicing, in vivo selection, SELEX, SMA, SMN, EXINCT, 3´-Cluster, Conserved tract, partial randomization, RNA structure, TSL1, TSL2, Review

Send correspondence to: Dr. Ravindra N. Singh, Department of Medicine, University of Massachusetts Medical School, Worcester, MA 01605-2324, USA, Tel: 508-856-1333, Fax: 508-856-6797, E-mail: Ravindra.Singh@umassmed.edu

http://www.bioscience.org/current/vol12.htm