CHAPTER

THE ROLE OF SR AND SR-RELATED PROTEINS IN pre-mRNA SPlicing

William F. Mueller and Klemens J. Hertel*

Department of Microbiology and Molecular Genetics, University of California, Irvine, California, USA

*Corresponding Author  Klemens J. Hertel—Email khetel@uci.edu

Abstract: Pre-mRNA splicing requires the activities of small nuclear ribonucleoproteins and other essential splicing factors. Among these are members of the SR protein family and SR-related proteins, which are integrally involved in regulating exon recognition, spliceosomal assembly, and spliceosomal re-arrangements to promote intron excision. This chapter will focus on the discovery of SR proteins and SR-related proteins, their structural organization, and their activities in mediating pre-mRNA splicing. The importance of SR-proteins in spliceosomal recruitment will be contrasted to other mechanisms resulting in alternative splicing. An additional aspect of the chapter will be a discussion on how reversible phosphorylation of the RS domain dictates SR protein activity and localization.

INTRODUCTION

The splicing of nuclear pre-mRNAs is carried out by the spliceosome, a large dynamic macromolecular complex that recognizes splicing signals and catalyzes the removal of noncoding intronic sequences to assemble protein coding sequences into mature mRNA.1 A critical step in pre-mRNA splicing is the recognition and pairing of 5’ and 3’ splice sites. It is possible to evaluate the formation of the mammalian spliceosome in vitro.2,3 At least four distinct complexes can be resolved by nondenaturing gel electrophoresis in the order of EABBC.4,5 These complexes differ in composition and order of appearance. Based on these findings, a sequential model of spliceosome assembly was developed. This progression requires the activity of more than 150 distinct protein factors and the U1, U2, U4, U5, and U6 small nuclear RNAs (snRNA) complexed with proteins into small nuclear ribonucleoproteins (snRNPs).5 The first intermediate of this assembly reaction
is E complex, which is characterized by ATP independent, stable interactions of U1 snRNP with the 5′ splice site and U2 auxiliary factor (U2AF) with the polypyrimidine tract. ATP hydrolysis then leads to the formation of A complex, which is characterized by the stable association of U2 snRNP with the branch point/3′ splice site sequence and functional commitment to splice site pairing. B complex and the catalytically active C complex form after the incorporation and re-arrangement of the U4•U6/U5 tri-snRNP and facilitate the excision of the intron and ligation of exons.

HISTORY AND DISCOVERY

In addition to snRNPs, pre-mRNA splicing requires other protein factors to efficiently remove introns and ligate exons. One class of regulatory proteins indispensable in this process is the SR protein family. Fractionation and purification experiments of proteins associated with spliceosomes in addition to genetic screens in Drosophila led to the discovery of SR proteins. Fractions from mammalian cell extracts were used in various inactivation and complementation assays to demonstrate that certain proteins, as well as snRNPs, were necessary for splicing. These studies laid the foundation for the classical complementation studies that continue to be used to characterize splicing factors and, eventually, to identify SRSF1 and SRSF2, the most widely studied of the human SR proteins. Sequence analysis of these proteins revealed the presence of extended arginine and serine dipeptides, termed the arginine-serine (RS) domain, in addition to at least one RNA binding domain of the RNA Recognition Motif (RRM)-type. The SR protein involvement in splicing became more obvious when additional studies in Drosophila identified SWAP (suppressor-of-white-apricot), Tra (transformer), and Tra-2 (transformer-2) as promoting specific splicing patterns in mature mRNAs. The identification of these factors, all of which contained RS domains, suggested that RS domain-containing proteins are integrally involved in splice site choice.

The original family of SR proteins was classified following the identification of other RS domain-containing proteins based on: The presence of a phosphoepitope that is recognized by the monoclonal antibody mAb104, their conservation across vertebrates and invertebrates, their ability to precipitate in 20 mM Magnesium Chloride, and their activity in splicing complementation. Later on it was discovered that SR proteins can also function outside of their canonical splicing roles in other aspects of gene expression, such as transcription and mRNA shuttling. Due to the increase in activities and the confusing nature of SR protein classification, the definition of an SR protein has recently been altered to only be based on specific sequence properties: "One or two N-terminal RBDs (PF00076), followed by a downstream RS domain of at least 50 amino acids with >40% RS content, characterized by consecutive RS or SR repeats." In humans, the SR protein family is encoded by twelve genes, designated Serine/Arginine-rich Splicing Factor (SRSF) 1-12 (Table 1). All twelve members of the SR protein family have a common structural organization and fulfill the domain requirement of the definition (Fig. 1).

The classification of SR proteins excludes many proteins that contain RS domains but do not meet the other SR protein classification criteria. These proteins are referred to as SR-related proteins. Bioinformatic surveys have found multiple proteins that contain RS domains, but may have different or no RNA binding domains and may lack the ability to complement splicing reactions (Table 2, Fig. 2). These SR-related proteins are less highly conserved but still may interact with the spliceosome and alter...
### Table 1. List of SR proteins, their current and previous nomenclature, RNA consensus binding sequences, and functions

<table>
<thead>
<tr>
<th>SR Protein</th>
<th>Previous Name</th>
<th>Consensus Binding Sequence</th>
<th>Function</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRSF1</td>
<td>ASF/SF2, SRp30a</td>
<td>RGAAGAAC, AGGACRRAGC, SRSASGA, UGRWG</td>
<td>Constitutive and alternative splicing</td>
<td>51, 56, 204</td>
</tr>
<tr>
<td>SRSF2</td>
<td>SC35, SRp30b</td>
<td>AGSAGAGUA, GUUCGAGUA, UGUUCSAGWU, GWUWCCUGCUA, GGGUAUGCUG, GACGAGUAGKS, AGGAGAU, GRYYCSYR</td>
<td>Constitutive and alternative splicing</td>
<td>52, 56, 81, 82</td>
</tr>
<tr>
<td>SRSF3</td>
<td>SRp20</td>
<td>GGUCCUCUUC WCWWC CUCKUCY</td>
<td>Constitutive and alternative splicing</td>
<td>81, 82, 133</td>
</tr>
<tr>
<td>SRSF4</td>
<td>SRp75</td>
<td>GAAGGA</td>
<td>Constitutive and alternative splicing</td>
<td>149</td>
</tr>
<tr>
<td>SRSF5</td>
<td>SRp40</td>
<td>GAGCAGUCGGCU, ACDGS</td>
<td>Constitutive and alternative splicing</td>
<td>51, 176</td>
</tr>
<tr>
<td>SRSF6</td>
<td>SRp55</td>
<td>USCGKM, UCAACCAGGGCGAC</td>
<td>Constitutive and alternative splicing</td>
<td>51, 92</td>
</tr>
<tr>
<td>SRSF7</td>
<td>9G8</td>
<td>UCAACA ACGAGAGAY GGACGACGAG</td>
<td>Constitutive and alternative splicing</td>
<td>81, 82, 159</td>
</tr>
<tr>
<td>SRSF8</td>
<td>Human SRp46</td>
<td>Not determined</td>
<td>Constitutive and alternative splicing</td>
<td>139</td>
</tr>
<tr>
<td>SRSF9</td>
<td>SRp30c</td>
<td>GACGAC, AAGAGCUCGG, CUGGAAU</td>
<td>Constitutive and alternative splicing</td>
<td>65, 140</td>
</tr>
<tr>
<td>SRSF10</td>
<td>SRp38, SRrp40, hTra2Beta</td>
<td>(GAA)n</td>
<td>Constitutive and alternative splicing</td>
<td>93</td>
</tr>
<tr>
<td>SRSF11</td>
<td>SRp54, p54</td>
<td>C rich regions</td>
<td>General splicing repressor</td>
<td>104</td>
</tr>
<tr>
<td>SRSF12</td>
<td>SRrp35, SRp86</td>
<td>Not determined</td>
<td>Negative of alternative splicing</td>
<td>148</td>
</tr>
</tbody>
</table>

N: any amino acid; Y: pyrimidine; R: purine; D: A,G,U; K: U,G; M: A,C; S: G,C; W: A,U.

SR and SR related tables reviewed in references 25,27,29.
RNA BINDING PROTEINS

Figure 1. Domain organization of SR-proteins. SR proteins identified by their most recent nomenclature (left column), followed by their previous nomenclature (middle column), followed by a cartoon representation of their major domains: RNA Recognition Motifs (RRM), Pseudo RRM domains (pRRM), RS domains (RS), Zinc finger domains (Zn) and linker regions.

Table 2. List of selected SR-related proteins, their current nomenclature and functions.

<table>
<thead>
<tr>
<th>SR Related Protein</th>
<th>Function</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>U170K</td>
<td>Splicing factor</td>
<td>68, 70</td>
</tr>
<tr>
<td>U2AF65</td>
<td>Splicing factor</td>
<td>11</td>
</tr>
<tr>
<td>U2AF35</td>
<td>Splicing factor</td>
<td>153</td>
</tr>
<tr>
<td>U5100K</td>
<td>Splicing factor</td>
<td>43</td>
</tr>
<tr>
<td>DDX46</td>
<td>Spliceosomal re-arrangement</td>
<td>151</td>
</tr>
<tr>
<td>ClkSty1-3</td>
<td>SR protein kinase</td>
<td>150</td>
</tr>
<tr>
<td>Urp</td>
<td>Splicing factor</td>
<td>158</td>
</tr>
<tr>
<td>SRm160</td>
<td>Splicing activator</td>
<td>15</td>
</tr>
<tr>
<td>SRm300</td>
<td>Splicing activator</td>
<td>30</td>
</tr>
</tbody>
</table>
The role of SR and SR-related proteins in pre-mRNA splicing. SR-related proteins function in multiple RNA processing pathways, expanding the SR protein family and increasing the complexity with which splicing can be regulated. By differential binding to RNA elements and differential interactions with other splicing factors through RS domain interactions, SR family proteins influence the way introns and exons are recognized and spliced.

While introns are common to eukaryotes, the complexity of alternative splicing varies among species. Accordingly, SR proteins exist in all metazoan species22 as well as in some lower eukaryotes, such as Schizosaccharomyces pombe.33,34 However, classical SR proteins are not present in all eukaryotes and are apparently missing from Saccharomyces cerevisiae, which also lacks alternative splicing. Instead, three SR-like proteins have been identified in S. cerevisiae, one of which, Npl3, has been shown to modulate the efficiency of splicing.35 Of the SR-related proteins found in higher eukaryotes, some (mostly snRNP associated proteins) have orthologs in yeast that lack RS domains, but many lack yeast orthologs all together.29 In general, the species-specific presence of SR family proteins correlates with the presence of RS domains within other components of the general splicing machinery allowing interaction between them. The observation that the density of RS repeats correlates with the conservation of the branch-point signal, a critical sequence element of the 3' splice site, argues for an evolutionary origin of SR proteins.36 As such, SR family proteins appear to be ancestral to eukaryotes, but were independently lost in some lineages. Phylogenetic tree analyses further suggest that successive gene duplications played an important role in SR protein evolution.37 These duplication events are coupled with high rates of nonsynonymous substitutions that promoted positive selection favoring the gain of new functions, supporting the hypothesis that the expansion of RS repeats during evolution had a fundamental role in the relaxation of the splicing signals and in the evolution of regulated alternative splicing.
STRUCTURAL ELEMENTS AND BINDING SPECIFICITIES OF SR AND SR-RELATED PROTEINS

All SR proteins share two main structural features: The RS domain and at least one RRM (Fig. 1). It is believed that RS domains facilitate protein-protein and protein-RNA interactions and by doing so mediate spliceosomal recruitment, assembly, and function (discussed in the SR family activity section).\textsuperscript{10,38-41} The RS domains of SR family proteins participate in protein interactions with a number of RS domain containing splicing factors.\textsuperscript{38,42} These factors include other SR proteins, SR-related proteins, and components of the general splicing machinery.\textsuperscript{29,38,42-45} Furthermore, the RS domain interacts with RNA\textsuperscript{46} and it can function as a nuclear localization signal by mediating the interaction with nuclear import receptor transportin-SR.\textsuperscript{47-49} These activities can be modulated through post-translational modifications, mainly via phosphorylation (discussed in the SR family activity section).\textsuperscript{29,31}

The presence of RRM domains is the other structural feature defining SR proteins. For the majority of SR proteins with two RNA binding domains, the second domain is a poor match to the RRM consensus and is referred to as an RRM homolog called a pseudo RRM (pRRM). The only exception is SRSF7, which contains an RRM and a zinc-knuckle domain that is thought to contact the RNA.\textsuperscript{50} In the cases where it has been determined, SR proteins have specific, yet degenerate RNA binding specificities.\textsuperscript{51,52} Originally, it was shown that SRSF1 and 2 bind to 5\' and 3\' splice sites to facilitate U1 snRNP, U2AF, and U2 snRNP interactions with the splice sites and with each other in E complex.\textsuperscript{18,53-55} However, this binding was also shown to be low affinity binding\textsuperscript{56} suggesting that there were better targets for SR proteins outside of splice sites. Soon after this, purine-rich consensus sequences were found that displayed higher affinity binding for SRSF1 and 2.\textsuperscript{56} These purine-rich sequences were discovered through mutational analysis as exonic splicing enhancers (ESEs), exonic sequences that activate specific exon inclusion.\textsuperscript{57-59} It was soon found that SR proteins promiscuously interact with multiple different ESEs,\textsuperscript{59-65} demonstrating that they have degenerate binding specificity (see the SR protein binding section). More recently, chromatin immunoprecipitation experiments were used to show that SR proteins are recruited cotranscriptionally and that the site of recruitment to the nascent RNA is mainly dictated by their RRM interactions with the pre-mRNA.\textsuperscript{66}

SR-related proteins have only an SR domain to define them (Fig. 2). Many SR-related proteins lack RRMs all together and instead have domains that mediate interactions with other cellular machineries or contain domains that modify other RNA processing machineries (such as kinases with SR domains).\textsuperscript{29,31} In the case of the SR-related protein U1-70K, the small RS domain is conserved from \textit{Drosophila} to humans, but not in yeast.\textsuperscript{67,71} The RRM of U1-70k binds to U1 snRNA through a conserved 8 amino acid sequence that was found in multiple RNA binding proteins.\textsuperscript{72} U1 snRNP binds to 5\' splice sites through RNA/RNA and protein/RNA interactions aided by SR proteins through U1-70k’s RS domain. Additionally, U1-70k’s RS domain assists in bridging splice sites during early spliceosomal steps.\textsuperscript{38,42} Despite recent structural analyses revealing extensive interaction throughout U1 snRNP, the RS domain binding abilities have yet to be characterized structurally.\textsuperscript{73,74}

For 3\' splice site recognition, the 65 and 35 kD subunits of U2AF mediate polypyrimidine tract binding and subsequent interaction with U2 snRNP.\textsuperscript{38,75-77} Through their interaction with the pre-mRNA, both of these essential SR-related proteins allow for snRNP and other SR protein recruitment. In an interaction reminiscent of U2AF 35 and 65
scenario, the *Drosophila* Tra protein, which has no RRM, must interact with Tra 2, which has an RRM, to bind to the *doublesex* pre-mRNA.\textsuperscript{78} Other RS-related proteins contain DExD/H-box ATPase/helicase domains, CTD interacting domains, kinase domains, and many others, all important in modulating splicing activity.\textsuperscript{29}

Few SR-related or SR proteins have been characterized structurally, mainly due to solubility issues, probably involving the large exposed hydrophobic regions. Unfortunately, no structural information detailing the RS domain is available to date. This may be explained by the poor solubility of these proteins in their free state and the unknown phosphorylation state of the serines within the RS domain. In addition, the degenerate RNA-binding sequences recognized by SR proteins may have prevented their study in the bound form. Consequently, only isolated RRMs of SR proteins have been analyzed structurally by nuclear magnetic resonance spectroscopy. To tackle the solubility issues, the RRMs of SRSF3 and SRSF7 were fused to the immunoglobulin G-binding domain \textsuperscript{1} of Streptococcal protein G (GB1) solubility tag\textsuperscript{79} or overexpressed RRMs were suspended in a solution containing charged amino acids.\textsuperscript{80} Using these manipulations it was possible to obtain solution structures of the free SRSF7 and SRSF3 RRMs and of the SRSF3 RRM in complex with the RNA sequence 5’-CAUC-3’. When analyzing the unbound RRMs of SRSF3 and SRSF7, one is struck by an unusually large exposed hydrophobic surface, which could explain why the solubility of SR proteins is so low. The SRSF3 RRM complex with RNA shows that although all four nucleotides present are contacted by the RRM only the 5’ cytosine is recognized in a specific manner. These structural insights provided a potential explanation for the seemingly low RNA binding specificity exhibited by SRSF3 \textsuperscript{81,82}

RS domain activities are regulated through phosphorylation, and thus require kinase interaction (see the SR protein regulation section). In some cases this could be mediated through RS domains found in certain SR protein kinases (Clk/Sty proteins 1-3), but in the case of SRPK1, interactions are established through a unique docking groove to SRSF1 directing phosphorylation to specific areas of its RS domain. This docking groove seems to cancel out the charges on the RS domain with opposing charges allowing deep and tight binding and, eventually, allowing processive phosphorylation.\textsuperscript{83-86} These observations suggest a structurally related regulatory mechanism for phosphorylation of RS domains, allowing specific kinases to recognize specific RS domains based on specific docking motives. However, due to the problems inherent to crystallizing RS domain containing proteins, it seems that more data regarding how RS domains interact will be long in coming.

**SPLICING ACTIVITY OF SR PROTEINS**

**Splicing Activation**

For classical cases of alternative splicing, it was shown that cis-acting RNA sequence elements increased exon inclusion by serving as binding sites for the assembly of multi-component splicing enhancer complexes. For example, a specific exonic sequence recruited Tra, Tra 2, and RBP1 (the *Drosophila* homolog of SRSF3) to form a heterotrimeric complex regulating sex specific alternative splicing of the *Drosophila doublesex* gene.\textsuperscript{60} Additional studies suggested that the RRMs directed SR proteins to sequences on the pre-mRNA through which they could interact with other proteins and influence splicing.\textsuperscript{59,62,87} These RNA sequence elements were usually located within the
regulated exon leading to their definition as ESEs.\textsuperscript{1,88} In general, ESEs are recognized by at least one member of the SR protein family and recruit the splicing machinery to the adjacent intron.\textsuperscript{1,27,88} This recruitment has been shown to activate weaker splice sites or to alter splice site choice.\textsuperscript{15,16} Surprisingly, SR protein binding sites are not only limited to alternatively spliced exons, but they have also been verified for exons of constitutively spliced pre-mRNAs.\textsuperscript{89,90} It is therefore likely that SR proteins bind to sequences found in most, if not all, exons and are thus active in the greater majority of all exon recognition.

It was soon discovered that the same SR proteins could display slightly different binding sequences in different contexts, thus new methods were established for finding ESEs. Using a modified binding SELEX technique,\textsuperscript{91} degenerate consensus RNA sequences were identified that interact with SRFS1 and 2.\textsuperscript{56} Subsequent enrichment approaches were based on the ability of an SR protein to activate intron removal.\textsuperscript{51} These functional SELEX techniques were used to generate binding consensus sequences for all SR proteins and for some SR-related proteins.\textsuperscript{81,82,89,92,93} In combination with the deciphering of the human genome, this information allowed global computational analyses of pre-mRNA sequences to determine the density of ESEs within the genome.\textsuperscript{94,95} The results from these bioinformatic approaches showed that potential SR protein binding sites are found in far more places than previously thought, supporting the notion that SR protein activity is paramount for splice site selection throughout the human genome.

Using recombinant SR proteins to complement splicing deficient cell extracts it was demonstrated that SR proteins functionally overlapped in many cases.\textsuperscript{22,50,96,97} SRSF1 and 2 cannot replace each other and they bind to different ESEs.\textsuperscript{50,98} RS domain swapping experiments further displayed different activation potentials or even splicing repression (see also the Splicing Repression section below).\textsuperscript{99} The strongest argument for nonredundant functions in vivo originate from the early embryonic lethal phenotypes of SR protein knock out mice, clearly demonstrating the developmental significance of these splicing factors in different pathways.\textsuperscript{25}

SR proteins facilitate the binding of regulatory complexes to the pre-mRNA and promote splicing through mediating the recruitment of prespliceosomal or E complex components such as U1 snRNP, and the proteins SF1 and the U2AF heterodimer (Fig. 3A). Multiple SR proteins have been shown to increase the kinetics of E complex formation.\textsuperscript{42,100} It was even found that the need for U1 snRNP binding to the 5' splice site could be bypassed in the presence of excessive amounts of recombinant SR proteins.\textsuperscript{101,102} In addition to their enhancement of splice site recognition, it was demonstrated that SR proteins bridge across introns or exons to facilitate splice site pairing\textsuperscript{18,38} through interacting with components from both U1 and U2 snRNPs.\textsuperscript{103,104} These splice site pairing interactions can be recapitulated in trans-splicing reactions, in which SR proteins also play an essential role.\textsuperscript{55,105-109}

The action of SR-mediated splice site recognition does not go unchecked. In addition to splicing enhancers, sequences have been identified that decrease exon inclusion efficiency (referred to as splicing silencers). In general, hnRNPs or other similar splicing inhibitory proteins recognize these RNA elements.\textsuperscript{110-112} Competition between the activities of SR proteins and hnRNP A1 have been implicated in the regulation of HIV TAT exon\textsuperscript{3113} as well as IgG exons 1 and 2\textsuperscript{77} pointing to a yin/yang model of splicing regulation. While SR proteins generally promote the formation of functional spliceosomes, hnRNPs or other proteins generally act against exon recognition. Interestingly, the role of SR proteins as activators or as inhibition blockers seems to change depending on the type of splicing
THE ROLE OF SR AND SR-RELATED PROTEINS IN pre-mRNA SPLICING

occurring, alternative or constitutive. In vitro tests of mutant SR proteins have shown that the RS domain is dispensable for certain alternative splice site selection events, but necessary for constitutive splicing.\textsuperscript{114-116} This data supports the idea that SR proteins recruit the spliceosome to activate alternative splice sites via their RS domains, and that they also inhibit repressive splicing mechanisms through the substrate specific binding of their RRM domains. These two modes of activity are not mutually exclusive as evidence points to SR proteins having multiple context dependent roles.\textsuperscript{117}

Figure 3. Pre-mRNA splicing activities of SR proteins. A) SR proteins assist in the recruitment of spliceosomal complexes to the pre-mRNA through RRM interaction with the RNA and RS domain interaction with other proteins. SR-related proteins can display equivalent recruitment functions (not shown). B) Directional activity of SR proteins. SR proteins can simultaneously exhibit activating and inhibitory functions based on their orientation relative to competing splice sites. Because these activities require interactions of SR proteins with the regulated exon (A, B), they are referred to as exon-dependent actions of SR proteins.\textsuperscript{51-128,134-149} C) SR mediated recruitment of the tri-snRNP and stabilization of previously recruited snRNPs. These interactions are essential for spliceosomal re-arrangements and proper function. SR-related proteins such as SRm160/300, DDX46, and others function similarly (not shown). D) SR-mediated splicing repression can occur when SR proteins bind with high affinity to the intron. The precise mechanism of this inhibition is unknown. Because these SR protein activities do not require interactions with the regulated exon (C, D),\textsuperscript{129-132} they are referred to as exon-independent actions of SR proteins.\textsuperscript{5,15,16}
An alternative mode of spliceosomal recruitment was suggested by experiments demonstrating that the RS domains of U2AF65 and SRSF1 contact the branch point and 5′ splice site in the mRNA. As these contacts are established within the functional spliceosome, it is likely that SR family proteins contact the pre-mRNA throughout the splicing process. Modulation of RNA binding by the RS domain seems to make SR protein binding less specific and more promiscuous. Irrespective of the RS domain activation mode, SR proteins facilitate the recruitment of spliceosomal components to the regulated splice site. Thus, SR proteins bound to ESEs function as general activators of exon definition. Kinetic analyses demonstrated that the relative activity of ESE-bound SR proteins determine the magnitude of splicing promotion. This activity depended on the number of SR proteins assembled on ESEs and the distance between ESEs and the intron. It was also shown that splicing activation was proportional to the number of serine-arginine repeats contained within the RS domain of the bound SR protein. Thus, the quantity of serine-arginine repeats appears to dictate the activation potential of SR proteins.

In cases of alternative splicing, the activation by SR proteins was found to promote the pairing of the most proximal splice sites across the intron. Interestingly, SR proteins do not do this by simply enhancing recognition of the proximal splice site alone, they also inhibit the use of the distal splice site through yet unknown mechanisms. This directional activation of splice sites seems to depend on splice site strength and is probably modulated by the presence of other enhancing or silencing signals. Even in the presence of SR proteins, neighboring splice sites play a significant role during splice site selection, as demonstrated in cases where a nearby 5′ splice site increases the usage of another 5′ splice site.

EXON INDEPENDENT ACTIVATION

Characterization of SRSF1′s activity showed that 5′ splice site cleavage, branch point choice, and lariat formation were all affected by its presence. These observations pointed to SR protein activity in more than just the initial splice site recognition and pairing steps of splicing. Similar results were obtained with other SR proteins. Mechanistic hints as to why SR protein action is important for later spliceosomal assembly steps came soon after it was found that the U4/U6•U5 tri-snRNP required SR proteins to be incorporated into the spliceosome. Because tri-snRNP incorporation succeeds initial splicing steps, no exonic sequences were found to be specifically required. In fact, SR proteins have essential activities that do not require interactions with exon sequences to initiate the first step of splicing. The role of the exon-independent function may be to promote the pairing of 5′ and 3′ splice sites across the intron or exon, facilitate cross talk between RNA processing events and/or to facilitate the incorporation of the U4/U6•U5 tri-snRNP into the spliceosome. While the RRM is essential for its exon-independent activity, it is likely that SR proteins interact with the partially assembled spliceosome or the tri-snRNP through RS domain contacts.
SPLICING REPRESSION

One striking feature of SR proteins is their prevalent location within the pre-mRNA. In nearly all cases SR proteins have been found to interact with exonic sequences of the pre-mRNA. This is a surprising finding considering the fact that their relatively promiscuous binding specificity predicts that introns are littered with potential SR protein binding sites. The fact that SR proteins are mainly observed to bind within exonic sequences suggests that additional requirements need to be met for functional SR protein binding to the pre-mRNA. However, some SR proteins do bind within the intron where they generally function as negative regulators of splicing (Fig. 3D). The best-characterized example of this is observed during adenovirus infection. In this case, splicing is repressed by the binding of the SR protein SRSF1 to an intronic repressor element located upstream of the 3' splice site branchpoint sequence in the adenovirus pre-mRNA. When bound to the repressor element, SRSF1 prevents the recruitment of U2 snRNP to the branchpoint sequence, thereby inactivating the 3' splice site. Other studies provided further support for the idea that SR proteins placed within introns generally interfere with the productive assembly of spliceosomes. SRSF2 interacts with an intronic sequence of the Beta-tropomyosin gene antagonizing the action of SRSF1 in the recognition of exon 6A. These observations suggest that expression levels of SR proteins may be very important in regulating splicing decisions, as the balance between competing SR protein binding sites, exonic and intronic, will affect the splicing outcome.

Some SR proteins seem to generally exhibit inhibitory actions or consistently oppose the "canonical" SR protein functions. SRSF11 competes with human Tra2β to promote exon skipping and distal splice site activation. In addition, SRSF5, 10, and 12 have also been implicated in negative regulation of splicing, although SRSF10 acts only after dephosphorylation in response to mitosis or cellular stress (see SR protein regulation below). SRSF12 was found to interact with all the SR proteins in both enhancing and suppressing contexts. Curiously, while SRSF12 modulates SR protein function, its activity may also be modified by interactions with SR proteins and hnRNPs. These observations demonstrated that SR proteins, through context dependent binding of RNA and context dependent protein interactions, act to facilitate spliceosomal complex formation, as barriers to prevent exon skipping, or as inhibitors of splice site usage.

SR-RELATED PROTEIN ACTIVITY

Due to the multitude of RS domain containing proteins, actions of SR-related proteins influence many cellular processes (Fig. 4). They are involved in bridging splicing reactions, regulation of other RS domain containing proteins, and are components in the main splicing machinery itself. Even though they are not studied as extensively as SR proteins, it is appreciated that their functions in splicing are important and, in many cases, necessary.

Like SR proteins, SR-related proteins act early in splicing to aid exon recognition and complex formation. The best studied of these early acting splicing proteins are components of the early spliceosomal complexes themselves: U1-70k and both subunits of U2AF. Multiple SR proteins interact with these SR-related proteins to promote their recruitment to splice sites and interaction across introns and exons. U1 snRNP has been shown to stabilize pre-mRNAs and to influence splice site choice.
RNA BINDING PROTEINS

competes for pre-mRNA binding with other polypyrimidine tract binding proteins that interfere with spliceosome assembly. Interestingly, multiple genes exist that have a high similarity to U2AF. Presumably, these U2AF related proteins function in subsets of pre-mRNA splicing, providing unique ways to identify and interact with the 3′ splice site.

The SR-related proteins SWAP, Tra, and Tra 2, were among the first proteins implicated in alternative splicing regulation and their human homologues have been shown to have similar activities. SRm160/300 is a large matrix protein that interacts with multiple RNA interaction proteins with multiple RS domains and contains a PWI RNA recognition domain. It is critical for bridging splice sites on certain pre-mRNAs and has been found to complex with transcription related proteins and chromatin support proteins. These observations suggest a mechanistic link between splicing and transcription.

Some SR-related proteins have also been shown to regulate SR protein activity by directing their phosphorylation via kinase domains. One of the best characterized of these are the Clk/Sty protein kinases. Clk/Sty 1 interacts with SR proteins and phosphorylates their RS domains, thus altering their activity and localization (see the SR protein regulation section).

Similarly, the hPrp4 kinase interacts with components of the tri-snRNP and phosphorylation of U4/U6 associated PRP31 and U5 associated PRP6 is necessary for successful transition into spliceosomal B complex. The U4/U6•U5 tri-snRNP contains many more RS domain containing proteins, all of which are potential targets for similar regulation mechanisms.

Figure 4. The functional repertoire of SR-related proteins in pre-mRNA splicing. SR-related proteins have been shown to participate in spliceosomal complex formation, phosphorylation of RS domains, splice site recognition, exon inclusion and the coupling of RNA processing events (transcription, polyadenylation, capping, nuclear export). In addition, some SR-related proteins have essential helicase or ATPase functions.
SR PROTEIN REGULATION AND LOCALIZATION

SR proteins are expressed throughout development in most tissues and organs. To control the production of the appropriate transcriptome, SR proteins are regulated at the translational and post-translational levels, either through self-regulation (Sxl and SWAP), or through interactions with other SR family proteins. In all of these cases, the control over SR protein expression is necessary for maintaining efficient splicing in all cell types.

Like other proteins involved in pre-mRNA splicing, SR proteins are enriched in nuclear compartments, termed speckles. Speckles consist of two distinct structures: Interchromatin granule clusters (IGCs), storage/re-assembly sites for pre-mRNA splicing factors that are 20-25 nm in diameter, and perichromatin fibrils (PFs), the site of actively transcribing genes and cotranscriptional splicing that are approximately 5 nm in diameter. The SR proteins are one prominent component of nuclear speckles and biochemical analyses have indicated that RS domains are responsible for targeting the SR family proteins to speckles and that this localization can be slightly different for each SR protein. Because the nuclear organization of SR proteins is dynamic, SR proteins are recruited from IGC storage clusters to the site of cotranscriptional splicing (PFs). Interestingly, both the RNA binding domains and RS domains are required for the recruitment of SR proteins from IGCs to PFs, as is phosphorylation of the RS domain.

SR proteins act at several steps during the splicing reaction and require phosphorylation and dephosphorylation for spliceosomal assembly. It was found that hyper- or hypo- phosphorylation inhibited splicing. This is evident in the case of SRSF1 where phosphorylation is required for interactions with U1-70k, yet dephosphorylation is required for splicing to continue to catalysis. A number of SR protein kinases have been shown to specifically phosphorylate serine residues within the RS domain of SR proteins. This seems to occur in a temporal and region specific manner, i.e., certain kinases prefer to act on certain areas of the RS domain with different specificities at different times. These include SR protein kinase 1 (SRPK1), Clk/Sty kinase (see the SR related protein activity section), cdc2p34 and topoisomerase. SRPK1 has been crystallized in complex with the docking domain of SRSF1 and seems to preferentially phosphorylate the N-terminal segment of the RS domain. These two examples allow a simplified view of how SRSF1 may be regulated (Fig. 5).

It has been shown that SR proteins may be regulated by their localization within the cell. The phosphorylation state of the RS domain triggers the shuttling of SR proteins between the cytoplasm and the nucleus (Fig. 5). From their location in nuclear speckles, phosphorylated SR proteins perform their functions in splicing after which a subset of SR proteins are active in mRNA export as adaptor molecules. Rephosphorylation is required for these shuttling SR proteins to return to the nucleus to carry out their nuclear splicing functions (Fig. 5). Several kinases found in the cytoplasm are responsible for this process, supporting an SR protein localization cycle that is intimately linked to mRNA production and the cell cycle. Unique pathways of this cycling exist in response to heat shock, cellular growth, and signaling to change SR protein concentrations within the nucleus, thus modulating mRNA splice patterns.
CONCLUSION

The SR protein family engages in ubiquitous and promiscuous interactions throughout the splicing pathway. Understanding of the action and regulation of these proteins seems daunting in face of all their cellular activities, which extends beyond those mentioned in this chapter. Global analyses of SR proteins have fueled the ability to predict potential binding sites of SR proteins.\textsuperscript{95,199–202} While the resulting computational tools have helped to suggest location from which SR proteins function, it has also lead to an embarrassment of riches, with SR binding sequences found frequently throughout the genome. When these results are compared with the most recent attempts to understand splicing through an analysis of cis-acting elements, the ‘splicing code’, it appears that the presence of SR protein binding sites are not as important to exon inclusion as many other RNA elements investigated.\textsuperscript{203} Using CLIP-seq techniques it was further demonstrated that some programs to predict SR protein binding sites may be over ambitious.\textsuperscript{204} It is clear that SR proteins are necessary for accurate splicing to occur, presumably within a subset of unique pre-mRNAs. The challenge to further understand the actions of this diverse family of proteins

Figure 5. A simplified model of phosphorylation and cellular localization of SR proteins. SR proteins require phosphorylation to mediate spliceosomal complex formation. However, dephosphorylation is required for splicing catalysis to occur, as well as nuclear export. Re-entry into the nucleus requires rephosphorylation where the SR protein can act again in the splicing pathway. As most SR proteins shuttle between the nucleus and cytoplasm, this cycle may provide a simple mode for functional regulation. Dephosphorylation of SR proteins leads to sequestration to the cytoplasm, altering SR protein levels in the nucleus and changing splicing patterns. By contrast, hyperphosphorylation of SR proteins represses splicing.\textsuperscript{169–198}
will rest on the continued analysis of actual binding sites using genome-wide techniques. This sort of global analysis may help to define different situations where SR proteins are most active and important. While the current understanding of this family of proteins is taking a shape that allows sufficient understanding of many splicing and splicing-related disease phenotypes, there is still a great amount of work to be done to create a more complete picture of their activities and regulation during pre-mRNA splicing.

ACKNOWLEDGEMENTS

The authors wish to acknowledge support from the NIH (RO1 GM62287 and R21 CA149548 to K.J.H.).

REFERENCES

70. Query CC, Bentley RC, Keene JD. A common RNA recognition motif identified within a defined U1 RNA binding domain of the 70K U1 snRNP protein. Cell 1989; 57(1):89-101.


93. Staknis D, Reed R. SR proteins promote the first specific recognition of pre-mRNA and are present together with U1 snRNP in a general splicing enhancer complex. Mol Cell Biol 1994; 14:7670-7682.


THE ROLE OF SR AND SR-RELATED PROTEINS IN pre-mRNA SPLICING


182. Cao W, Jamison SF, Garcia-Blanco MA. Both phosphorylation and dephosphorylation of ASF/SF2 are required for pre-mRNA splicing in vitro. RNA 1997; 3(12):1456-1467.