Another step forward for SELEXive splicing

Emanuele Buratti and Francisco E. Baralle

International Centre for Genetic Engineering and Biotechnology, Padriciano 99, Trieste 34012, Italy

Splicing alterations are being increasingly reported to cause human diseases. However, predicting beforehand whether a given mutation might lead to aberrant splicing is often hampered by insufficient knowledge of which *cis*-acting sequences affect exon recognition. To better define these sequences, experimental methods have been developed that integrate pre-mRNA splicing with sequence selection assays. Recently, a novel *in vivo* selection method based on a partially randomized full-length exon has enabled the identification of new splicing-controlling elements in *SMN* exon 7. Skipping of this exon results in an aberrant protein and is involved in proximal spinal muscular atrophy (SMA). Hopefully, this approach will provide novel targets for nascent RNA molecular medicine approaches for the recovery of abnormal pre-mRNA splicing events.

Introduction

During recent years, SELEX-based methodologies (systematic evolution of ligands by exponential enrichment) have become a useful tool for researchers working with RNA because they enable the simultaneous screening of binding affinities from diverse pools of RNAs for target sequences or other activities, such as ribozyme function or drug research [1]. In this context, the recent development of novel therapeutic strategies [2] aimed at correcting aberrant RNA processing has spurred the search for as many potential RNA target sequences as possible. The new selection method proposed by Singh *et al.* [3] investigated exon 7 of the survival of motor neuron (*SMN*) gene, a crucial exon that is involved in spinal muscular atrophy (SMA) disease.

Selection methods have often been used to identify the RNA-binding requirements of selected splicing factors or search for novel *cis*-acting motifs that are capable of regulating splicing efficiency. Although initially confined to basic research, the practical applications of these studies for molecular medicine have begun to emerge. In fact, although the number of disease-causing mutations that can be related to splicing dysfunctions is increasing [2], it is becoming more difficult for screening studies to identify which mutations might affect the splicing of a particular gene and which only represent innocuous substitutions or polymorphisms [4]. It is therefore important for future diagnostic and therapeutic procedures to identify beforehand which regions within any given exon are important for its definition. This will not only be useful for predicting or identifying which substitutions or polymorphisms might cause splicing dysfunctions but also for defining which regions might be preferentially targeted by new therapeutic approaches [2]. Here, the experimental approaches that have been used for identifying sequences essential for splicing control will be discussed.

Corresponding author: Baralle, F.E. (baralle@icgeb.org).
Available online 15 December 2004

The classical SELEX approach: selection by protein binding

The initial information regarding which exonic or intronic sequences might be essential for splicing control can be obtained by accurately mapping the binding sequences of the already known splicing factors. Using the classical selection approach, the recombinant splicing factor of choice is added to a pool of RNA molecules, each containing two fixed flanking sequences and a random central portion (Figure 1a). The pool of bound RNA molecules can then be gel purified and enriched and high-affinity binders can be isolated by performing several rounds of selection and RT–PCR amplifications. This selection approach has been used for a variety of splicing factors, with a particular emphasis towards the splicing factors that up- or down-regulate exonic inclusion. These factors include several proteins that are usually (but not exclusively) associated with enhancer activity, such as the serine–arginine rich (SR) proteins ASF/SF2 [5], SC35 [5], 9G8 [6], SRp20 [6], B52 (the *Drosophila* homologue of SRp55) [7], the two mammalian Tra2 homologues [8], the TIA-1 splicing factor [9] or the neuron-specific splicing regulator NOVA-1 [10]. In addition, several factors that are normally associated with the promotion of exon skipping have been analyzed, such as the polypyrimidine tract binding protein (PTB) [11] and heterogeneous ribonucleoprotein (hnRNP) A1 [12]. These studies, in addition to providing useful data regarding the consensus RNA binding sequences for each factor, have advanced our knowledge of the splicing pathway. For example, the finding that a component of the yeast U1 small-nuclear ribonucleoprotein (U1snRNP) complex, the U1C protein, selects a 5′ splice-site-like sequence (GUAU) in SELEX experiments [13] has suggested that the interaction between this protein and the RNA acts as a preliminary 'screening' step before commitment by duplex formation between U1snRNP and the donor site sequence.

However, a potential drawback of these assays is represented by the fact that many splicing factors lack well-defined binding consensus sequences and that functionality might not always involve very tight interactions. In addition, intrinsic limitations of these *in vitro* selection procedures [1] that involve, for example, binding kinetics or the reverse transcription

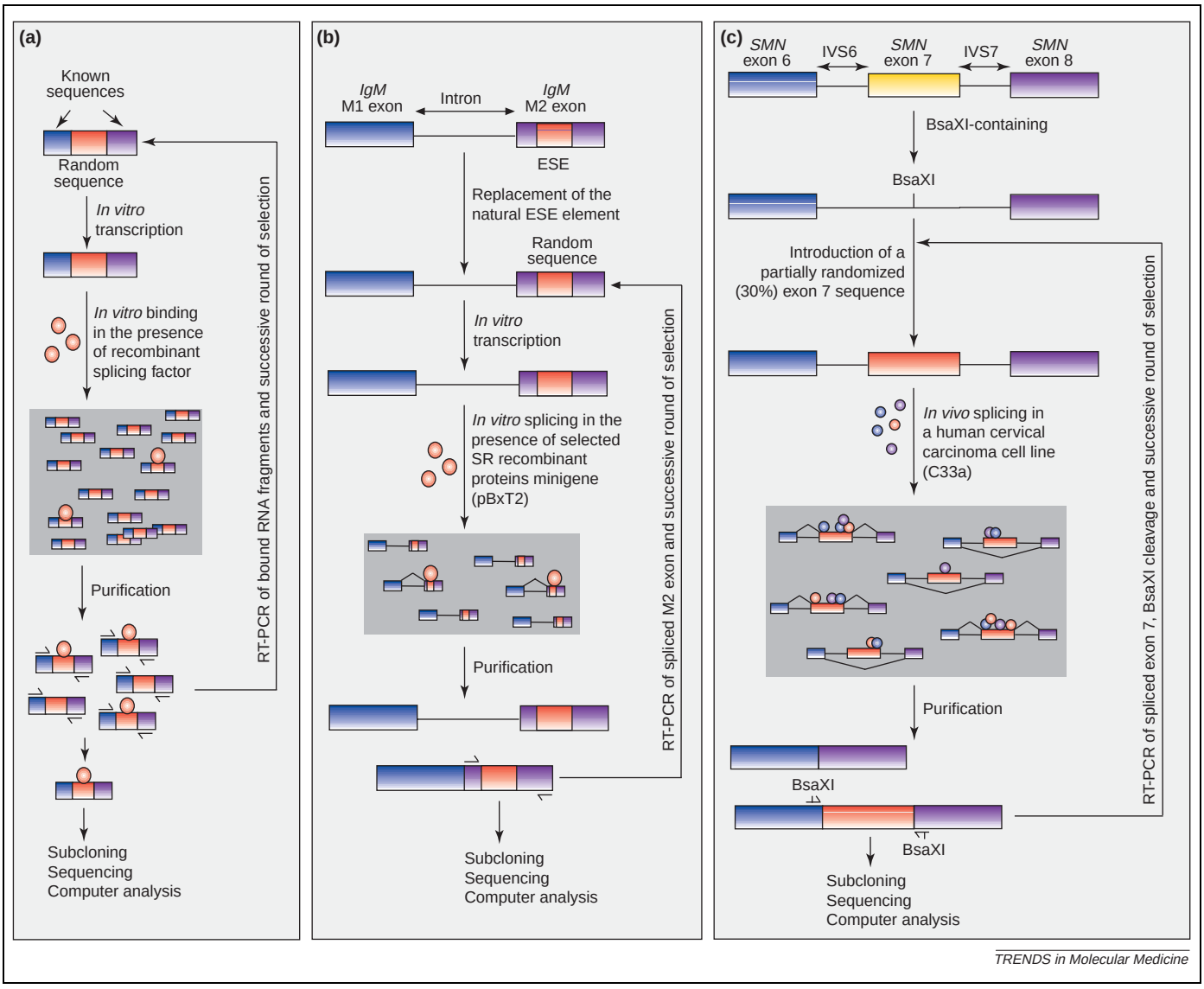

Figure 1. SELEX-based approaches applied to the splicing field. **(a)** Schematic diagram of a classic SELEX procedure to determine the RNA binding specifities of RNA binding proteins. The recombinant splicing factor is incubated with a library of short RNA molecules that contain two fixed flanking sequences and a 20–25 nucleotide-long random core, providing a potential number of 10^{14}–10^{15} different RNA molecules (i.e. potential target sites). High-affinity binders can then be identified by separating the protein–RNA bound complexes from the unbound RNA and protein pool. During each successive selection cycle, the pool of specific RNA binders is enriched by RT–PCR, taking advantage of the fixed flanking sequences to perform successive rounds of amplification. **(b)** This approach has also been combined with the splicing process to take into account functionality as well as RNA binding efficiency. Exonic splicing enhancer (ESE) sequences can be functionally investigated by coupling the SELEX basic methodology with an *in vitro* splicing RNA template. In this case, a naturally occurring ESE in exon M2 of the *IgM* gene is replaced by a random 20-nucleotide sequence. The template is then transcribed to obtain a pool of pre-mRNA molecules that are used in a standard *in vitro* splicing assay. To obtain selection specificity for a specific splicing factor, the splicing reaction can be performed in a depleted nuclear extract or S100 HeLa extract that can drive the splicing process only if complemented by a recombinant splicing factor of interest. After every round of selection, the spliced products are purified by gel electrophoresis and the M2 exon sequences that contain functional ESEs can be amplified and enriched by RT–PCR to insert them back in the original pre-mRNA template. After ~3–4 rounds of selection, the pool of enhancer sequences is used to build a scoring matrix that can be used to analyze any RNA sequence for potential high-score binding motifs. **(c)** Alternatively, RNA templates can be used in transfection assays or with a splicing-competent nuclear extract to identify novel enhancer sequences. This functional selection technique has recently been applied to whole exonic sequences. A minigene system based on the human *SMN* gene is shown, in which the entire exon 7 is partially randomized to provide ~10^{11} different exonic sequences that can be selected for inclusion in the final spliced products following several rounds of selection *in vivo* by transfecting in a human cervical carcinoma cell line. The borders of this randomized exon are broken to indicate that these sequences will not always be used as an exon in this technique. Most importantly, after each round of selection, the spliced exon 7 sequences are recovered by RT–PCR using specially designed primers that carry a BsaXI restriction enzyme site, enabling their reinsertion into the minigene splicing cassette without carrying over any undesired nucleotides from *SMN* exons 6 and 8. This selection process applied has enabled the precise mapping, at the single nucleotide level, of all the major determinants involved in the control of exon 7 splicing.

and amplification steps might lead to the over- or under-representation of selected motifs. It is presumably for these reasons that sequences belonging to naturally occurring enhancer or silencer elements do not always correspond to the optimal binding sites that have been identified using the selection techniques described in Figure 1a and that many of these splicing factors are capable of recognizing RNA sequences that deviate considerably from the optimal *in vitro* selected motifs.

Alternative approaches: coupling splicing functionality with protein binding

Attempts to obtain a closer connection between the mapping of RNA sequences capable of affecting splicing and the sequence-selection process have modified the

basic methodology to account for splicing functionality and protein binding [14–18]. Figure 1b shows a schematic of an RNA template that is used in this kind of experiment. In one of its most basic forms, it consists of two exons separated by an intronic sequence, which can be transcribed *in vitro* to obtain a pre-mRNA molecule that can undergo the splicing process. The peculiarity that sets this pre-mRNA template apart from other *in vitro* splicing templates is the fact that in one of the exonic sequences, a naturally occurring exonic splicing enhancer (ESE) sequence, is replaced by a randomized sequence (usually 20–25 nucleotides long). Following *in vitro* transcription, this arrangement yields a library of many different pre-mRNAs. This pool of pre-mRNAs can then be subjected to a splicing reaction *in vitro* and the spliced mRNA products, after purification, can be amplified by RT–PCR and rebuilt in the exon–intron–exon template for a new round of selection. The initial selection was performed by complementing a single recombinant SR protein with a HeLa cell cytoplasmic fraction (S100), which lacks all SR proteins but is otherwise competent for splicing. To minimize potential S100-specific selection bias, later selection rounds are performed using an SR-depleted nuclear extract. In this way, it has been possible to study the sequence specificity of several SR proteins, such as ASF/SF2 [17], SRp40 [17], SRp55 [17] and SC35 [18], under splicing conditions. (For a direct comparison between the high-affinity binding sites determined by classical SELEX methodologies for these SR proteins and the RNA motifs identified using functional ESE selection, see [19].) In general, however, although the motifs obtained by these two different methods share some basic similarities for each factor (such as the preference for purine-rich regions by SF2/ASF), there are also significant differences. One of the reasons for these differences might be the fact that these functional approaches are not immune to potential drawbacks. In fact, it has to be considered that adding back a single SR protein might not properly compensate for functions that might be performed by the mixture of SR proteins present in the undepleted nuclear extract.

However, these functional experiments have enabled the elaboration of binding matrices, which can then be used to search for potential SR-binding motifs in any RNA sequence of interest using a web-based application, ESE-Finder (http://rulai.cshl.edu/tools/ESE/) [20].

How reliable are web-based predictive approaches?

The reliability of web-based applications in splicing research has been tested using a well-characterized alternatively spliced exon of the human *fibronectin* gene, the EDA exon. The splicing of this exon is totally dependent on the presence of an ESE sequence (the GAAGAAGA boxed region; Figure 2), which can bind to most SR protein family members and is localized in a RNA stem-loop sequence [21]. The deletion of this region results in the complete abrogation of all SR protein binding, as determined by immunoprecipitation studies, and this leads to complete exon skipping in a minigene system [21]. In addition to the ESE-finder application, another experimental approach for predicting the occurrence of enhancer motifs [22] (but based on the statistical analysis of exonic sequences rather than selection procedures) that has recently become available as a web-based application, RESCUE-ESE (http://genes.mit.edu/burgelab/rescue-ese), was tested [23]. Both programs predict hits along the second half of the fibronectin EDA exon sequence (Figure 2). In this system, it is the RESCUE-ESE program that presents the better clustering of potential hits in the region corresponding to the GAAGAAGA sequence. However, this program does not provide any insight regarding the identity of the *trans*-acting factors that might affect that position. However, the ESE Finder application yields a more scattered pattern of hits over the whole EDA sequence but, significantly, the binding predictions for SF2/ASF, which is the SR protein with the highest stimulatory activity on the EDA exon inclusion [24], are limited to two regions, one with a fair correspondence to the GAAGAAGA sequence (Figure 2a).

Therefore, although each approach has limited predictive value when considered on its own, the combination of both gives the greatest chance of predicting real ESE sequences in the EDA system and establishing the potential identity of the *trans*-acting factors that bind in that position. Nevertheless, it should also be noted that these programs might show a variable degree of reliability according to the system analyzed. For example, it has been reported that the ESE-finder scoring motif can only partially explain the complex regulation of CFTR exon 9 splicing and, in this context, individual changes in SR binding scores cannot predict the splicing efficiency [25]. Hence, although promising, at present there is no substitute for functional experiments to validate splicing predictions. An improved knowledge of splicing mechanisms will undoubtedly make these approaches more reliable.

SELEX approaches in the search for new *cis*-acting motifs

No prior knowledge regarding which proteins are interacting with the randomized sequences is necessary to set up a functional selection assay. Consequently, similar approaches to those described for the individual SR protein factors (Figure 1b) have been performed using a set of pre-mRNA templates that carry a randomized sequence within one of the component exons [14,15,26–28]. These templates can then be subjected to rounds of splicing either *in vitro* or *in vivo* by transfection in eukaryotic cells to select novel conserved sequence families that can act as splicing modifiers. Some of the sequence families or enhancer sequences identified in this way have been shown to bind to a variety of SR protein factors, such as SRp30 [26] and SRp20 [28], whereas others remain to be characterized. Finally, an interesting development is represented by an RNA-ligase-based selection strategy to investigate donor-site functionality from a pool of randomized 5′ splice sites in the presence or absence of particular factors, such as U1snRNP [16].

Functional selection applied to whole exons

In the cases discussed above, the randomized sequences used in the selection approaches are generally set in a

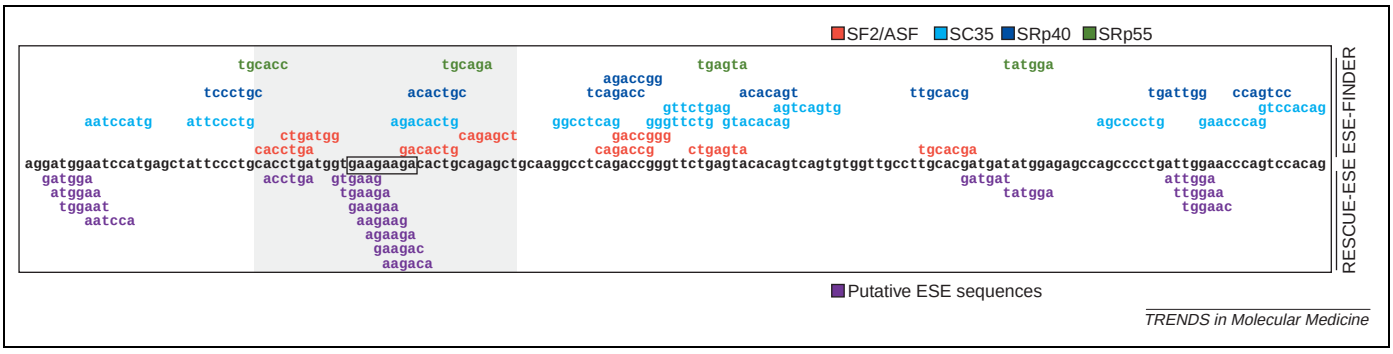

Figure 2. Estimating the reliability of web-based ESE predictors in the EDA exon of the fibronectin gene. The nucleotide sequence of the second half of the alternatively spliced fibronectin EDA exon [nucleotides (nt.) 102–270] in black with its well-characterized ESE sequence (gaagaaga, highlighted by the box). The upper lettering shows the results of an ESE-finder analysis for several SR proteins – SF2/ASF, SC35, SRp40 and SRp55 – performed on this sequence. The results correspond to all the 'hits' above the default thresholds defined for each protein. Although, within the context of each SR protein, the scores for the different hits contained some variability, there was no apparent clustering of scoring values for any particular position (data not shown). In addition, the RESCUE-ESE prediction for the identification of candidate ESE hexameric sequences are reported below the wild-type EDA sequence. The results obtained from both programs can be compared with functional analysis using immunoprecipitation and minigene studies.

heterologous context and are of limited length, never extending to whole exonic sequences. Singh and coworkers [3] have addressed this issue by devising a functional selection technique to identify the controlling elements that regulate the splicing of an entire exon: exon 7 of the *SMN* gene (Figure 1c). This exon was selected because of its key role in a common autosomal-recessive disorder, SMA. In humans, there are two copies of the *SMN* gene, *SMN1* and *SMN2*. The inactivation of the *SMN1* gene leads to disease because the *SMN2* gene cannot compensate for its absence. The reason for this is a critical C to U substitution in exon 7 of the *SMN2* gene that does not change the codon, but either disrupts an ASF/SF2 binding site [29] or creates an hnRNP A1 binding site [30]. As a result, the *SMN2* gene produces a transcript in which exon 7 is predominantly skipped, leading to the production of a truncated protein and inability to compensate for *SMN1* inactivation. The aim of the work by Singh *et al.* [3] was to develop a functional selection method that enabled the identification of all the important splicing-controlling sequences that are contained within *SMN* exon 7. This was achieved by engineering a pre-mRNA template that contained the full sequence of *SMN* exon 7 flanked by intervening sequence (IVS) 6 and IVS7 intronic sequences and *SMN* exons 6 and 8 (Figure 1c). The wild-type exon 7 sequence was then replaced by a pool of partially randomized (30%) exon 7 sequences to substantially preserve the wild-type characteristics of this exon but also enable the mutability of each nucleotide. This pool of pre-mRNA molecules was then transfected into a human cervical carcinoma cell line and the spliced products were purified to select the exon 7 sequences that could be recognised correctly by the splicing machinery. These exonic sequences were then amplified by RT–PCR using specially designed primers that enabled their reinsertion into the minigene splicing cassette without carrying over any undesired nucleotides from *SMN* exons 6 and 8. After four successive rounds of *in vivo* selection, the different exon 7 sequences were subsequently analyzed to evaluate the range and type of nucleotide substitutions that had occurred with respect to the initial wild-type sequence. The results of this analysis principally identified the weak 5′ splice site consensus sequence of wild-type exon 7 as a crucial factor in regulating *SMN* exon 7 recognition. It also confirmed the inhibitory context surrounding the first fifteen nucleotides of this exon, a region that they named Exinct (extended inhibitory context), and the splicing-enhancer role of the Tra2-β1 ESE region [31]. This result contributes to the validation of recent efforts to predict the presence of enhancer elements in exonic sequences. In fact, the computer-based RESCUE-ESE approach [22,23] indicates that these two regions (Exinct and Tra2-β1 ESE) contain many potentially splicing functional sequences. In addition, this *in vivo* selection of the entire exon identified a novel additional inhibitory region (3′-cluster) near to the 3′ end of exon 7, which remains to be characterized.

However, there are limitations to this selection process, in addition to the obvious drawback that it provides no information regarding which *trans*-acting factors are binding to the newly identified controlling regions. There is also the possibility that distinct variants might arise very early during the selection process (such as, in this case, the nearly invariable selection of G at position 54, which greatly improves the quality of the *SMN* exon-7 5′ splice site). These variants might result in different requirements for the splicing enhancers and silencers within the exon, and for this reason it might be appropriate that future applications of this technique should exclude from the randomization the obvious crucial sequences that might prove to be dominant.

Concluding remarks

The targeting of splicing-controlling regions has recently become an attractive possibility for novel molecular medicine approaches aimed at recovering splicing function. The loss of exon recognition following the C6U substitution [29] has already been the target of several strategies to recover *SMN2* function in the individuals in whom *SMN1* is inactivated [2]. The approach of Singh *et al.* [3] has provided new targets for therapeutic strategies, such as ESSENCE (exon-specific splicing enhancement by small chimeric effectors), TOES (targeted oligonucleotide enhancer of splicing) and antisense oligonucleotides, which were recently reviewed by Garcia-Blanco *et al.* [2], or by the development of specific drugs,

rybozymes or siRNA strategies to block negative-factor interactions. On a more general note, however, the greater significance of this work is to have demonstrated the feasibility of applying a functional *in vivo* selection and splicing method to entire exonic sequences through careful experimental design. Although many details remain to be investigated, these results are encouraging regarding the potential future trends of this field. There is no doubt that the greater benefit regarding the treatment of splicing defects will require the progressive integration of all the predictive and functional techniques that have recently been developed.

Acknowledgements

This work was supported by grants from the Telethon Onlus Foundation (Italy) (grant GGP02453) and FIRB (RBNE01W9PM).

References

1 Klug, S.J. and Famulok, M. (1994) All you wanted to know about SELEX. *Mol. Biol. Rep.* 20, 97–107
2 Garcia-Blanco, M.A. *et al*. (2004) Alternative splicing in disease and therapy. *Nat. Biotechnol.* 22, 535–546
3 Singh, N.N. *et al*. (2004) *In vivo* selection reveals combinatorial controls that define a critical exon in the spinal muscular atrophy genes. *RNA* 10, 1291–1305
4 Pagani, F. and Baralle, F.E. (2004) Genomic variants in exons and introns: identifying the splicing spoilers. *Nat. Rev. Genet.* 5, 389–396
5 Tacke, R. and Manley, J.L. (1995) The human splicing factors ASF/SF2 and SC35 possess distinct, functionally significant RNA binding specificities. *EMBO J.* 14, 3540–3551
6 Cavaloc, Y. *et al*. (1999) The splicing factors 9G8 and SRp20 transactivate splicing through different and specific enhancers. *RNA* 5, 468–483
7 Kim, S. *et al*. (2003) Specific SR protein-dependent splicing substrates identified through genomic SELEX. *Nucleic Acids Res.* 31, 1955–1961
8 Tacke, R. *et al*. (1998) Human Tra2 proteins are sequence-specific activators of pre-mRNA splicing. *Cell* 93, 139–148
9 Dember, L.M. *et al*. (1996) Individual RNA recognition motifs of TIA-1 and TIAR have different RNA binding specificities. *J. Biol. Chem.* 271, 2783–2788
10 Buckanovich, R.J. and Darnell, R.B. (1997) The neuronal RNA binding protein Nova-1 recognizes specific RNA targets *in vitro* and *in vivo*. *Mol. Cell. Biol.* 17, 3194–3201
11 Perez, I. *et al*. (1997) Mutation of PTB binding sites causes misregulation of alternative 3′ splice site selection *in vivo*. *RNA* 3, 764–778
12 Burd, C.G. and Dreyfuss, G. (1994) RNA binding specificity of hnRNP A1: significance of hnRNP A1 high-affinity binding sites in pre-mRNA splicing. *EMBO J.* 13, 1197–1204
13 Du, H. and Rosbash, M. (2002) The U1 snRNP protein U1C recognizes the 5′ splice site in the absence of base pairing. *Nature* 419, 86–90
14 Coulter, L.R. *et al*. (1997) Identification of a new class of exonic splicing enhancers by *in vivo* selection. *Mol. Cell. Biol.* 17, 2143–2150
15 Tian, H. and Kole, R. (1995) Selection of novel exon recognition elements from a pool of random sequences. *Mol. Cell. Biol.* 15, 6291–6298
16 Lund, M. and Kjems, J. (2002) Defining a 5′ splice site by functional selection in the presence and absence of U1 snRNA 5′ end. *RNA* 8, 166–179
17 Liu, H.X. *et al*. (1998) Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins. *Genes Dev.* 12, 1998–2012
18 Liu, H.X. *et al*. (2000) Exonic splicing enhancer motif recognized by human SC35 under splicing conditions. *Mol. Cell. Biol.* 20, 1063–1071
19 Cartegni, L. *et al*. (2002) Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat. Rev. Genet.* 3, 285–298
20 Cartegni, L. *et al*. (2003) ESEfinder: a web resource to identify exonic splicing enhancers. *Nucleic Acids Res.* 31, 3568–3571
21 Buratti, E. *et al*. (2004) RNA folding affects the recruitment of SR proteins by mouse and human polypurinic enhancer elements in the fibronectin EDA exon. *Mol. Cell. Biol.* 24, 1387–1400
22 Fairbrother, W.G. *et al*. (2002) Predictive identification of exonic splicing enhancers in human genes. *Science* 297, 1007–1013
23 Fairbrother, W.G. *et al*. (2004) RESCUE-ESE identifies candidate exonic splicing enhancers in vertebrate exons. *Nucleic Acids Res.* 32, W187–190
24 Cramer, P. *et al*. (1999) Coupling of transcription with alternative splicing: RNA pol II promoters modulate SF2/ASF and 9G8 effects on an exonic splicing enhancer. *Mol. Cell* 4, 251–258
25 Pagani, F. *et al*. (2003) Missense, nonsense, and neutral mutations define juxtaposed regulatory elements of splicing in cystic fibrosis transmembrane regulator exon 9. *J. Biol. Chem.* 278, 26580–26588
26 Tian, H. and Kole, R. (2001) Strong RNA splicing enhancers identified by a modified method of cycled selection interact with SR protein. *J. Biol. Chem.* 276, 33833–33839
27 Woerfel, G. and Bindereif, A. (2001) *In vitro* selection of exonic splicing enhancer sequences: identification of novel CD44 enhancers. *Nucleic Acids Res.* 29, 3204–3211
28 Schaal, T.D. and Maniatis, T. (1999) Selection and characterization of pre-mRNA splicing enhancers: identification of novel SR protein-specific enhancer sequences. *Mol. Cell. Biol.* 19, 1705–1719
29 Cartegni, L. and Krainer, A.R. (2002) Disruption of an SF2/ASF-dependent exonic splicing enhancer in SMN2 causes spinal muscular atrophy in the absence of SMN1. *Nat. Genet.* 30, 377–384
30 Kashima, T. and Manley, J.L. (2003) A negative element in SMN2 exon 7 inhibits splicing in spinal muscular atrophy. *Nat. Genet.* 34, 460–463
31 Hofmann, Y. *et al*. (2000) Htra2-β 1 stimulates an exonic splicing enhancer and can restore full-length SMN expression to survival motor neuron 2 (SMN2). *Proc. Natl. Acad. Sci. U. S. A.* 97, 9618–9623

1471-4914/$ - see front matter
doi:10.1016/j.molmed.2004.11.007