REVIEW

Sorting out the complexity of SR protein functions

BRENTON R. GRAVELEY
Department of Genetics and Developmental Biology, University of Connecticut Health Center, Farmington, Connecticut 06030, USA

INTRODUCTION

Members of the serine/arginine-rich (SR) protein family have multiple functions in the pre-mRNA splicing reaction. In addition to being required for the removal of constitutively spliced introns, SR proteins can function to regulate alternative splicing both in vitro and in vivo (Ge & Manley, 1990; Krainer et al., 1990a; Fu et al., 1992; Zahler et al., 1993a; Caceres et al., 1994; Wang & Manley, 1995). In the cell, SR proteins migrate from speckles—subnuclear domains that may function as storage sites for certain splicing factors—to sites of active transcription (Misteli et al., 1997; Misteli & Spector, 1999) and some SR proteins have been found to shuttle in and out of the nucleus (Caceres et al., 1998). The subcellular localization of SR proteins can be modulated by phosphorylation (Misteli & Spector, 1998; Misteli et al., 1998) and this undoubtedly underlies some regulated splicing events. However, once in the nucleus and localized to the nascent pre-mRNA, exactly how SR proteins engage the general splicing machinery to recognize specific splice sites is unclear and is an area of intense investigation.

All SR proteins have a modular organization and contain an N-terminal RNA-binding domain that interacts with the pre-mRNA and a C-terminal RS domain that functions as a protein interaction domain. SR proteins have been proposed to function by binding to the pre-mRNA and recruiting a number of different general splicing factors to the pre-mRNA during spliceosome assembly (Wu & Maniatis, 1993; Kohtz et al., 1994; Roscigno & Garcia-Blanco, 1995; Zuo & Maniatis, 1996). In addition, SR proteins are thought to mediate interactions between splicing factors bound to the 5’ and 3’ splice sites (Wu & Maniatis, 1993). However, it has not been clear whether all of these activities occur during the removal of each intron. Recent studies now suggest that all of the proposed SR protein functions are carried out during each round of splicing, and at least some of these functions are performed by independent SR protein molecules. This review discusses recent advances in understanding the diverse functions of SR proteins in metazoan pre-mRNA splicing and presents a model that takes these new findings into account. Although the reader should keep in mind that the activity of SR proteins in vivo can be influenced by modulating their subcellular localization, this review focuses on biochemical experiments that specifically address the mechanisms by which SR proteins directly function in the splicing reaction. (The reader may also refer to related reviews recently published elsewhere (Tacke & Manley, 1999; Blencowe, 2000).)

spliceosome assembly

Considerable progress has been made in the understanding of spliceosome assembly (Reed & Palandjian, 1997; Burge et al., 1999). The spliceosome consists of five small nuclear ribonucleoprotein particles (snRNPs), designated U1, U2, U4, U5, and U6, and a large number of non-snRNP proteins. Spliceosome assembly is directed, in part, by the RNA sequences at the splice sites. In mammals, the 5’ splice site consensus sequence is AG/GURAGU (where / denotes the exon/intron boundary). Three distinct sequence elements are found at the 3’ splice site—the branchpoint (YNYURAC), a polypyrimidine tract, and the actual 3’ splice site (YAG/N). Spliceosome assembly proceeds in a step-wise fashion (Fig. 1) and is initiated upon the binding of U1 snRNP to the 5’ splice site, SF1/mBBP to the branchpoint sequence and U2 auxiliary factor (U2AF) to the pyrimidine tract and 3’ AG to form the E, or early, complex. Subsequently, U2 snRNP binds to the branchpoint to form A complex, followed by the

Reprint requests to: Brenton R. Graveley, Department of Genetics and Developmental Biology, University of Connecticut Health Center, 263 Farmington Avenue, Farmington, Connecticut 06030, USA; e-mail: graveley@neuron.uchc.edu.
FIGURE 1. Spliceosome assembly pathway. In the first step of spliceosome assembly, U1 snRNP binds to the 5’ splice site (Michaud & Reed, 1991), SF1/mBBP binds to the branchpoint (Abovich & Rosbash, 1997), and U2 auxiliary factor (U2AF) binds to the pyrimidine tract (Ruskin et al., 1988; Zamore & Green, 1989; Bennett et al., 1992) and 3’YAG (Merendino et al., 1999; Wu et al., 1999; Zorio & Reddy, 1988). This complex commits the pre-mRNA to the splicing pathway and is called the E, or early, complex (Michaud & Reed, 1991). Next, E complex is converted to A complex when the U2 snRNP binds to the branchpoint (Bennett et al., 1992). Subsequently, B complex is formed when the U4, U5, and U6 snRNPs enter the spliceosome as a tri-snRNP particle (Reed & Palandjian, 1997). Finally, a massive rearrangement occurs in which U6 replaces U1 at the 5’ splice site, U6 and U2 interact, U5 bridges the splice sites, and U1 and U4 become destabilized. This rearranged spliceosome is called the C complex and is catalytically active.

The association of the U4/U6+U5 tri-snRNP to form B complex. Next, the spliceosome rearranges to form the catalytically active C complex. Some of the SR protein functions, such as the recruitment of U1 snRNP and U2AF to the 5’ and 3’ splice sites, respectively, are thought to act at the stage of E complex formation (Reed, 1996). SR proteins also appear to function at later steps, such as the transition from A to B complex (Roscigno & Garcia-Blanco, 1995) and even perhaps after the first catalytic step of splicing (Chew et al., 1999).

THE SR PROTEIN FAMILY OF SPlicing FACTORS

SR proteins were independently discovered by a number of groups taking very different approaches. For instance, SF2/ASF (Splicing Factor 2/Alternative Splicing Factor; Ge et al., 1991; Krainer et al., 1991) was purified from HeLa cell nuclear extract as a factor required to reconstitute splicing in S100 splicing-deficient HeLa cell extract (Krainer et al., 1990b) and to induce splice site switching (Ge & Manley, 1990). In another approach, monoclonal antibodies directed against purified spliceosomes were used to identify SR proteins. In one case, this led to the identification of SC35 (Fu & Maniatis, 1990, 1992b). Another group, using the monoclonal antibody mAb104, identified an entire family of related proteins—including SC35 and SF2/ASF—that they termed the SR protein family based on their high serine and arginine content (Roth et al., 1991; Zahler et al., 1992). Subsequently, additional members of the SR protein family were identified (Ayane et al., 1991; Champlin et al., 1991; Diamond et al., 1993; Zahler et al., 1993b; Cavaloc et al., 1994; Screaton et al., 1995; Zhang & Wu, 1996; Soret et al., 1998), and the human SR protein family currently contains 10 known members (Fig. 2A). Although SR proteins have been identified in all metazoan species examined and plants (Lazar et al., 1995; Lopato et al., 1996a, 1996b, 1999; Golovkin & Reddy, 1998, 1999), they are not found in all eukaryotes. For example, Schizosaccharomyces pombe contains at least two SR proteins (Gross et al., 1998; Lutzlerberger et al., 1999), while Saccharomyces cerevisiae contains none.

Until recently, metazoan SR proteins were thought to be encoded by essential genes. For instance, in Drosophila melanogaster, deletion of the gene encoding the B52 protein, which corresponds to human SRp55, results in lethality in the first or second instar larval stage (Ring & Lis, 1994). Likewise, conditional depletion of the ASF/SF2 protein in chicken DT40 B-cells results in cell death (Wang et al., 1996). Moreover, a late embryonic lethal phenotype was observed when the Caenorhabditis elegans SF2/ASF homolog, rsp-3, was targeted by RNAi (Longman et al., 2000). Surprisingly, no phenotype was observed when six other C. elegans SR protein genes were individually targeted by RNAi (Longman et al., 2000). However, when multiple C. elegans SR protein genes were simultaneously targeted by this method, developmental defects or lethality was observed (Longman et al., 2000). It will be important to determine the reasons why some SR pro-
Proteins are not essential. Do the nonessential SR proteins simply not participate in the splicing of essential genes, or can other SR proteins functionally substitute for the missing SR protein? The absence of an obvious phenotype in the worms treated with dsRNA, however, does not exclude the possibility that certain SR proteins are completely dispensable. For example, some SR proteins may be required for the proper function of specific neurons, and worms lacking these SR proteins may display subtle phenotypes that would be difficult to assess without rigorous behavioral or electrophysiological assays. Complete sets of knockouts in other metazoans must be created and analyzed to establish the generality of the results obtained in worms. The recent completion of the genomic DNA sequence of D. melanogaster should greatly facilitate these types of experiments.

### Structural organization of SR proteins

SR proteins have a characteristic structural organization (Fig. 2A). They contain one or two RNA-binding domains at their N-terminals and a variable-length arginine/serine-rich domain at their C-terminals (the RS domain). The RNA-binding and RS domains are modular. For example, the RNA-binding domains can be exchanged between SR proteins (Chandler et al., 1997; Mayeda et al., 1999) and can bind RNA in the absence of the RS domain (Caceres & Krainer, 1993; Zuo & Manley, 1993; Tacke & Manley, 1995). Likewise, the RS domains can be exchanged between SR proteins (Chandler et al., 1997; Wang et al., 1998b) and can even function when fused to a heterologous RNA-binding domain (Graveley & Maniatis, 1998).

A number of additional RS-domain-containing proteins, distinct from the SR proteins, are involved in various aspects of pre-mRNA metabolism and are collectively referred to as SR-related proteins or SRrps (Ford et al., 1999). Examples include both subunits of U2AF, [U2AF35 (Zhang et al., 1992) and U2AF65 (Zamore et al., 1992)], snRNP components [(U1-70K (Query et al., 1989), U5-100K (Teigelkamp et al., 1997), U4/U6,U5-27K (Fetzer et al., 1997)], and hLuc7p (Fortes et al., 1999; Nishii et al., 2000)], splicing regulators [Tra (Boggs et al., 1987), Tra2 (Amrein et al., 1988), hTra2α, hTra2β (Dauwalder et al., 1997)].

### FIGURE 2. Schematic diagram of human SR proteins and SR related proteins. A: The domain structures of the known members of the human SR protein family are depicted. RRM: RNA recognition motif; RRMH: RRM homology; Z: zinc knuckle, RS: arginine/serine-rich domain. B: The domain structures for some of the human SRrps that participate in pre-mRNA splicing are depicted. All proteins, with the exception of SRm300, are drawn to scale. RRM: RNA recognition motif; RS: arginine/serine-rich domain; Zn: zinc finger; DEXD/H Box: motif characteristic of RNA helicases.
et al., 1996; Beil et al., 1997), and SWAP (Denhez & Lafyatis, 1994; Spikes et al., 1994)), splicing coactivators [SRm160/SRm300 (Blencowe et al., 1998, 2000)], RNA helicases [hPrp16 (Zhou & Reed, 1998), HRH1 (Ono et al., 1994), and Hel117 (Sukegawa & Blobel, 1995)], and protein kinases [Clk/Sty 1, 2, and 3 (Johnson & Smith, 1991; Hanes et al., 1994)]. Some of the human SRrps that participate in pre-mRNA splicing are depicted in Figure 2B. Interestingly, additional SRrps have been identified, and these proteins participate in a number of processes other than pre-mRNA splicing. For a detailed review on SRrps see Blencowe et al. (1999).

The RNA-binding domain

To fully understand how SR proteins function, it is critical to know where they bind to the pre-mRNA. A number of studies over the last few years have therefore attempted to elucidate the RNA-binding specificities of each SR protein (Tacke & Manley, 1995; Shi et al., 1997; Tacke et al., 1997, 1998; Liu et al., 1998, 2000; Cavaloc et al., 1999; Schaal & Maniatis, 1999b). A surprising conclusion from these experiments is that SR proteins recognize a vast array of RNA sequences (see Table 1). Although SR proteins do display distinct RNA-binding specificities, the consensus sequences that they recognize are rather degenerate. In fact, in several cases, sequences identified as binding sites for one SR protein can also be recognized by other SR proteins (Tacke & Manley, 1995; Liu et al., 1998; Tacke et al., 1999). The overlapping and promiscuous RNA-binding specificities of SR proteins may partially account for their apparent redundancy in their function. The possible significance of the degenerate SR protein recognition sequences will be discussed in more detail below.

### The RS domain

It is now widely accepted that the RS domains of SR proteins function as protein interaction domains. SR proteins were initially shown to interact with one another, the splicing regulators Tra and Tra2, and with other components of the splicing machinery that contain RS domains such as U1-70K and U2AF35 (Wu & Maniatis, 1993; Amrein et al., 1994; Kohtz et al., 1994). Importantly, these protein interactions all require the RS domains of each protein (Wu & Maniatis, 1993; Amrein et al., 1994; Kohtz et al., 1994). Only recently, however, have the RS domains been shown to be sufficient to mediate protein interactions. For instance, the RS domain of SF2/ASF has been shown to be sufficient to interact with RSF1, a Drosophila splicing repressor (Labourier et al., 1999). In addition, when artificially tethered to the pre-mRNA, the RS domains of several human SR proteins are sufficient to activate enhancer-dependent splicing (Graveley & Maniatis, 1998), an activity that presumably requires protein interactions. In contrast, the RS domain of SF2/ASF is unable to interact with U1-70K (Xiao & Manley, 1997). These studies indicate that the RS domains of SR pro-

### TABLE 1. RNA sequences identified as SR protein-binding sites by in vitro selection methods.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Binding site</th>
<th>Method*</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>SF2/ASF</td>
<td>RGAAGAAC</td>
<td>SELEX</td>
<td>Tacke &amp; Manley, 1995</td>
</tr>
<tr>
<td></td>
<td>AGGAACRGAC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SRSASGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SC35</td>
<td>AGSAGAGUA</td>
<td>SELEX</td>
<td>Tacke &amp; Manley, 1995</td>
</tr>
<tr>
<td></td>
<td>GGUCCAGUA</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>UGUCUCAGWU</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GWUWCCUGCUA</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GGGAUAUGCG</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GAGCAGUAQKS</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AGGAGAU</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9G8</td>
<td>GRYYGSR</td>
<td>Functional</td>
<td>Liu et al., 2000</td>
</tr>
<tr>
<td></td>
<td>AGACKACGAY</td>
<td>SELEX</td>
<td>Cavaloc et al., 1999</td>
</tr>
<tr>
<td></td>
<td>AGGAGAGAY</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRp40</td>
<td>UGGGAGCGUVRGCUCGY</td>
<td>SELEX</td>
<td>Tacke et al., 1997</td>
</tr>
<tr>
<td></td>
<td>ACDGS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRp55</td>
<td>USCGKM</td>
<td>SELEX</td>
<td>Liu et al., 1998</td>
</tr>
<tr>
<td>B52</td>
<td>GRUCACCDNGGCGAAGNG</td>
<td>SELEX</td>
<td>Shi et al., 1997</td>
</tr>
<tr>
<td>hTra2β</td>
<td>(GAA)n</td>
<td>SELEX</td>
<td>Tacke et al., 1998</td>
</tr>
</tbody>
</table>

* N: any nucleotide; R: purine; Y: pyrimidine; S: G or C; K: U or G; W: A or U; D: A, G, or U; M: A or C.

* “SELEX” indicates that the RNA sequence was determined to be a high affinity binding site for a purified SR protein; “functional” indicates that the RNA sequence was determined to function as an SR protein-specific splicing enhancer.
proteins are indeed protein interaction domains, but suggest that different protein interactions may have distinct RS domain sequence requirements.

The modular domain organization of SR proteins has made it possible to assess the specificity of RS domain function in a variety of contexts by exchanging RS domains between different SR proteins. For example, it was found that the RS domains from other SR proteins, as well as the splicing regulator Tra2, can substitute for the SF2/ASF RS domain in chicken cells lacking endogenous SF2/ASF (Wang et al., 1998b). Likewise, it was found that several RS domains can replace the Tra2 C-terminal RS domain in D. melanogaster (Dauwalder & Mattox, 1998). Moreover, the RS domain of Tra can also substitute for the SWAP RS domain in Drosophila (Li & Bingham, 1991). Several in vitro studies have also been used to analyze RS domain specificity. Chandler et al. (1997) found that exchanging the RS domains of SF2/ASF and SC35 had no effect on the ability of the protein to complement splicing-deficient S100 extracts, suggesting that RS domains are fully interchangeable. However, when the RS domains of six different human SR proteins were tested for their ability to activate splicing when artificially tethered to the pre-mRNA, it was found that the activity of RS domains differed quantitatively (Graveley et al., 1998). In fact, the potency of an RS domain varies directly with the number of RS dipeptides present (Graveley et al., 1998). In summary, these results show that the RS domains of SR proteins are, for the most part, fully interchangeable and seem to differ from one another only quantitatively. However, subtle differences not detected in these rather crude assays could lead to profound differences in the animal.

**RS domain phosphorylation**

A number of studies have shown that the serine residues within the RS domains of SR proteins are extensively phosphorylated. This phosphorylation appears to influence the subcellular localization of SR proteins (Gui et al., 1994a; Colwill et al., 1996) as well as protein interactions involving SR proteins (Wang et al., 1998a; Xiao & Manley, 1998), both of which may change the ability of SR proteins to function in splicing. In fact, it has recently been shown that both hyper- and hypo-phosphorylated SR proteins are unable to support splicing (Kanopka et al., 1998; Prasad et al., 1999; Sanford & Bruzik, 1999). The modulation of phosphorylation levels appears to be used to control the activity of SR proteins in adenovirus infection (Kanopka et al., 1998) and early development in the nematode Ascaris lumbricoides (Sanford & Bruzik, 1999).

Three types of protein kinases have been identified that can phosphorylate RS domains in vitro. These include the SR protein kinase family [SRPK1 (Gui et al., 1994b) and SRPK2 (Wang et al., 1998a)], the Clk/Sty family (see, for example, Colwill et al., 1996), and DNA topoisomerase I (Rossi et al., 1996). The most direct evidence for an in vivo role of SR protein kinases comes from an elegant series of experiments in D. melanogaster. Drosophila contains a homolog of Clk/Sty called Doa. Doa was found to phosphorylate the Drosophila SR protein RBP1, as well as Tra and Tra2, in vitro, and RBP1 in vivo (Du et al., 1998). Interestingly, Doa mutations disrupt the splicing of the doublesex pre-mRNA, but have no effect on the splicing of the fruitless pre-mRNA (Du et al., 1998)—both targets of regulation by Tra, Tra2, and RBP1. This result is of particular interest because Tra, Tra2, and RBP1 control alternative 3' splice site selection in the doublesex pre-mRNA, and alternative 5' splice site selection in the fruitless pre-mRNA. This observation suggests that the regulation of alternative 5' and 3' splice sites by SR proteins have different requirements for RS domain phosphorylation. In addition, these results demonstrate that an SR protein kinase can control the developmentally regulated alternative splicing of specific pre-mRNAs.

**THE FUNCTIONS OF SR PROTEINS IN PRE-mRNA SPLICING**

Despite the plethora of proposed SR protein functions, they can essentially be divided into two categories— exon dependent and exon independent.

*The exon-dependent functions of SR proteins*

Perhaps the best-characterized functions of SR proteins are those that involve binding to exon sequences. The exon-dependent functions can be subdivided into regulated 3' splice site selection, regulated 5' splice site selection, and constitutive functions. Although most of the work in this area has focused on how SR proteins function to regulate alternative splicing, it is now clear that the binding of SR proteins to constitutive exons also plays an important role in the splicing reaction.

*Regulated 3' splice site selection*

Soon after SR proteins were discovered, it was appreciated that they function not only as essential splicing factors, but also as splicing regulators. A flurry of papers demonstrated that SR proteins could bind to exon sequences, where they functioned to enhance the splicing of the adjacent intron (see, for example, Lavigueur et al., 1993; Sun et al., 1993; Tian & Maniatis, 1993). These RNA elements have since been termed exonic splicing enhancers and have been identified in a number of metazoan exons. A well-characterized example of an exonic splicing enhancer is provided by the D. melanogaster sex-determination gene doublesex, which is alternatively spliced in males and females (Baker,
In males, exon 4 is skipped, whereas in females, exon 4 is included. This splicing event requires a complex exonic splicing enhancer in exon 4 that consists of six 13-nt repeats, collectively referred to as the dsx repeat element (dsxRE). Each repeat of the dsxRE is recognized by the SR protein RBP1 and the splicing regulator Tra2 (Lynch & Maniatis, 1996). However, the binding of these factors to the splicing enhancer requires the presence of the splicing regulator Tra, which is expressed only in females. Tra, Tra2, and RBP1 interact with each other to cooperatively bind the dsx splicing enhancer (Lynch & Maniatis, 1996). This splicing enhancer complex, in turn, functions to activate the splicing of the upstream intron. The dsx splicing enhancer has served as a paradigm for regulated splicing enhancers.

The most controversial question in the SR protein field is how enhancer-bound SR proteins function to activate splicing, and several mechanisms have been proposed. The observations that many enhancer-dependent introns contain suboptimal pyrimidine tracts and that SR proteins can physically interact with U2AF35 (Wu & Maniatis, 1993) led to the proposal that splicing enhancers function by recruiting U2AF65 to the adjacent pyrimidine tract of regulated introns (Tian & Maniatis, 1993; Wu & Maniatis, 1993) (Fig. 3A). Subsequent in vitro experiments using purified recombinant proteins have shown that SR proteins can, in fact, recruit U2AF65 to the upstream intron and that this recruitment requires U2AF35 (Zuo & Maniatis, 1996). Similar results have also been obtained in experiments performed in nuclear extracts (Wang et al., 1995; Bouck et al., 1998; B.R. Graveley, K. Hertel, & T. Maniatis, unpubl. data). However, other studies in nuclear extracts have failed to detect a difference in U2AF binding in the presence or absence of a splicing enhancer, raising the possibility that enhancer-bound SR proteins may act through an alternative mechanism (Kan & Green, 1999; Li & Blencowe, 1999). For instance, it has been suggested that splicing enhancer complexes may function in conjunction with the splicing coactivator SRm160/300 (Eldridge et al., 1999) (Fig. 3B). In at least one case, splicing enhancers may function to counteract splicing inhibitory complexes (Kan & Green, 1999; Fig. 3C). Finally, although splicing enhancers appear to function primarily at the early stages of spliceosome assembly—prior to, or at the step of splice site selection—one set of recent experiments indicates that splicing enhancers may also function to enhance the second step of splicing (Chew et al., 1999).

Although SRm160/300 clearly participates in enhancer-dependent splicing and some enhancers may function in part by negating the effects of splicing inhibitors, at least one set of experiments is difficult to reconcile with models that do not invoke U2AF recruitment. It has been shown for several enhancer-dependent introns that improving the weak pyrimidine tract can relieve the requirement for a splicing enhancer (Tian & Maniatis, 1994; Lorson & Androphy, 2000; B.R. Graveley, K. Hertel, & T. Maniatis, unpubl. results). For example, replacing the normal pyrimidine tracts of either the Drosophila dsx or the mouse IgM pre-mRNAs with sequences that have a 5–10-fold higher affinity for U2AF render those RNAs enhancer independent (B.R. Graveley, K. Hertel, & T. Maniatis, unpubl. data). Likewise, decreasing the strength of a pyrimidine tract of a constitutively spliced pre-mRNA, and therefore the affinity of U2AF, can make that intron enhancer dependent (Tian & Maniatis, 1992; Graveley & Maniatis, 1998). Thus, the requirement for a splicing enhancer is directly related to the intrinsic affinity of U2AF for the pyrimidine tract. The simplest explanation for these results is that splicing enhancers do indeed function to recruit U2AF. The failure to observe differences in U2AF binding in some cases may be a reflection of differences in experimental systems and the highly cooperative nature of spliceosome assembly. It is also possible that U2AF recruitment is not sufficient to activate splicing. Future experiments must more clearly address the detailed mechanism of enhancer function and definitively identify the target for the putative protein–protein interactions between enhancer complexes and the splicing machinery.

Regulated 5′ splice site selection

In addition to promoting the recognition of alternative 3′ splice sites, splicing enhancers also appear to promote the recognition of alternative 5′ splice sites. Again, the best example comes from a Drosophila gene—fruitless. Just as in the case of dsx pre-mRNA, the fruitless pre-mRNA is alternatively spliced in males and females, except the latter case involves the selection of alternative 5′ splice sites (Ryner et al., 1996). An analysis of the sequence of the fruitless gene revealed the presence of three nearly perfect copies of the dsxRE 13-nt repeats, immediately upstream of the female-specific 5′ splice site (Ryner et al., 1996). This observation suggests that the same splicing enhancer complex that regulates dsx alternative splicing may regulate fruitless alternative splicing. In fact, Tra and Tra2 mutant females fail to utilize the female-specific 5′ splice site (Ryner et al., 1996). This result directly implicates the Tra, Tra2, RBP1 heterotrimeric complex in 5′ splice site selection in the fruitless pre-mRNA. Importantly, three other recent examples suggest that SR proteins bound to upstream splicing enhancers stimulate downstream 5′ splice site utilization in other pre-mRNAs (Bourgeois et al., 1999; Côté et al., 1999; Selvakumar & Helfman, 1999). These results strongly suggest that upstream splicing enhancer complexes function to stimulate the binding of U1 snRNP to the downstream 5′ splice site through an interaction with U1-70K (Fig. 3D).
However, it remains to be shown if this is the mechanism by which upstream SR proteins promote the recognition of the 5’ splice site.

The constitutive functions

In addition to their exon-dependent role in regulating alternative splicing, SR proteins have recently been shown to have an exon-dependent function in constitutive splicing. The exon-definition model (Berget, 1995) proposes that interactions between components bound to the 5’ and 3’ splice sites flanking an exon serve to distinguish exons, which are typically small, from introns, which are typically large. SR proteins have been proposed to participate in this process by binding to the exon where they simultaneously interact with U2AF35 bound to the upstream 3’ splice site and the 70K protein of U1 snRNP bound to the downstream 5’ splice site. It is thought that the majority of constitutively spliced exons are defined by this mechanism (Reed, 1996). If correct, constitutively spliced exons should contain binding sites for one or more SR proteins. In support of this
model, a number of SR protein binding sites have recently been identified in constitutive exons (Mayeda et al., 1999; Schaal & Maniatis, 1999a). In fact, these sites have been shown to function as constitutive splicing enhancers (Mayeda et al., 1999; Schaal & Maniatis, 1999a).

Thus it appears that SR proteins have similar exon-dependent functions in regulated and constitutive splicing. In constitutive splicing, the SR proteins bind to the exon constitutively, whereas in alternative splicing the SR protein association is regulated. The functional interactions between SR proteins and constitutive exons place a significant constraint on the types of RNA sequences that can be present in coding exons, because codons and SR protein binding sites would overlap. Perhaps the degenerate RNA binding specificities of SR proteins have evolved to cope with this functional conflict. It is interesting to note that for *doublesex* and *fruitless*—two alternative splicing events that are essential to the survival of the species—the splicing enhancers, which are highly conserved, are present in untranslated regions. Thus, in instances where a high degree of RNA-binding specificity is necessary to achieve proper regulation, it may not be possible for splicing enhancers to coexist with coding sequence.

The exon-independent functions of SR proteins

In addition to their exon-dependent functions, SR proteins have some exon-independent functions. Most of these functions are considered exon-independent simply because it is not known if SR proteins must bind to exons to perform these functions. However, a recent study has shown that at least some essential SR protein functions do not require binding to exon sequences. Based in part on the observation that the second exon is not necessary for the first step of splicing (Smith et al., 1989; Anderson & Moore, 1997), Hertel and Maniatis (1999) performed an interesting set of experiments in which they determined the minimal number of nucleotides in the first exon that are required for the first step of splicing. They found that a pre-mRNA that contains only a single nucleotide in the first exon, and completely lacks the second exon, can undergo the first step of splicing. Interestingly, SR proteins are required for this splicing reaction with the minimal substrate. Moreover, all SR proteins tested in this reaction could promote splicing, indicating that there is functional redundancy with regards to the exon-independent function(s) of SR proteins. Thus, SR proteins perform some essential function(s) that does not require their binding to exons—perhaps mediating interactions between the 5’ and 3’ splice sites.

Cross-intron interactions

SR proteins have been proposed to play an important role in the pairing of 5’ and 3’ splice sites (Fig. 4A). This idea was originally based on the observation that the SR protein SC35 can promote an interaction between U1 and U2 snRNPs on the pre-mRNA (Fu & Maniatis, 1992a). Further evidence for this model was provided by the observation that SR proteins can simultaneously interact with U1-70K and U2AF35, components bound to the 5’ and 3’ splice sites respectively (Wu & Maniatis, 1993). In addition, SR proteins are required for trans-splicing, a reaction in which the 5’ and 3’ splice sites are contained on separate RNA molecules (Bruzik & Maniatis, 1995; Chiara & Reed, 1995). Although SR proteins clearly have the potential to function in splice site pairing, it remains to be determined whether SR proteins actually participate in this process in the spliceosome.

Tri-snRNP incorporation

SR proteins have also been shown to play a role in the recruitment of the U4/U6/U5 tri-snRNP into the spliceosome (Roscigno & Garcia-Blanco, 1995; Tarn & Steitz, 1995) (Fig. 4A). Although the precise role of SR proteins in this activity has not been determined, it is most likely that this recruitment is mediated by an interaction between an SR protein present in the partially assembled spliceosome and a component of the tri-snRNP. There are at least two proteins within the human U4/U6/U5 tri-snRNP that contain RS domains—U4/U6/U5-27K (Fetzer et al., 1997) and U5-100K (Teigelkamp et al., 1997)—and are therefore potential interaction targets of SR proteins.

5’ splice site recognition

As described earlier, enhancer-bound SR proteins can promote the utilization of a downstream 5’ splice site, and this activity may involve the recruitment of U1 snRNP to the 5’ splice site. However, SR proteins may also mediate U1 snRNP recruitment by a mechanism that does not involve binding to exon sequences. Experiments performed with purified components have shown that SF2/ASF can recruit U1 snRNP to an RNA containing a 5’ splice site (Kohtz et al., 1994; Jamison et al., 1995). This activity requires an intact 5’ splice site sequence, and at least one functional RNA binding domain and the RS domain of SF2/ASF (Kohtz et al., 1994; Jamison et al., 1995). Moreover, SF2/ASF has been shown to bind RNAs containing wild-type, but not mutant 5’ splice sites (Zuo & Manley, 1994). Together, these results suggest that SF2/ASF binds directly to the 5’ splice site, interacts with U1-70K, and recruits U1 snRNP (Fig. 4B). However, many details of this model still need to be resolved, such as whether SF2/
ASF can actually bind to the pre-mRNA:U1 snRNA duplex, and if not, where SF2/ASF binds to carry out this function.

SR proteins may also participate in the recognition of the 5’ splice site after U1 snRNP binding. This insight came from experiments in which U1 snRNP was either debilitated by antisense oligonucleotides or depleted from nuclear extracts (Crispino et al., 1994; Tarn & Steitz, 1994). Extracts treated by either method are unable to support splicing on their own. However, the addition of high concentrations of SR proteins can restore splicing (Crispino et al., 1994; Tarn & Steitz, 1994). Thus, SR proteins appear to have the ability to bypass the requirement for U1 snRNP in 5’ splice site selection. Subsequent experiments have suggested that the added SR proteins assist U6 snRNP in recognizing the 5’ splice site (Crispino & Sharp, 1995; Tarn & Steitz, 1995)—a step that normally occurs late in spliceosome assembly. However, this function of SR proteins may be indirect. SR proteins have been shown to recruit the U4/U6•U5 tri-snRNP to the spliceosome (Roscigno & Garcia-Blanco, 1995; Tarn & Steitz, 1995). Thus, the
High concentrations of SR proteins added to U1 snRNP-depleted extracts may increase the number of spliceosomes that contain the U4/U6•U5 tri-snRNP. This, in turn, would increase the probability that U6 snRNA could functionally recognize the 5′ splice site in a step that does not require SR proteins. Additional studies are required to determine whether SR proteins actually play a direct role in the U6 snRNP-5′ splice site interaction in spliceosomes that contain U1 snRNP, and if so, how this is achieved.

**MULTIPLE SIMULTANEOUS FUNCTIONS**

Aside from the obvious question of how SR proteins function from a mechanistic point of view, there are a number of important questions that have not been given much consideration until recently. For instance, are all of the proposed functions of SR proteins performed during the removal of each intron or do different introns require SR proteins to perform different subsets of these functions? If all of the functions are carried out for each intron, does a single SR protein molecule perform all of these functions, or do different SR protein molecules carry out each function? At one extreme, it is possible that a single SR protein molecule bound to the downstream exon can simultaneously interact with U2AF, U1-70K, and a component of the U4/U6•U5 tri-snRNP. These interactions would serve to define and pair the 5′ and 3′ splice sites, and recruit the tri-snRNP, effectively performing all of the SR protein functions in one fell swoop. At the other extreme, each of the SR protein functions could be performed by individual SR protein molecules. Although these issues have not been fully answered, some recent experimental systems have allowed some of these questions to be addressed. Four papers contain results that together strongly indicate that multiple SR protein molecules participate in the removal of each intron (Graveley & Maniatis, 1998; Tacke et al., 1998; Xiao & Manley, 1998; Hertel & Maniatis, 1999). Specifically, these results suggest that each splicing event requires at least one SR protein molecule to perform a regulated exon-dependent function, and a second SR protein molecule to perform one or more of the other SR protein functions. Three of these papers have used pre-mRNAs containing SR protein-binding sites in the downstream exon. Thus, for the purposes of this discussion, the regulated exon-dependent functions of SR proteins will apply only to 3′ splice site selection and not 5′ splice site selection, which is likely to be constitutive in these cases.

Hybrid proteins consisting of the MS2 coat protein and an RS domain are sufficient to activate the splicing of enhancer-dependent pre-mRNAs containing an MS2 binding site in HeLa cell nuclear extracts (Graveley & Maniatis, 1998). In S100 extracts, however, both an SR protein and an MS2-RS protein, neither of which can function alone, are required for splicing (Graveley & Maniatis, 1998). The MS2-RS proteins are probably only able to substitute for the regulated exon-dependent function of SR proteins for several reasons. First, because they can only bind to the MS2 site, their activity is spatially restricted. Second, the RS domain alone has been shown to be necessary but not sufficient to interact with the U1-70K protein (Xiao & Manley, 1997). Thus, the MS2-RS proteins are probably unable to recruit U1 snRNP to the 5′ splice site or mediate cross-intron interactions. The additional SR protein that is required in this situation is most likely performing one or both of these functions, and may also be involved in tri-snRNP incorporation. Thus, one interpretation of this experiment is that the MS2-RS proteins are performing the regulated exon-dependent function and the intact SR proteins are performing the constitutive exon-dependent and exon-independent functions (Fig. 5A). These results, as well as those described below, also indicate that the regulated exon-dependent functions of SR proteins can be uncoupled from the other SR protein functions.

The human homologs of the *Drosophila* Tra2 protein (hTra2α and hTra2β) function as activators of pre-mRNA splicing in vitro (Tacke et al., 1998). In S100 extracts, neither hTra2α nor hTra2β are sufficient to activate the splicing of a mouse IgM pre-mRNA containing a binding site for hTra2 as a splicing enhancer. However, the addition of either protein to a dilute nuclear extract, which alone does not support splicing, activates splicing of the IgM pre-mRNA containing the hTra2 splicing enhancer. These experiments are consistent with a model in which the activity of the hTra2 proteins is restricted to the regulated exon-dependent functions and the additional SR proteins present in the nuclear extract are required to perform the constitutive exon-dependent and exon-independent functions (Fig. 5B).

A study of the phosphorylation requirements for SR protein function has provided additional evidence that multiple SR proteins function in the splicing of each intron (Xiao & Manley, 1998). The HIV-1 *tat* pre-mRNA is known to contain SF2/ASF and to lack SC35-responsive elements (Fu, 1993; Mayeda et al., 1999). In S100 extracts, *tat* pre-mRNA splicing can be activated by the addition of SF2/ASF, but not SC35. Xiao and Manley (1998) have found that thiophosphorylated SF2/ASF is unable to activate *tat* pre-mRNA splicing in an S100 extract. However, they found that a combination of thiophosphorylated SF2/ASF and unmodified SC35 reconstituted *tat* pre-mRNA splicing in S100 extracts. Consistent with the model developed above, these results suggest that the thiophosphorylated SF2/ASF is performing the regulated exon-dependent functions and SC35 is performing the constitutive exon-dependent and exon-independent functions of SR proteins (Fig. 5C). Moreover, these results indicate that the regulated exon-dependent functions of SR
FIGURE 5. Summary of results indicating that multiple individual SR proteins participate in the removal of each intron. Throughout this figure the regulated exon-dependent functions of SR proteins in 3' splice site selection are illustrated as U2AF recruitment and the exon-independent functions are illustrated as cross-intron bridging and tri-snRNP recruitment. Note that the exon-dependent function of SR proteins in 5' splice site recognition may be occurring, but is not depicted in this illustration. A: MS2-RS proteins and SR proteins are required for enhancer-dependent splicing (Graveley & Maniatis, 1998). In this model, the MS2-RS protein is performing the regulated exon-dependent SR protein functions and the SR protein is performing the exon-independent and constitutive exon-dependent functions of SR proteins. B: hTra2 proteins and SR proteins are required for enhancer-dependent splicing (Tacke et al., 1998). In this model, hTra2 is performing the regulated exon-dependent functions and SR proteins are performing the exon-independent and constitutive exon-dependent functions. C: The phosphorylation requirements for the regulated exon-dependent functions are distinct from the other SR protein functions (Xiao & Manley, 1998). In this model, thiophosphorylated SF2/ASF is performing the regulated exon-dependent functions and SC35 is performing the exon-independent and constitutive exon-dependent functions. D: SR proteins have exon-independent functions (Hertel & Maniatis, 1999). In this model, a pre-mRNA substrate containing only 1 nt in the first exon and completely lacking the second exon can undergo the first step of splicing in a reaction that requires SR proteins. The SR protein in this case must be performing an exon-independent function.
proteins have phosphorylation requirements distinct from the other SR protein functions—the regulated exon-dependent functions do not require dephosphorylation whereas at least one of the other SR protein functions does.

The last experiments that support this model are those discussed earlier showing that a pre-mRNA containing only a single nucleotide in the first exon, and completely lacking the second exon, can undergo the first step of splicing (Hertel & Maniatis, 1999). The finding that SR proteins are required for the splicing of this minimal substrate demonstrates that SR proteins have an essential exon-independent function that, at least for some introns, is sufficient for splicing (Fig. 5D).

Taken together, this series of results provides very strong functional evidence that separate SR protein molecules perform the regulated exon-dependent functions and constitutive exon-dependent and exon-independent functions in the removal of each intron. This model can account for some of the apparent discrepancies in the literature. For instance, it was originally proposed that SR proteins were functionally redundant based on the observations that any individual SR protein could restore splicing in an S100 splicing-deficient extract (Fu et al., 1992; Mayeda et al., 1992; Zahler et al., 1992). Hertel and Maniatis (1999) have shown that the splicing of strong introns requires only the exon-independent functions of SR proteins and that all SR proteins tested for this activity were functionally redundant. Perhaps, then, only the exon-dependent functions of SR proteins will display SR protein specificity. This model could therefore account for earlier observations that SR proteins are functionally redundant in vitro.

This model can also explain some discrepancies in the interchangeability of RS domains. Although not discussed earlier, the RS domain of the large U2AF sub-unit can substitute for the C-terminal RS domain of Transformer 2 of Drosophila, and can function as a splicing activation domain in vitro when fused to the MS2 coat protein (B.R. Graveley, unpubl. data). In each case, the RNA binding domain of the fusion proteins restricts the activity of the RS domain to the regulated exon-dependent functions because Tra2 is a natural splicing regulator and MS2 is an artificial splicing regulator. In contrast, the U2AF65 RS domain is unable to replace the SF2/ASF RS domain in vivo (DT40 chicken cells) or in vitro (Wang et al., 1998b). Although the in vivo result may be explained if the fusion protein failed to localize to the nucleus (which was not examined in this study), the in vitro results are not complicated by this possibility. Importantly, SF2/ASF can perform the exon-independent functions of SR proteins (Hertel & Maniatis, 1999). Therefore, the ability of the U2AF65 RS domain to function when fused to Tra2 or MS2, but not when fused to SF2/ASF, may reflect differences in the RS domain sequence requirements for the regulated exon-dependent and the exon-independent functions of SR proteins. This idea is consistent with the observation that thiophosphorylated SF2/ASF can perform the regulated exon-dependent functions of SR proteins, but is unable to perform the other SR protein functions (Xiao & Manley, 1998). Thus, the regulated exon-dependent and the other SR protein functions not only have different RS domain sequence requirements, but they also have different phosphorylation requirements.

CONCLUDING REMARKS

The last few years have seen several advances in our understanding of SR protein functions. The finding that SR proteins bind to both constitutive and alternative exons provides a nice biological explanation for their rather promiscuous binding specificities. The technological advances that have allowed for the uncoupling of SR protein activities have led to the discovery that different functional RS domain-mediated protein interactions have distinct protein sequence and phosphorylation requirements, and that multiple independent SR proteins participate in the removal of each intron. Hopefully, work over the next few years will be equally fruitful and will help to further sort through some of the confusing areas in the field.

ACKNOWLEDGMENTS

I would like to thank Javier Cáceres, Adrian Krainer, and Juan Valcárcel for communicating results prior to publication. In addition, I would like to thank Javier Cáceres, Klemens Hertel, Kristen Lynch, Tom Maniatis, and members of my laboratory for their useful comments on the manuscript.

Received May 11, 2000; returned for revision June 9, 2000; revised manuscript received June 23, 2000

REFERENCES


Sorting out the complexity of SR protein functions


Sorting out the complexity of SR protein functions.

B R Graveley

RNA 2000 6: 1197-1211