Review TRENDS in Plant Science Vol.9 No.11 November 2004

Plant serine/arginine-rich proteins and their role in pre-mRNA splicing

Anireddy S.N. Reddy

Department of Biology and Program in Cell and Molecular Biology, Colorado State University, Fort Collins, CO 80523, USA

Pre-messenger RNA (pre-mRNA) splicing, a process by which mature mRNAs are generated by excision of introns and ligation of exons, is an important step in the regulation of gene expression in all eukaryotes. Selection of alternative splice sites in a pre-mRNA generates multiple mRNAs from a single gene that encode structurally and functionally distinct proteins. Alternative splicing of pre-mRNAs contributes greatly to the proteomic complexity of plants and animals and increases the coding potential of a genome. However, the mechanisms that regulate constitutive and alternative splicing of pre-mRNA are not understood in plants. A serine/arginine-rich (SR) family of proteins is implicated in constitutive and alternative splicing of pre-mRNAs. Here I review recent progress in elucidating the roles of serine/arginine-rich proteins in pre-mRNA splicing.

The seminal discovery in 1977 that eukaryotic genes are split into coding (exons) and noncoding (introns) sequences added another step to the process of gene regulation: introns in the primary transcripts are precisely excised and exons are spliced in the nucleus to generate functional mRNAs [1,2]. In ensuing years it has been found that primary transcripts with multiple introns can undergo splicing in many different ways leading to the generation of multiple proteins from a single gene. Furthermore, pre-mRNA splicing appears to imprint the information necessary for nucleocytoplasmic transport of spliced mRNA, nonsense-mediated mRNA decay and mRNA localization in the cytoplasm [3,4]. Multicellular organisms regulate splicing of pre-mRNAs depending on cell type, developmental state or external stimulus. Studies from non-plant systems have established that transcription, RNA processing, and mRNA transport are tightly coupled and function as an 'mRNA factory' [5]. Many proteins involved in mRNA processing also interact with transcription and mRNA export machinery.

Alternative splicing

Alternative splicing generates multiple proteins with different functions from a single gene and greatly enhances the coding potential of a genome. Multiple transcripts from a single gene can result from exon skipping, mutual exclusion of exons and retention of introns and/or selection of an alternative 5′ or 3′ site. Alternative splicing can affect the stability and translatability at the RNA level and produce truncated or extended proteins with altered (increased, decreased or loss of) activity, cellular localization, regulation, and/or stability. Comparison of completely sequenced genomes with full-length cDNAs and many ESTs coupled with genome-wide analysis of alternative splicing with exon junction microarrays has revealed that a high percentage of eukaryotic genes undergo alternative splicing [6,7]. At least 74% of human multi-exon genes undergo alternative splicing [7]. Brian Haas *et al.* [8] identified a total of 1188 genes in *Arabidopsis* that produce 2501 alternatively spliced transcripts. The actual number of alternative spliced genes in *Arabidopsis* is likely to be far greater than this because the current analysis was performed with a limited number of ESTs and full-length cDNAs (0.2 million in *Arabidopsis* and 5 million in human). Half of the genes encoding splicing variations are because of alternative donor and/or acceptor sites for the same intron (http://www.tigr.org/tdb/e2k1/ath1/altsplicing/splicing_variations.shtml). Retention of introns in plants appears to be fairly common. About 15% of alternatively spliced transcripts have introns retained either in their untranslated regions (190) or in the protein coding region (218 genes) or both (22) [8]. It is not clear whether the unspliced introns represent inefficient splicing or regulated alternative splicing to generate different transcripts and proteins with functional significance. These cytosolic mRNAs with retained introns are not removed by RNA surveillance mechanisms in the cytoplasm and the conservation of intron retention across species in some cases indicates a biological role for these RNAs. Of the 218 variants that retain introns, 99 encode truncated proteins and the rest encode variant proteins, suggesting a role for unspliced introns in generating proteomic complexity [8]. The truncated proteins can have dominant-negative effects.

Functional significance of alternative splicing

Although transcripts of many plant genes, including many disease resistance genes, are alternatively spliced, the functional significance is known for only a few genes [9,10]. The presence of both constitutively and alternatively spliced transcripts of the N gene in tobacco and *RPS4* in *Arabidopsis*, which confer resistance to TMV and *Pseudomonas syringae* pv tomato strain DC3000, respectively, was shown to be necessary for the function of these genes [11,12]. The *FCA*, which encodes an RNA binding nuclear protein and promotes the floral transition, produces four transcripts (α, β, γ and δ) by alternative

Corresponding author: Anireddy S.N. Reddy (reddy@lamar.colostate.edu).

 1360-1385/$ - see front matter doi:10.1016/j.tplants.2004.09.007

processing [13]. The FCA negatively regulates processing of its own transcript and this autoregulation prevents precocious flowering [14]. There are other instances in plants where alternative splicing regulates enzyme activity or cellular localization of a protein [9].

Pre-mRNA splicing machinery

Pre-mRNA splicing takes place in the spliceosome, a large RNA-protein complex, by two successive transesterification reactions [15]. The assembly of the spliceosome involves a series of RNA–RNA, RNA–protein and protein–protein interactions. Studies with metazoans have revealed the existence of two types (a major U2-type and a minor U12-type) of spliceosome that differ in their composition. Plants have both U2 and U12 type introns [16,17]. A recent large-scale comprehensive analysis of the human major spliceosome indicates that it consists of >300 distinct proteins [18,19]. The major spliceosome contains five small nuclear ribonucleoprotein particles (snRNPs) – U1, U2, U4, U5, U6 – and many non-snRNP splicing factors. The snRNPs recognize splice sites and branch point sequences in pre-mRNA and aid in splicing [15]. The first step in the assembly of the spliceosome is the recognition of the 5′ splice site by U1 snRNP to form a complex called the early (E) complex in mammalian cells that commits the pre-mRNA to spliceosome assembly [20]. In metazoans, U1 snRNP contains one U1 snRNA molecule and at least 11 proteins, including three U1snRNP-specific proteins (U1-70K, U1-A and U1-C) [21]. U1-70K interacts with other snRNP complexes including U2 snRNP [22]. U1-A and U1-70K have been characterized from plants [23–25] and U1-C has been found in the *Arabidopsis* genome database (http://www.*Arabidopsis*.org).

Many components of major and minor spliceosomes are present in plants [9,26] (http://www.life.umd.edu/labs/mount/factors/), suggesting that plant spliceosomes might have a composition that is similar to animal spliceosomes. However, important differences exist in the recognition of plant introns [9,26–28]. Plant 5′ and 3′ splice sites are more variable and the branch point is less well defined compared with the yeast and metazoans [29]. Another distinguishing feature of plant introns is their compositional bias for UA- or U-rich sequences, which is important for recognition of 5′ and 3′ splice sites and for efficient splicing of introns in plants [9,17,26,27,30,31]. A few proteins that specifically interact with U-rich sequences have been reported [32]. Overexpression of

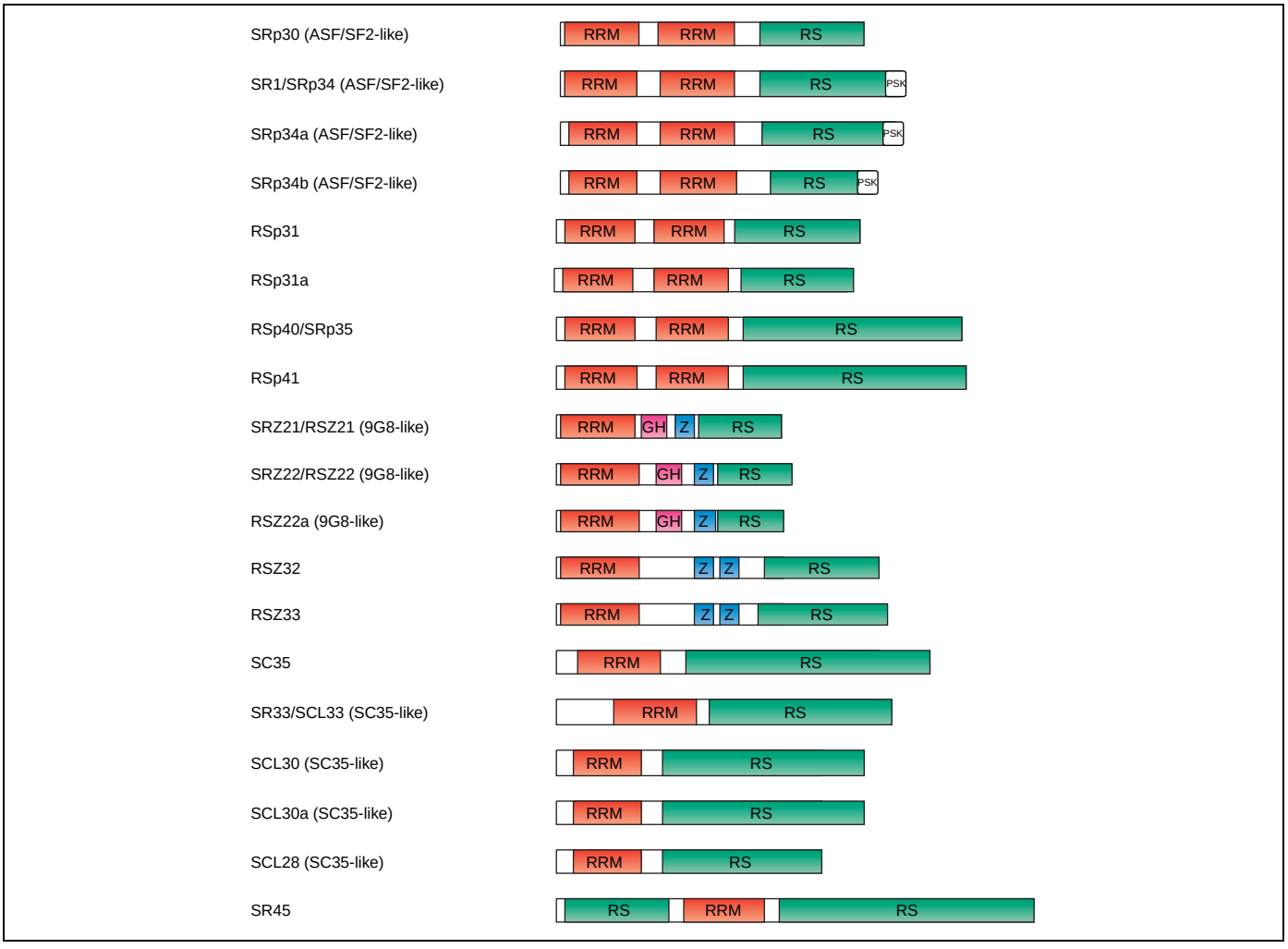

Figure 1. Structural features of *Arabidopsis* serine/arginine-rich (SR) proteins. Abbreviations: GH, glycine hinge; PSK, a domain rich in proline, serine and lysine; RRM, RNA recognition motif; RS, arginine/serine-rich domain; Z, zinc knuckle. Adapted, with permission from, Ref. [80].

one of these proteins (UBP) in protoplasts enhanced the splicing of pre-mRNAs that are otherwise inefficiently processed. The importance of the GC content in exons in efficient splicing, and an AG-rich exonic element in promoting downstream 5′ splice site selection, has also been reported in some cases in plants [33–35]. Kazuya Yoshimura *et al.* [36] have identified a *cis*-element in chloroplast ascorbate peroxidase that regulates alternative splicing.

Serine/arginine-rich proteins

The serine/arginine-rich proteins (SR proteins) are a highly conserved family of structurally and functionally related non-snRNP proteins with multiple roles in pre-mRNA splicing [37,38]. These proteins have modular domain structure with one or two RNA recognition motifs (RRMs) at the N-terminus and an arginine/serine-rich (RS) domain at the C-terminus. All SR proteins are phosphoproteins and are recognized by a monoclonal antibody (mAb104) [39]. SR proteins in animals function as essential splicing factors in constitutive pre-mRNA splicing and also regulate alternative splicing by influencing the splice site selection in a concentration-dependent manner [22]. Some animal SR proteins that shuttle between the nucleus and cytoplasm function in mRNA export, mRNA stability or translation [40–42]. SR proteins participate in RNA–protein and protein–protein interactions during spliceosome assembly [43]. In multicellular organisms, splice sites are degenerate, hence the splice site selection depends on additional sequence elements such as exonic splicing enhancers or silencers (ESEs/ESSs) or intron splicing enhancer or silencers (ISEs/ISSs). These sequence elements recruit SR proteins, which act as adapters and recruit other components of the splicing machinery. The RRM, which confers RNA-binding specificity, binds to specific regulatory sequences in pre-mRNA, and the RS domain functions as a splicing regulator by mediating protein–protein interactions with other components of the splicing machinery. SR proteins show functional redundancy *in vitro* because each single SR protein can restore splicing activity to a splicing-deficient cytoplasmic extract [37]. *In vivo* studies with non-plant systems indicate that some SR proteins are redundant whereas others are not [38].

The family of SR proteins in *Arabidopsis* comprises at least 19 proteins that range in size from 21 to 45 kDa [44] (http://www.plantgdb.org/PlantGDB/prj/SiP/SRGD/ASRG/Display.php?GID=2.2). The presence of either full-length cDNAs or ESTs for each SR gene suggests that they are all expressed. Some plant SR proteins appear to be orthologs of metazoan SR proteins whereas others are unique to plants with novel structural features. Four are similar to SF2/ASF but some contain a C-terminal extension that is rich in proline, serine and lysine (PSK domain). Another five (SC35 and SCLs) are similar to SC35, three are like 9G8 in humans and the rest (seven) have no counterparts in animal systems (Figure 1, Table 1). Six SR proteins in *Arabidopsis* have arisen by gene duplication [45]. Unlike most plant SR proteins that contain a single RS domain, SR45 has two distinct RS domains, one on either side of the RRM (Figure 1). SR45 has no orthologs in metazoans but is highly conserved in other plants. In rice there are at least two SR45 proteins. A few plant SR proteins have been shown to complement splicing-deficient S100 extract or influence splicing *in vivo* [46–48], whereas SR1 did not complement HeLa cells but promoted splice site switching in mammalian nuclear extracts [49].

Table 1. Alternative splicing and localization of *Arabidopsis* serine/arginine-rich (SR) proteins

Gene ID	Alternative splicing[a]	Size[b]	Localization
At1g09140 (SRp30)	Yes, 3	268	NP, S
At1g02840 (SR1/SRp34)	Yes, 7	303	NP, S
At3g49430 (SRp34a)	ND	300	ND
At4g02430 (SRp34b)	ND	278	NP, S
At3g61860 (RSp31)	Yes, 1	264	NP, S
At2g46610 (RSp31a)	Yes, 2	250	ND
At4g25500 (RSp40/RSp35)	Yes, 1	350	NP, S
At5g52040 (RSp41)	ND	356	ND
At1g23860 (SRZ21/RSZ21)	No	187	NP, S
At4g31580 (SRZ22/RSZ22)	No	200	NP, S
At2g24590 (RSZ22a)	ND	196	ND
At3g53500 (RSZ32)	Yes, 1	284	ND
At2g37340 (RSZ33)	Yes, 4	290	NP, S
At5g64200 (SC35)	ND	303	NP, S
At1g55310 (SR33/SCL33)	Yes, 4	287	NP, S
At3g55460 (SCL30)	Yes, 1	262	NP, S
At3g13570 (SCL30a)	Yes, 1	261	NP, S
At5g18810 (SCL28)	ND	236	NP, S
At1g16610 (SR45)	No	414	NP, S

Abbreviations: NP, nucleoplasm; S, speckles; ND, not determined.
[a]Number of alternatively spliced isoforms.
[b]Several of these proteins have alternatively spliced forms. The size of only the longest protein is shown for each gene.

Alternative splicing of SR proteins and regulation of this process by SR proteins

Northern, RT–PCR and promoter analyses have revealed differential expression of SR proteins in different tissues and cell types [48,50–55]. Most SR proteins showed distinct as well as overlapping expression patterns. Moreover, expression of some SR proteins is developmentally regulated. It is interesting that pre-mRNAs of several plant SR genes and other spliceosomal proteins produce multiple transcripts by alternative splicing (Table 1). Several alternatively spliced forms contain a premature stop codon and produce truncated proteins that lack one or more domains [50,53–55]. SRp30 produces at least three different transcripts and the level of these isoforms is controlled in a temporal and spatial manner [55]. SR1 pre-mRNA generates at least seven products [45,50]. Only one alternatively spliced form codes for the full-length protein and the rest code for proteins that differ in the RS domain. Additionally, the alternative splicing pattern of SR1 is regulated in an organ- and tissue-specific manner and by temperature. Results from transient gene expression assays indicate that alternative splicing of SR1 is not an autoregulatory mechanism used to control the transcript level of the full-length protein [50].

Overexpression of SRp30 influenced the splice site choice and changed the splicing pattern of its own pre-mRNA and other endogenous plant genes (RSp31, SRp34, U1-70K). High levels of SRp30 greatly reduced the amount of SR1 (also known as SRp34) full-length mRNA

and showed pleiotropic changes in morphology and development of transgenic plants [55]. Similar studies with RSZ33 altered the splicing pattern of two other SR proteins [SRp30 and SR1 (also known as SRp34)] [48]. The splicing of SRp30 is enhanced and correct splicing of the tenth intron of SR1/SRp34 was promoted in RSZ33 overexpression lines. Furthermore, it was shown that it regulates its own expression. The results indicate that SRp30 and RSZ33 are splicing factors and regulate splicing. In addition to splicing changes, the *RSZ33* overexpressor showed many developmental abnormalities [48].

Phosphorylation of SR proteins

Two families of protein kinases, SRPK (SR-specific protein kinase) and Clk/Sty (also called LAMMER-type kinases) that phosphorylate SR proteins have been extensively characterized in animals [56]. In *Arabidopsis* there are sixteen protein kinases that fall into the LAMMER-type kinase family [57]. This family contains both LAMMER-type as well as SRPKs. One of the kinases (AFC2) autophosphorylates and phosphorylates four plant SR proteins (SRZ21, SRZ22, SRp33 and SR45) [53] and the interaction between AFC2 and SR33 is modulated by the phosphorylation status of these proteins. Plant and animal LAMMER kinases differ in that plant kinases lack the RS domain that is implicated in the interaction of these kinases with the SR proteins. This suggests that structural features other than an RS domain are also important in the interaction of AFC2 with other proteins. The interaction of SRp34/SR1 with CypRS64 is strongly reduced by dephosphorylation whereas the phosphorylation status of SCL30 had no effect on its interaction with CypRS64. A member of the LAMMER family of protein kinases from tobacco (PK12), a nuclear protein, binds and phosphorylates both plant and animal SR proteins *in vitro* [58]. Overexpression of this kinase in *Arabidopsis* altered the splicing pattern of specific endogenous genes (SRp30, SR1/SRp34 and U1-70K) and the expression of several other genes and caused defects in development [59]. Based on these few reports it appears that phosphorylation of SR proteins modulates their interaction with other proteins and cellular localization (see below). Further studies are needed to evaluate fully the functional significance of the phosphorylation of SR proteins in regulating pre-mRNA splicing.

Localization and dynamics of SR proteins

Studies with metazoans show that the nucleus is highly organized with many different types of extremely dynamic subnuclear compartments such as Cajal bodies and speckles [60,61]. Speckles are intranuclear structures that correspond to interchromatin granules and contain a high concentration of pre-mRNA splicing factors. Although the precise function of speckles is not known, they are thought to be the sites of splicing factor storage and/or assembly [62]. During the past year, the localization of most plant SR proteins has been analyzed by expressing SR proteins fused to fluorescent reporters (GFP, YFP, CFP and RFP) in protoplasts, cultured cells and/or transgenic plants (Table 1). All SR proteins that have been tested were localized exclusively to the nucleus. Within the nucleus, in addition to diffused nucleoplasmic localization, all the tested SR proteins are localized to speckles [51,63–65]. Differences in the distribution of some SR proteins within the nuclei of different cell types were also observed. The number of speckles varied depending on cell type, cell cycle stage and developmental stage [51,63]. All SR proteins were excluded from the nucleolus and Cajal bodies [65]. Time-lapse microscopy of speckles that contain plant SR proteins have revealed that they move rapidly within the nuclear space [51,63,64]. Fluorescent recovery after photobleaching (FRAP) studies have shown that SR proteins in the speckles are rapidly exchanged with the nucleoplasmic pool [51]. Following transcriptional inhibition, the number of speckles was reduced with an increase in the size of the speckles [51,63]. Heat or a protein kinase inhibitor also caused dramatic relocalization of SR45 speckles [63]. However, the pattern of speckles is distinct with each treatment, indicating that the pathways affected by these treatments might be different. These studies clearly show that plant nuclei, like animal nuclei, contain highly dynamic subnuclear compartments. However, the mechanisms that regulate localization and dynamics of plant SR proteins are not known. The localization of most plant SR proteins, with a

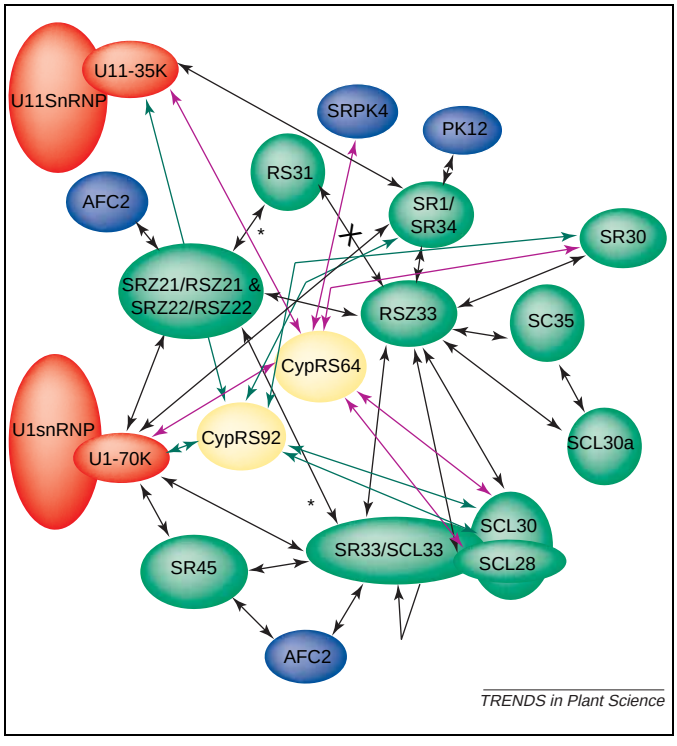

Figure 2. Network of interactions among serine/arginine-rich (SR) proteins and between SR proteins and other spliceosomal proteins. These interactions were identified using either yeast two-hybrid analysis and/or *in vitro* protein–protein interaction assays [52,53,59,67,68]. SRZ21/RSZ21 and SRZ22/RSZ22 are two different proteins that showed similar interactions with other proteins. The interaction of SR33/SCL33 with itself is indicated by an arrow turning back on itself. Asterisks indicate that SR33/SCL33 and RS31 interact only with SRZ21/RSZ21. RS31 and RSZ33 do not interact with each other (indicated by an 'X'). All SR proteins are shown in green. U1 and U11 snRNP proteins are indicated in red. Protein kinases and cyclophilin-like proteins are shown in blue and yellow, respectively. SR34, an SR protein with a molecular mass of 34 kDa (also called SR1); SR33, an SR protein with a molecular mass of 33 kDa (also called SCL33 for SC35-Like protein 33); SRZ21 and SRZ22 are 21 and 22 kDa SR proteins with one zinc knuckle (also called RSZ21 and RSZ22, respectively); AFC2, a LAMMER-type protein kinase; PK12, a LAMMER-type kinase from tobacco; CypRS, cyclophilin-like protein with RS domain.

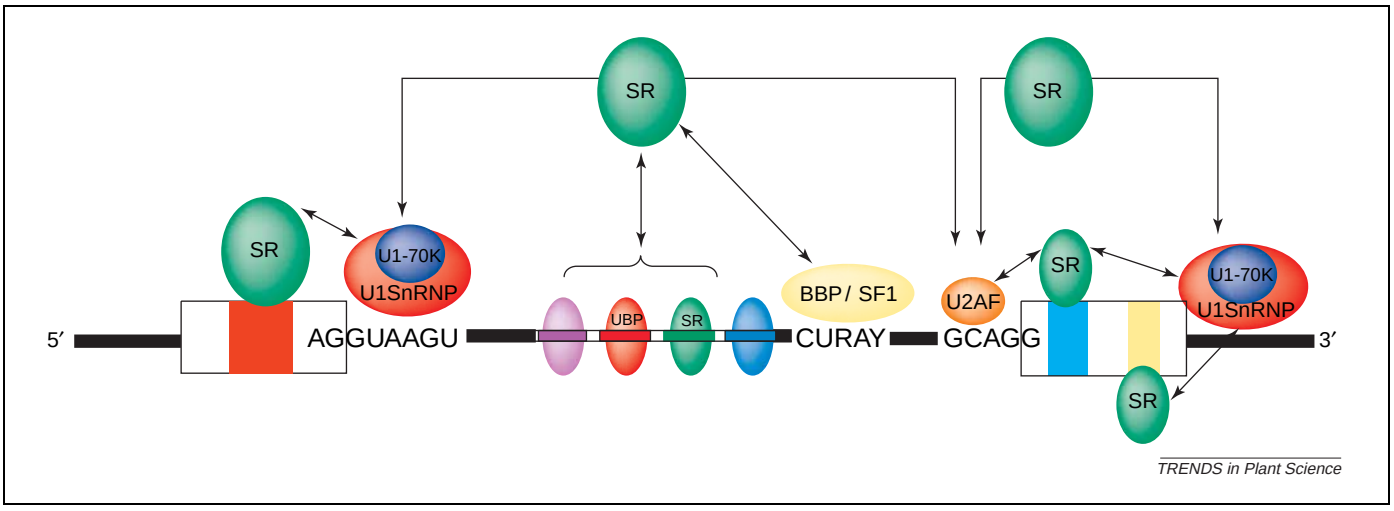

Figure 3. Potential roles of plant serine/arginine-rich (SR) and other RNA-binding proteins in intron and exon definition during the early stages of spliceosome assembly. White boxes indicate exons and the horizontal line between and on either side of each exon indicates introns. Consensus sequences of plant 5′ and 3′ splice sites and branch point are indicated. Colored boxes within each exon represent exonic splicing regulators [exonic splicing enhancers (ESE) shown in light red and blue and exonic splicing suppressor (ESS) shown in yellow]. Colored boxes within the intron represent intronic splicing regulators [intronic splicing enhancers (ISE) are shown in red and green; intronic splicing suppressors (ISS) are shown in purple and light blue] such as AU or U-rich region. SR and other RNA binding proteins that are likely to interact with exonic and intronic elements are shown. SR proteins can promote the recognition of 5′ and 3′ splice sites by recruiting U1 snRNP, U2AF and other spliceosomal proteins. In addition, they can bridge the spliceosomal components at 5′ and 3′ splice sites. Arrows indicate SR protein-mediated interactions. Proteins that bind to uridine-rich sequences and SR proteins are likely to play an important role in intron recognition. Several SR proteins, including some novel ones, and the proteins that bind uridine-rich sequences have already been identified in plants. Plant U1-70K is known to interact with at least five different SR proteins (see Figure 2). Some of the depicted interactions are based on animal literature. Abbreviations: BBP, branch point binding protein; SF1, splicing factor 1; SR, serine/arginine-rich protein; U2AF, U2 auxiliary factor, UBP, uridylate-specific binding protein.

few exceptions, was analyzed using protoplast transfection assays. Although the plant protoplast system is useful for gaining quick insight into the localization and dynamics of SR proteins, it is also necessary to study SR proteins *in planta* to appreciate fully the organ-, tissue- and cell-type-dependent changes in a developmental context. Further studies with fluorescent reporter fusion proteins using FRAP, FLIP (fluorescence loss in photobleaching) and FRET (fluorescence-resonance energy transfer) would provide an unprecedented look at the protein dynamics and interactions between spliceosomal proteins in live cells and be valuable for obtaining further insights into the functions of SR proteins [66].

Interaction network among SR proteins and other spliceosomal proteins

Several plant SR proteins have been shown to interact with other members in the SR family and other spliceosomal proteins. Interaction studies among plant SR proteins have shown a complex network of interactions [52,53,67,68] (Figure 2). In animals, only three SR proteins (SC35, ASF/SF2 and SR38) have been shown to interact with U1-70K [22,69,70]. The plant U1-70K interacts with more (five) SR proteins, including some novel SR proteins, suggesting that early stages of spliceosome formation or splice site selection might differ from that in animals. RSZ33 interacts with eight other SR proteins (Figure 2). It has been shown recently that four SR proteins, two U snRNP proteins (U1-70K and U11-35K) and an SR-specific protein kinase (SRPK4) interact with CypRS64, a protein that consists of an N-terminal peptidyl-prolyl *cis/trans* isomerase domain and a C-terminal domain with Ser-Arg/Ser-Pro (RS/SP) dipeptides [68]. CypRS92, a protein related to CypRS64 also interacts with SCL28, SCL30, SR1/SRp34, U1-70K and U11-35K [68]. *Arabidopsis* contains almost twice as many SR proteins than are found in humans (19 in *Arabidopsis* and 11 in humans), including several plant-specific ones; several of them interact with other SR proteins as well as with U1-70K and/or U11-35K, which suggests that these proteins are likely to play important roles during splice site selection and early stages of spliceosome assembly in plants [52,53,67]. The observation that one SR protein interacts with the U11 snRNP-specific protein indicates that SR proteins have a role in splicing U-12 introns also. From limited analyses of splicing of U2 and U12 introns it appears that the splicing efficiency of plant pre-mRNAs depends on both intronic elements (U or UA content or other ISEs/ISSs) and exonic signals (ESEs/ESSs) [9,31,36]. The complex interactions among plant SR proteins and between SR proteins and other spliceosomal proteins (Figure 2) suggest that they are likely to function in one or more of the following: defining exons and introns, recruiting spliceosomal proteins onto pre-mRNA, influencing splice site choice by binding to ESE/ISE or ESS/ISS, and/or bridging the spliceosomal components at 5′ and 3′ splice sites (Figure 3).

Concluding remarks and future directions

Understanding constitutive and regulated splicing is crucial, not only to understand its role in growth and development but also to optimize expression of transgenes in plants. We now know the full complement of SR proteins in *Arabidopsis* and rice. The structural features of SR proteins, their localization, and interaction between them and with other spliceosomal proteins strongly indicate that they are involved in early stages of spliceosome assembly and splice site selection. However, many

fundamental questions pertinent to recognition of introns and exons, and mechanisms involved in splice site choice, remain unanswered.

A major challenge now is to uncover the precise functions of each plant SR protein. Because of the large number of SR proteins, functional redundancy will be a problem in analyzing their function(s) using a single gene knockout approach. This is evident from studies with animals [71]. Hence, it would be necessary to generate knockouts in multiple SR genes in a single line to uncover their functions. In addition, overexpression using constitutive, regulated, or tissue-specific expression of individual SR proteins should also be useful in functional studies. Reduced levels of SR proteins might also show phenotype. In the absence of a plant-derived *in vitro* splicing system, the knockout or knockdown and overexpression approaches are particularly valuable in studying the role(s) of SR proteins.

Information about the interaction network of SR proteins is necessary to understand fully the roles of these proteins. It is possible to identify all proteins that interact directly or indirectly with each SR protein using yeast two hybrid screens and/or tandem affinity purification (TAP) and modified TAP methods [72]. Immunoaffinity selection using antibodies to a single protein in the spliceosome could also be useful in isolating protein complexes (as has been reported in animals [73]) and in identifying the protein components. These approaches should help develop a comprehensive protein–protein interaction network map for SR proteins and provide new insights into the function of SR proteins. In addition, systematic evolution of ligands by exponential enrichment (SELEX) and genomic SELEX approaches with each SR protein as well as bioinformatics analyses of all alternatively spliced genes across phylogenetically diverse plant species should aid in identifying the regulatory elements in plant pre-mRNAs [74–77].

Microarrays that distinguish spliced from unspliced mRNA from every intron-containing gene, such as those used with yeast and humans [7,78], should be useful in determining the genome-wide effects on splicing caused by knockout or knockdown, or overexpression of each SR protein. In cases where SR proteins are likely to be nonessential, splicing-sensitive microarrays offer a way to analyze subtle changes in pre-mRNA splicing on a global scale. Using splicing sensitive and/or genome-wide microarrays [7,79], it should be possible to develop a comprehensive catalog of all alternatively spliced pre-mRNAs, particularly in *Arabidopsis* and rice, in a few years. However, mechanistic understanding of constitutive and alternative splicing and elucidation of functions of each alternatively spliced isoform will take decades of research.

Acknowledgements

I thank Maxim Golovkin and Gul Shad Ali for their contributions to splicing research in my laboratory; Elisha Botnick for her help in preparing the figures; Gul Shad Ali, Irene Day and Saiprasad Goud for constructive comments on the manuscript. Work on Pre-mRNA splicing research in my laboratory is supported by a grant from the US Department of Energy (DE-FG02–04ER15556).

References

1 Chow, L.T. *et al*. (1977) An amazing sequence arrangement at the 5′ ends of adenovirus 2 messenger RNA. *Cell* 12, 1–8
2 Berget, S.M. *et al*. (1977) Spliced segments at the 5′ terminus of adenovirus 2 late mRNA. *Proc. Natl. Acad. Sci. U. S. A.* 74, 3171–3175
3 Le Hir, H. *et al*. (2000) The spliceosome deposits multiple proteins 20–24 nucleotides upstream of mRNA exon–exon junctions. *EMBO J.* 19, 6860–6869
4 Reichert, V.L. *et al*. (2002) 5′ exon interactions within the human spliceosome establish a framework for exon junction complex structure and assembly. *Genes Dev.* 16, 2778–2791
5 Maniatis, T. and Reed, R. (2002) An extensive network of coupling among gene expression machines. *Nature* 416, 499–506
6 Brett, D. *et al*. (2002) Alternative splicing and genome complexity. *Nat. Genet.* 30, 29–30
7 Johnson, J.M. *et al*. (2003) Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science* 302, 2141–2144
8 Haas, B.J. *et al*. (2003) Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* 31, 5654–5666
9 Reddy, A.S.N. (2001) Nuclear pre-mRNA splicing in plants. *CRC Crit. Rev. Plant Sci.* 20, 523–572
10 Jordan, T. *et al*. (2002) Alternative splicing of transcripts encoding Toll-like plant resistance proteins – what's the functional relevance to innate immunity? *Trends Plant Sci.* 7, 392–398
11 Dinesh-Kumar, S.P. and Baker, B.J. (2000) Alternatively spliced N resistance gene transcripts: their possible role in tobacco mosaic virus resistance. *Proc. Natl. Acad. Sci. U. S. A.* 97, 1908–1913
12 Zhang, X.C. and Gassmann, W. (2003) RPS4-mediated disease resistance requires the combined presence of RPS4 transcripts with full-length and truncated open reading frames. *Plant Cell* 15, 2333–2342
13 Macknight, R. *et al*. (2002) Functional significance of the alternative transcript processing of the *Arabidopsis* floral promoter FCA. *Plant Cell* 14, 877–888
14 Quesada, V. *et al*. (2003) Autoregulation of FCA pre-mRNA processing controls *Arabidopsis* flowering time. *EMBO J.* 22, 3142–3152
15 Burge, C.B. *et al*. (1999) Splicing of precursors to mRNAs by the spliceosomes. In *The RNA World* (Gesteland, R.F. *et al*., eds), pp. 525–560, Cold Spring Harbor Laboratory Press
16 Shukla, G.C. and Padgett, R.A. (1999) Conservation of functional features of U6atac and U12 snRNAs between vertebrates and higher plants. *RNA* 5, 525–538
17 Zhu, W. and Brendel, V. (2003) Identification, characterization and molecular phylogeny of U12-dependent introns in the *Arabidopsis thaliana* genome. *Nucleic Acids Res.* 31, 4561–4572
18 Rappsilber, J. *et al*. (2002) Large-scale proteomic analysis of the human spliceosome. *Genome Res.* 12, 1231–1245
19 Zhou, Z. *et al*. (2002) Comprehensive proteomic analysis of the human spliceosome. *Nature* 419, 182–185
20 Rosbash, M. and Séraphin, B. (1991) Who's on first? The U1 snRNP-5′ splice site interaction and splicing. *Trends Biochem. Sci.* 16, 187–190
21 Will, C.L. and Luhrmann, R. (1997) Protein functions in pre-mRNA splicing. *Curr. Opin. Cell Biol.* 9, 320–328
22 Manley, J.L. and Tacke, R. (1996) SR proteins and splicing control. *Genes Dev.* 10, 1569–1579
23 Golovkin, M. and Reddy, A.S.N. (1996) Structure and expression of a plant U1 snRNP 70K gene: alternative splicing of U1 snRNP 70K pre-mRNAs produces two different transcripts. *Plant Cell* 8, 1421–1435
24 Simpson, G.G. *et al*. (1995) Molecular characterization of spliceosomal proteins U1A and U2B″ from higher plants. *EMBO J.* 14, 4540–4550
25 Golovkin, M. and Reddy, A.S.N. (2003) Expression of U1snRNP 70K antisense transcript using *APETALA3* promoter suppresses the development of petals and stamens. *Plant Physiol.* 132, 1884–1891
26 Lorkovic, Z.J. *et al*. (2000) Pre-mRNA splicing in higher plants. *Trends Plant Sci.* 5, 160–167
27 Brown, W.W.S. and Simpson, C.G. (1998) Splice site selection in plant pre-mRNA splicing. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 49, 77–95
28 Schuler, M.A. (1998) Plant pre-mRNA splicing. In *A Look Beyond*

Transcription: Mechanisms Determining mRNA Stability and Translation in Plants (Bailey-Serres, J. and Gallie, D.R., eds), pp. 1–19, American Society of Plant Physiologists
29 Zhu, W. *et al*. (2003) Refined annotation of the *Arabidopsis* genome by complete expressed sequence tag mapping. *Plant Physiol.* 132, 469–484
30 Luehrsen, K.R. *et al*. (1994) Nuclear pre-mRNA processing in higher plants. *Prog. Nucleic Acids Res. Mol. Biol.* 47, 149–193
31 Lewandowska, D. *et al*. (2004) Determinants of plant U12-dependent intron splicing efficiency. *Plant Cell* 16, 1340–1352
32 Lambermon, M.H. *et al*. (2000) UBP1, a novel hnRNP-like protein that functions at multiple steps of higher plant nuclear pre-mRNA maturation. *EMBO J.* 19, 1638–1649
33 Carle-Urioste, J.C. *et al*. (1997) A combinatorial role for exon, intron and splice site sequences in splicing in maize. *Plant J.* 11, 1253–1263
34 Latijnhouwers, M.J. *et al*. (1999) Test of the combinatorial model of intron recognition in a native maize gene. *Plant Mol. Biol.* 41, 637–644
35 McCullough, A.J. and Schuler, M.A. (1997) Intronic and exonic sequences modulate 5′ splice site selection in plant nuclei. *Nucleic Acids Res.* 25, 1071–1077
36 Yoshimura, K. *et al*. (2002) Identification of a *cis* element for tissue-specific alternative splicing of chloroplast ascorbate peroxidase pre-mRNA in higher plants. *J. Biol. Chem.* 277, 40623–40632
37 Graveley, B.R. (2000) Sorting out the complexity of SR protein functions. *RNA* 6, 1197–1211
38 Sanford, J.R. *et al.* (2003) Multiple roles of the SR protein family in splicing regulation. In *Regulation of Alternative Splicing* (Vol. 31) (Jeanteur, P., ed.), pp. 33–58, Springer
39 Zahler, A.M. *et al*. (1992) SR proteins: a conserved family of pre-mRNA splicing factors. *Genes Dev.* 6, 837–847
40 Huang, Y. and Steitz, J.A. (2001) Splicing factors SRp20 and 9G8 promote the nucleocytoplasmic export of mRNA. *Mol. Cell* 7, 899–905
41 Lemaire, R. *et al*. (2002) Stability of a PKCI-1-related mRNA is controlled by the splicing factor ASF/SF2: a novel function for SR proteins. *Genes Dev.* 16, 594–607
42 Sanford, J.R. *et al*. (2004) A novel role for shuttling SR proteins in mRNA translation. *Genes Dev.* 18, 755–768
43 Tacke, R. and Manley, J.L. (1999) Determinants of SR protein specificity. *Curr. Opin. Cell Biol.* 11, 358–362
44 Lorkovic, Z.J. and Barta, A. (2002) Genome analysis: RNA recognition motif (RRM) and K homology (KH) domain RNA-binding proteins from the flowering plant *Arabidopsis thaliana*. *Nucleic Acids Res.* 30, 623–635
45 Kalyna, M. and Barta, A. (2004) A plethora of plant serine/arginine-rich proteins: redundancy or evolution of novel gene functions? *Biochem. Soc. Trans.* 32, 561–564
46 Lopato, S. *et al*. (1996) Pre-mRNA splicing in plants: characterization of Ser/Arg splicing factors. *Proc. Natl. Acad. Sci. U. S. A.* 93, 3074–3079
47 Lopato, S. *et al*. (1999) A novel family of plant splicing factors with a Zn knuckle motif: examination of RNA binding and splicing activities. *Plant Mol. Biol.* 39, 761–773
48 Kalyna, M. *et al*. (2003) Ectopic expression of atRSZ33 reveals its function in splicing and causes pleiotropic changes in development. *Mol. Biol. Cell* 14, 3565–3577
49 Lazar, G. *et al*. (1995) Identification of a plant serine–arginine-rich protein similar to the mammalian splicing factor SF2/ASF. *Proc. Natl. Acad. Sci. U. S. A.* 92, 7672–7676
50 Lazar, G. and Goodman, H.M. (2000) The Arabidopsis splicing factor SR1 is regulated by alternative splicing. *Plant Mol. Biol.* 42, 571–581
51 Fang, Y. *et al*. (2004) Tissue-specific expression and dynamic organization of SR splicing factors in *Arabidopsis*. *Mol. Biol. Cell* 15, 2664–2673
52 Golovkin, M. and Reddy, A.S.N. (1998) The plant U1 small nuclear ribonucleoprotein particle 70K protein interacts with two novel serine/arginine-rich proteins. *Plant Cell* 10, 1637–1648
53 Golovkin, M. and Reddy, A.S.N. (1999) An SC35-like protein and a novel serine/arginine-rich protein interact with *Arabidopsis* U1-70K protein. *J. Biol. Chem.* 274, 36428–36438
54 Lopato, S. *et al*. (1996) Characterization of a novel arginine/serine-rich splicing factor in *Arabidopsis*. *Plant Cell* 8, 2255–2264
55 Lopato, S. *et al*. (1999) atSRp30, one of two SF2/ASF-like proteins from *Arabidopsis thaliana*, regulates splicing of specific plant genes. *Genes Dev.* 13, 987–1001
56 Misteli, T. (1999) RNA splicing: what has phosphorylation got to do with it? *Curr. Biol.* 9, R198–R200
57 Shiu, S.H. and Bleecker, A.B. (2003) Expansion of the receptor-like kinase/Pelle gene family and receptor-like proteins in *Arabidopsis*. *Plant Physiol.* 132, 530–543
58 Savaldi-Goldstein, S. *et al*. (2000) The ethylene-inducible PK12 kinase mediates the phosphorylation of SR splicing factors. *Plant J.* 21, 91–96
59 Savaldi-Goldstein, S. *et al*. (2003) Alternative splicing modulation by a LAMMER kinase impinges on developmental and transcriptome expression. *Plant Cell* 15, 926–938
60 Lamond, A.I. and Spector, D.L. (2003) Nuclear speckles: a model for nuclear organelles. *Nat. Rev. Mol. Cell Biol.* 4, 605–612
61 Boudonck, K. *et al*. (1998) Coiled body numbers in the *Arabidopsis* root epidermis are regulated by cell type, developmental stage and cell cycle parameters. *J. Cell Sci.* 111, 3687–3694
62 Misteli, T. *et al*. (1997) The dynamics of a pre-mRNA splicing factor in living cells. *Nature* 387, 523–527
63 Ali, G.S. *et al*. (2003) Nuclear localization and *in vivo* dynamics of a plant-specific serine/arginine-rich protein. *Plant J.* 36, 883–893
64 Docquier, S. *et al*. (2004) Nuclear bodies and compartmentalization of pre-mRNA splicing factors in higher plants. *Chromosoma* 112, 255–266
65 Lorkovic, Z.J. *et al*. (2004) Use of fluorescent protein tags to study nuclear organization of the spliceosomal machinery in transiently transformed living plant cells. *Mol. Biol. Cell* 15, 3233–3243
66 Phair, R.D. and Misteli, T. (2000) High mobility of proteins in the mammalian cell nucleus. *Nature* 404, 604–609
67 Lopato, S. *et al*. (2002) Network of interactions of a novel plant-specific Arg/Ser-rich protein, atRSZ33, with atSC35-like splicing factors. *J. Biol. Chem.* 277, 39989–39998
68 Lorkovic, Z.J. *et al*. (2004) Interactions of *Arabidopsis* RS domain containing cyclophilins with SR proteins and U1 and U11 snRNP-specific proteins suggest their involvement in pre-mRNA splicing. *J. Biol. Chem.* 279, 33890–33898
69 Wu, J.Y. and Maniatis, T. (1993) Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell* 75, 1061–1070
70 Shin, C. *et al*. (2004) Dephosphorylated SRp38 acts as a splicing repressor in response to heat shock. *Nature* 427, 553–558
71 Longman, D. *et al*. (2000) Functional characterization of SR and SR-related genes in *Caenorhabditis elegans*. *EMBO J.* 19, 1625–1637
72 Gavin, A.C. and Superti-Furga, G. (2003) Protein complexes and proteome organization from yeast to man. *Curr. Opin. Chem. Biol.* 7, 21–27
73 Makarov, E.M. *et al*. (2002) Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome. *Science* 298, 2205–2208
74 Turek, C. and Gold, L. (1990) Systemic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* 249, 505–510
75 Schneider, D. *et al*. (1992) Selection of high affinity RNA ligands to the bacteriophage R17 coat protein. *J. Mol. Biol.* 228, 862–869
76 Singer, B.S. *et al*. (1997) Libraries for genomic SELEX. *Nucleic Acids Res.* 25, 781–786
77 Kim, S. *et al*. (2003) Specific SR protein-dependent splicing substrates identified through genomic SELEX. *Nucleic Acids Res.* 31, 1955–1961
78 Clark, T.A. *et al*. (2002) Genomewide analysis of mRNA processing in yeast using splicing-specific microarrays. *Science* 296, 907–910
79 Yamada, K. *et al*. (2003) Empirical analysis of transcriptional activity in the *Arabidopsis* genome. *Science* 302, 842–846
80 Reddy, A.S.N. *et al*. *Arabidopsis* U1 snRNP 70K protein and its interacting proteins: nuclear localization and *in vivo* dynamics of a novel plant-specific serine/arginine-rich protein. In *The Nuclear Envelope* (Evans, D. and Bryant, J., eds), BIOS Scientific Publisher (in press)