Spliceosomal Proteins in Plants

Y. Ru, B.-B. Wang, and V. Brendel(✉)

Contents

Abstract The spliceosome is a large nuclear structure consisting of dynamically interacting RNAs and proteins. This chapter briefly reviews some of the known components and their interactions. Large-scale proteomics and gene expression studies may be required to unravel the many intricate mechanisms involved in splice site recognition and selection.

Introduction

The spliceosome is a large ribonucleoprotein (RNP) complex whose highly orchestrated assembly at each intron involves small nuclear RNAs (snRNAs) and hundreds of proteins. Two types of spliceosomes have been identified in both

V. Brendel
Department of Genetics, Development and Cell Biology, Iowa State University,
Ames, IA 50011-3260, USA
e-mail: vbrendel@iastate.edu

A.S.N. Reddy, M. Golovkin (eds.), *Nuclear pre-mRNA Processing in Plants:*
Current Topics in Microbiology and Immunology 326.

vertebrates and plants and are referred to as major (or U2 type) and minor (or U12 type) spliceosomes. They differ in their snRNA composition, but differences in their protein composition are not completely understood, although some specific proteins were identified from the minor spliceosome (Will et al. 1999, 2004; Lorković et al. 2005). The minor spliceosome is involved in the splicing of a small fraction of introns with distinct donor site and branchpoint motifs (see the chapter by C. G. Simpson and J. W. S. Brown, this volume). A computational screen of the Arabidopsis genome identified 74 snRNA genes and about 400 genes encoding known or putative splicing-related proteins (Wang and Brendel 2004). Here we briefly review the snRNA genes characteristic for the two types of spliceosomes and then discuss the major classes of plant splicing-related proteins.

Small Nuclear RNAs

Like its counterparts in yeasts and metazoa, the major spliceosome in plants also contains five types of U-rich snRNAs, referred to as U1, U2, U4, U5, and U6 snRNA (reviewed in Lorković et al. 2000b; Reddy 2001). All five types were identified experimentally in Arabidopsis (Vankan et al. 1988; Vankan and Filipowicz 1988, 1989; Waibel and Filipowicz 1990; Hofmann et al. 1992). A genome-wide survey in Arabidopsis identified a total of 70 genes encoding these snRNAs, including 14 U1, 18 U2, 11 U4, 14 U5, and 13 U6 snRNA genes (Wang and Brendel 2004). Most of these genes seem to be active, as their promoter regions contain both TATA box and conserved upstream element (USE) motifs (Wang and Brendel 2004). Four types of snRNA genes (U1, U2, U4, and U5) are transcribed by RNA polymerase II (Pol II) (Vankan et al. 1988; Vankan and Filipowicz 1989; Connelly and Filipowicz 1993), while U6 snRNA genes are transcribed by polymerase III (Pol III) (Waibel and Filipowicz 1990). snRNA genes were also experimentally identified from other plant species, including bean (van Santen and Spritz 1987), pea (Hanley and Schuler 1991), potato (Vaux et al. 1992), wheat (Musci et al. 1992), and maize (Leader et al. 1993). For the minor spliceosome, four type of snRNAs named U11, U12, U4atac, and U6atac replace the U1, U2, U4, and U6 snRNA in the major spliceosome, while U5 snRNA is used in both major and minor spliceosomes (Tarn and Steitz 1997). Arabidopsis U12, U6atac, and U4atac were identified both computationally and experimentally (Shukla and Padgett 1999; Wang and Brendel 2004; Lorković et al. 2005), while U11 snRNA has only been computationally predicted (Wang and Brendel 2004). The four minor spliceosome snRNA genes also have the conserved TATA box and USE motifs in their promoter regions (Wang and Brendel 2004). Plant snRNAs likely play similar functions as their mammalian homologs, because their secondary structure and all functional domains are well conserved between plants and mammals (Lorković et al. 2000b). Recently, there is accumulating evidence that the snRNAs are involved in the catalysis of splicing, indicating that the spliceosome may be an RNA-centric enzyme similar to the ribosome (reviewed in Valadkhan 2005).

Small Nuclear Ribonucleoproteins

The major spliceosome is assembled dynamically at each intron, facilitated by interactions of the spliceosomal small nuclear ribonucleoproteins (snRNPs) with the pre-mRNA. Proteomic studies isolated five different snRNPs (U1, U2, U5, U4/U6, and U4/U6.U5 tri-snRNP) from human and yeast (Fabrizio et al. 1994; Caspary et al. 1999; Gottschalk et al. 1999; Krämer et al. 1999; Stevens and Abelson 1999; Stevens et al. 2001). Spliceosome assembly starts with U1 snRNP binding to the 5′-splice site, followed by U2 snRNP binding to the intron branch site, association of the U4/U6.U5 tri-snRNP, and release of U1 and U4 snRNPs to form the catalytic complex (reviewed in Burge et al. 1999). Many common (Sm core proteins) and specific proteins binding to the snRNAs were identified in each snRNP (reviewed in Will and Lührmann 2001). No proteomics approach has been applied to plant spliceosomes so far. Computational genome screening using human homologs as query sequences identified 91 snRNP protein-coding genes conserved in Arabidopsis, fewer than 10 of which have been experimentally studied (Wang and Brendel 2004).

Sm Core Proteins

Seven common proteins were identified in U1, U2, U4, and U5 snRNPs. All of them contain an Sm domain and thus were named as SmB, SmD1, SmD2, SmD3, SmE, SmF, and SmG (reviewed in Will and Lührmann 2001). Seven "like Sm" proteins (Lsm2–8) are the counterparts of Sm proteins in U6 snRNP. The Sm proteins bind to the snRNAs to form the structural core of the snRNPs. The Sm domain mediates protein interactions with other core proteins and snRNP-specific proteins. One additional Lsm protein exists (Lsm1) but cannot bind to snRNA (reviewed in Will and Lührmann 2001). All 15 Sm core proteins are conserved in Arabidopsis, with nine of them encoded by duplicated genes. All 24 genes are expressed based on EST data (Wang and Brendel 2004), but only Lsm5 (SAD1) has been experimentally characterized. *Lsm5 (SAD1)* is expressed universally at a low level in Arabidopsis. Mutation of Lsm5 increases plant sensitivity to drought stress and ABA (Xiong et al. 2001). Lsm5 protein may not be essential to the spliceosome but instead may affect the specificity or efficiency of the splicing machinery.

U1 snRNP-Specific Proteins

Three U1 snRNP-specific proteins (U1-70K, U1-A, and U1-C) are shared between human and yeast (Gottschalk et al. 1998). All three proteins exist in plants, and two of them (U1-70K and U1-A) are well characterized (Simpson et al. 1995; Golovkin and Reddy 1996). Arabidopsis U1-70K (atU1-70K) is an essential protein of 427

amino acid residues containing an RNA binding domain and an RS domain (Golovkin and Reddy 1996). The N-terminus of atU1-70K is conserved in eukaryotic organisms, while the C-terminal RS domain has much more variation. Four SR proteins (SR33, SR45, SRZ21, SRZ22) interact with atU1-70K, possibly through their RS domains (Golovkin and Reddy 1998, 1999). Reduced expression levels of atU1-70K in floral organs cause abnormal petal and stamen development (Golovkin and Reddy 2003). atU1-70K can be alternatively spliced by retaining the long sixth intron (910 nt), which leads to a truncated protein (Golovkin and Reddy 1996). Recently the gene structure and alternative splicing pattern of U1-70K gene were found to be conserved in rice and maize (Gupta et al. 2006; Wang and Brendel 2006a). U1-A gene was identified from potato and Arabidopsis. Similar to their human counterparts, plant U1-A proteins can bind specifically to the U1 snRNA loop II sequence (Simpson et al. 1995). U1-70K, U1-A, and the yet uncharacterized U1-C gene are all single copy in Arabidopsis (Golovkin and Reddy 1996; Wang and Brendel 2004), while potato may have multiple copies of U1-A (Ibrahim et al. 2001). Several yeast-specific proteins such as SNU71, SNU65/Prp42, SNU56, NAM8, Prp39, and Prp40 make the yeast U1 snRNP a more complex structure than its human counterpart (Gottschalk et al. 1998). Computational searches identified two homologs for each of Prp39 and Prp40 in Arabidopsis (Wang and Brendel 2004). The remaining proteins are apparently absent from the plant genome, indicating unique features in 5′-splice site recognition in plants.

Other snRNP-Specific Proteins in the Major Spliceosome

Similarity searches identified 23 U2, 14 U5, 10 U4/U6, and six U4/U6.U5 snRNP-specific protein-coding genes in the Arabidopsis genome (Wang and Brendel 2004). U2-A′ and U2-B″ from U2 snRNP were experimentally studied in plants (Simpson et al. 1995; Lorković et al. 2004). Potato U2-B″ can interact with human U2-A′ to enhance sequence-specific binding (Simpson et al. 1995). Similar to the cases in human and yeast, plant U2-B″ and U1-A (from U1 snRNP) are very similar to each other, with over 50% identity and 70% similarity between their protein sequences. Only one member of the U1A/U2B″ family (SNF) was found in *Drosophila melanogaster*, which can interact with both U1 and U2 snRNAs (Polycarpou-Schwarz et al. 1996). Functional redundancy was recently discovered between U1-A and U2-B″ in *Caenorhabditis elegans* (Saldi et al. 2007). In Arabidopsis, two copies of U2-B″ exist, with 80% identity and 90% similarity between their protein sequences. Both U2-B″ genes are expressed and can be alternatively spliced, increasing the complexity of this protein family two- to fourfold (Wang and Brendel 2004). Other proteins from these snRNPs are also highly conserved. A few genes, including atSF3a120/SAP114 (Prp21) from U2 snRNP, U5-116 KD (Snu114), U5-200 KD (Brr2), U4/U6-90 KD (Prp3), U4/U6-15.5 KD (Snu13), and tri-65 KD (Snu66), are duplicated multiple times in Arabidopsis (Wang and Brendel 2004). It is not known how the functions of these redundant proteins coordinate with each other in plants.

U11/U12 snRNP-Specific Proteins

Little is known about the protein composition of the minor spliceosome. The only purified snRNPs are 18S U11/U12 snRNP and 12S U11 snRNP in humans. Seven minor spliceosome specific proteins, all Sm core proteins, and seven subunits of heteromeric SF3b from U2 snRNP were identified from U11/U12 snRNP, but none of the U1 snRNP-specific proteins were detected (Will et al. 2004). Four (25KDa, 31KDa, 35KDa, and 65KDa) minor spliceosome proteins are present in the Arabidopsis genome, while the other three (20KDa, 48KDa, and 59KDa) are absent (Wang and Brendel 2004). Sequence similarity detected between At2g46200 and U11-59KD and between At3g04160 and U11-48K suggests they are possible orthologs (Lorković et al. 2005). The U11/U12-35 KD has strong sequence similarity and analogous function to U1-70K (Lorković et al. 2005). Contrasting with the duplication events in the major spliceosome, all minor spliceosome-specific genes exist as single copy. The Arabidopsis U11/U12-specific genes are highly conserved in rice, indicating a similar mechanism of splicing of AT-AC introns throughout plants (Lorković et al. 2005).

Splicing Factors

In a previous survey (Wang and Brendel 2004), splicing factors were divided into eight subgroups: splice site selection proteins; serine/arginine-rich (SR) proteins; 17S U2-associated proteins; 35S U5-associated proteins; BΔU1 complex-specific proteins; exon junction complex (EJC) proteins; second-step splicing factors; and other known splicing factors. Here we focus on the first two subgroups because they are more thoroughly studied than other subgroups.

Proteins Involved in Splice Site Selection

Cap-Binding Proteins

The nuclear cap-binding complex (CBC) is required for effective spliceosome assembly and pre-mRNA splicing in both yeast and mammalian systems (Colot et al. 1996; Lewis et al. 1996a). More detailed experiments showed that CBC facilitates the recruitment of U1 snRNP to the 5′-splice site of the first intron in HeLa cell nuclear extracts (Lewis et al. 1996b). Two subunits of CBC in Arabidopsis [cap-binding proteins (CBP) AtCBP20 and AtCBP80] were identified and characterized (Kmieciak et al. 2002). Both *AtCBP20* and *AtCBP80* are single-copy genes in the Arabidopsis genome. AtCBP20 plays a role in ABA regulation and drought response (Papp et al. 2004). Another subunit, AtCBP80, is also called abscisic acid

(ABA) hypersensitive 1 (ABH1). Studies of several *ABH1* mutants showed that ABH1 is a modulator in early ABA signal transduction (Hugouvieux et al. 2001) and acts in the flowering pathway (Bezerra et al. 2004; Kuhn et al. 2007) (see the chapter by J. M. Kuhn et al., this volume). Specifically, *ABH1* may affect flowering time by influencing the splicing patterns of some flowering pathway genes. The intron-1-retaining transcripts of *FLOWERING LOCUS C* (*FLC*) accumulate in *abh1-7* (Kuhn et al. 2007), suggesting that *ABH1* may regulate *FLC* intron 1 removal. However, it is still unknown whether plant CBC, like its animal counterparts, facilitates the recognition of the 5′-splice site of the first intron by U1 snRNP.

U2 snRNP Auxiliary Factor

U2 snRNP Auxiliary Factor (U2AF) is a dimeric splicing factor with a small 35-kDa and a large 65-kDa subunit. In animals, the small subunit $U2AF^{35}$ binds to the 3′-splice site (Merendino et al. 1999; Wu et al. 1999; Zorio and Blumenthal 1999). The large subunit $U2AF^{65}$ binds to the polypyrimidine tract between the branchpoint and the 3′-splice site (Zamore et al. 1992). The bindings of U2AF help to bring U2 snRNP to the branchpoint (Ruskin et al. 1988). It is not well understood, however, how U2AF functions in plants.

Two isoforms of $U2AF^{65}$, named $U2AF^{65}$a and $U2AF^{65}$b, have been identified in both *Nicotiana plumbaginifolia* (Tex-Mex tobacco) and Arabidopsis (Domon et al. 1998). $U2AF^{65}$a localizes to the nucleus. Both proteins interact with plant introns and prefer binding to poly(U) instead of poly(G), poly(C), or poly(A) (Domon et al. 1998). The U-rich sequence in plants was shown to function as either polypyrimidine tract or splicing signal (Simpson et al. 2004). In addition, both plant isoforms can complement the lack of human $U2AF^{65}$ in HeLa cell splicing extracts and stimulate splicing (Domon et al. 1998).

$U2AF^{35}$ homologs are found in Arabidopsis, rice, maize, and other flowering plants (Domon et al. 1998; Wang and Brendel 2004, 2006b). Two homologs were identified in Arabidopsis, at$U2AF^{35}$a and at$U2AF^{35}$b, and both have nuclear localization. Altered levels of at$U2AF^{35}$ cause the change of *FLC* expression level and *FCA* (*FLOWERING TIME CONTROL PROTEIN*) splicing pattern and thus lead to the change of flowering time (Wang and Brendel 2006b). Plants with altered levels of at$U2AF^{35}$ also show other pleiotropic phenotypes. A widely conserved C-terminal motif (SERE) in seed plant $U2AF^{35}$ homologs suggests that it may have plant-specific function (Wang and Brendel 2006b).

Serine/Arginine-Rich Proteins

The serine/arginine-rich (SR) proteins are essential in several steps of both constitutive and alternative splicing (see the chapters by A. Barta et al. and G. S. Ali and A. S. N. Reddy, this volume). All SR proteins are phosphoproteins. Typical characteristics

of SR proteins include one or two N-terminal RNA recognition motifs (RRMs) and a C-terminal arginine/serine-rich (RS) domain (Kalyna and Barta 2004; Reddy 2004). During splice site recognition and spliceosome formation, the RRM recognizes and binds to particular exon and/or intron motifs, and the RS domain interacts with other proteins in the spliceosome. There are 11 and seven genes encoding SR proteins in human and *Caenorhabditis elegans*, respectively (Kalyna and Barta 2004), while there are many more plant SR genes, with at least 19 in Arabidopsis and 24 in rice (Iida and Go 2006; Isshiki et al. 2006; Reddy 2007). Some SR proteins are highly conserved, whereas others are species specific. The fact that both plant- and animal-specific SR proteins exist indicates that the mechanism of pre-mRNA splicing is different by some means in the two kingdoms.

Discovery of Plant SR Proteins

Plant SR proteins were discovered by various approaches. For example, the gene of the first plant SR protein, atSRp34/atSR1, was identified by searching Arabidopsis cDNAs for highly conserved RRMs in animal SR genes (Lazar et al. 1995). Similar approaches based on sequence similarity search were applied later to discover additional SR proteins along with other plant splicing factors (Lopato et al. 1996b, 1999a, 2002, 2006; Lorković and Barta 2002; Gao et al. 2004; Wang and Brendel 2004; Iida and Go 2006). Another approach was to isolate clones interacting with U1-70K or known SR proteins in yeast two-hybrid assays (Golovkin and Reddy 1998, 1999; Lopato et al. 2002, 2006). In the third approach, SR proteins were detected with human SF2/ASF-specific antibody and monoclonal antibody mAb104, which interacts with a shared phosphoepitope in all SR proteins (Lopato et al. 1996a, 1999b).

SR Proteins in Arabidopsis

Based on sequence similarities with human SR proteins, the 19 Arabidopsis SR proteins can be grouped into four families: SF2/ASF, SC35, 9G8, and plant-specific families (Wang and Brendel 2004). Four SR proteins, atSRp34/SR1, atSRp34a, atSRp34b, and atSRp30, have gene structures similar to human SF2/ASF. They have two N-terminal RRMs and a C-terminal RS domain. Unlike human SF2/ASF, all four of these SR proteins except atSRp30 have a PSK domain that is rich in proline, serine, and lysine (Reddy 2004). The splicing activity of atSRp34/atSR1 has been shown by its influence on alternative 5′-splice site selection in HeLa nuclear extract, which is comparable with the activity of human SF2/ASF (Lazar et al. 1995). Regulation of *atSRp34*/*SR1* splicing is organ- and stage-specific and temperature-dependent but is not autoregulated (Lazar and Goodman 2000). Full-length *atSRp34*/*SR1* mRNA is downregulated when another SR gene, *atSRp30*, is overexpressed (Lopato et al. 1999b). Overexpression of *atSRp30* also changes alternative splicing of endogenous *atRSp31*, *atSRp34*/*RS1*, *U1-70K*, and its own pre-mRNA.

Transgenic plants overexpressing *atSRp30* show pleiotropic phenotype changes including delayed flowering and larger flowers and rosette leaves (Lopato et al. 1999b).

The SC35 family comprises five SR proteins (atSC35, atSR33/atSCL33, atSCL30, atSCL30a, and atSCL28). All of them have typical RRM and RS domains. In yeast two-hybrid assay and coprecipitation analysis, atSR33/atSCL33 was shown to interact with itself and atSR45, a plant-specific SR protein, but not with atSRZ21 or atSRZ22 from the 9G8 family (Golovkin and Reddy 1999).

Proteins from the 9G8 family are similar to human 9G8 splicing factor. Three of them (atRSZp22/atSRZ22, atRSZp22a, and atRSZp21/atSRZ21) have a glycine hinge and a CCHC-type zinc finger motif in between the RRM and RS domain. The other two members, atRSZ33 and atRSZ32, each have two CCHC-type zinc fingers between the RRM and RS domain (Reddy 2004). atRSZp22 can restore splicing activities in both SR protein-depleted HeLa cell S100 extract and 9G8-depleted HeLa nuclear extract (Lopato et al. 1999a). In yeast two-hybrid screens and in vitro binding assays, atRSZ33 interacts with atRSZp21 and atRSZp22 from the same family, atSRp34/SR1 from the SF2/ASF family, and all five members of the SC35 family (Lopato et al. 2002).

Five Arabidopsis SR proteins do not have animal counterparts and belong to the plant-specific family. atRSp41, atRSp40/atSRp35, atRSp32/atRSp31a, and atRSp31 have two RRMs and a C-terminal RS domain. Unlike all other plant SR proteins with only one RS domain, atSR45 has two RS domains separated by an RRM (Golovkin and Reddy 1999). Both atRSp31 and atSR45 can complement SR protein-deficient HeLa cell S100 extract (Lopato et al. 1996b; Ali et al. 2007). The mutant *sr45-1* shows a pleiotropic phenotype including delayed flowering and change of flower and leaf morphologies. On the molecular level, *FLC* is upregulated in *sr45-1*, consistent with the late flowering phenotype. Compared with wild type, *sr45-1* has a different splicing pattern of pre-mRNAs of other SR genes, including *atSRp30*, *atRSp31*, *atRSp31a*, *atSRp34*, and *atSRp34b* (Ali et al. 2007).

SR Proteins in Other Plant Species

Many SR proteins from maize (*Zea mays*), wheat (*Triticum aestivum*), and rice (*Oryza sativa*) have also been identified and characterized (Gao et al. 2004; Iida and Go 2006; Isshiki et al. 2006; Lopato et al. 2006). There are at least three SF2/ASF-like SR proteins in maize (zmSRp30, zmSRp31, and zmSRp32) (Gao et al. 2004). They show strong structural similarity to Arabidopsis atSRp34/SR1 and atSRp30. Like *atSRp34*/*SR1*, all three maize SR genes are alternatively spliced. Overexpression of *zmSRp32* in transiently transformed maize BMS cells enhances the selection of a weak 5′-splice site. In wheat, two SR proteins, taRSZ38 and taRSZ38a, have high similarity to Arabidopsis atRSZ33, a 9G8 family protein (Lopato et al. 2006). Additional wheat proteins involved in pre-mRNA splicing were identified with the yeast two-hybrid system by using taRSZ38 and several subsequent positive clones as baits. Among the resulting proteins are taU1-70K, taU2AF^{65}, taU2AF^{35}, and a number of SR proteins (taSRp30,

taSRp30a, taRSZ22, and taRSZ22a) (Lopato et al. 2006). To date, 24 rice SR proteins have been found; four of them are homologs of the Arabidopsis plant-specific family (Iida and Go 2006; Isshiki et al. 2006). Several rice SR genes undergo alternative splicing events, some of which are conserved between Arabidopsis and rice (see below) (Iida and Go 2006; Kalyna et al. 2006). Transient assays in rice Oc cell protoplasts indicate that osRSp29 and osRSZp23 enhance splicing and favor different 5′-splice sites of the same intronic region (Isshiki et al. 2006). Domain-swapping experiments show that the first RRM is essential for osRSp29's efficient splicing activity (Isshiki et al. 2006). Overexpression of *osRSZ36* in transgenic rice changes its own RNA's splicing pattern, suggesting a feedback regulation. Overexpression of *osSRp33b* alters the splicing of *osSRp33a* and *osSRp32* pre-mRNAs (Isshiki et al. 2006).

Conserved Alternative Splicing of Plant SR Genes

Alternative splicing is observed for 15 of the 19 Arabidopsis SR genes, from which about 95 transcripts are produced (Palusa et al. 2007). This greatly increases the flexibility and complexity of the spliceosome. Seventeen of the 24 rice SR genes are subjected to alternative splicing as well (Iida and Go 2006). Alternative splicing events also occur in some SR genes from other plant species (see above). The observation that some events are conserved among SR homologs from different plant species (Iida and Go 2006; Kalyna et al. 2006) highlights the importance of alternative splicing in plant SR gene transcripts.

Studies reveal that long introns (>400 nt) are present in most (12 of 19) Arabidopsis SR genes, and these long introns are alternatively spliced in nine of such genes (Kalyna and Barta 2004). This is remarkable considering that the average length of Arabidopsis intron is 173 nt (Reddy 2007) and about 22% of the genes in Arabidopsis are alternatively spliced (Wang and Brendel 2006a). More detailed analyses show that the alternatively spliced long introns are also found in rice, maize, gymnosperms, moss (*Physcomitrella patens*), and green alga (*Chlamydomonas reinhardtii*) transcripts encoding SR proteins (Iida and Go 2006; Kalyna et al. 2006). These alternatively spliced long introns are located in the RRM-encoded regions, resulting in mRNAs encoding SR proteins with partial RRMs. Such conserved alternative splicing events are found in the plant-specific and SC35 families and the 9G8 members with two zinc knuckles (Iida and Go 2006; Kalyna et al. 2006) and may be critical to splicing regulation.

Regulation of Splicing in Plants

Splicing is a dynamic process affecting the expression of all intron-containing genes. Recent studies revealed that 20%–30% expressed genes are alternatively spliced in Arabidopsis and rice, demonstrating the critical roles of splicing in

gene expression (Campbell et al. 2006; Wang and Brendel 2006a). Splicing must therefore be tightly regulated to ensure accuracy and efficiency. Many proteins are involved in splicing regulation. Some proteins can assist and promote splicing, while others may have a negative effect. We defined proteins assisting splicing as splicing factors, such as SR proteins and U2AF, which were discussed in the previous section. Other proteins that can either modify splicing factors or have negative effect on splicing were defined as splicing regulators (Wang and Brendel 2004). Nevertheless, this classification is tentative, and sometimes it is hard to define clearly whether a protein is a splicing factor or a splicing regulator.

The main difference in splicing-related genes between mammals and plants is the expansion of splicing regulators in plants (Wang and Brendel 2004). One type of splicing regulator is hnRNP protein, which can bind to pre-mRNA and block the binding site for splicing factors (Wang and Brendel 2004). PTB/hnRNP I, for instance, can compete with the U2AF large subunit for the polypyrimidine tract of introns (Lin and Patton 1995). Three homologs of PTB/hnRNP I were identified in Arabidopsis (Wang and Brendel 2004). Moreover, a family of 21 glycine-rich RNA binding proteins were identified in Arabidopsis as homologs of human hnRNP A1 and hnRNP A2/B1(Wang and Brendel 2004). Five of these genes were also experimentally identified, including AtGRP7, AtGRP8, and three UBA2 homologs (Heintzen et al. 1994; Lambermon et al. 2002). AtGRP7 can influence alternative splicing of its own transcripts as well as AtGRP8 transcripts (Staiger et al. 2003). UBA2 proteins can interact with UBP1 and UBA1 proteins to recognize U-rich sequences (Lambermon et al. 2002). Homologs for other hnRNP proteins such as hnRNP F and CUG-BP were also identified in Arabidopsis (Lorković and Barta 2002; Wang and Brendel 2004). In addition, 15 hnRNP-like proteins were identified in Arabidopsis, possibly representing the plant-specific hnRNPs (Lambermon et al. 2000; Lorković et al. 2000a; Landsberger et al. 2002; Wang and Brendel 2004). The UBP1 proteins can strongly enhance splicing of some introns in protoplasts (Lambermon et al. 2000), whereas UBA1, RBP45, and RBP47 proteins have no such function reported (Lorković et al. 2000a; Landsberger et al. 2002).

SR protein kinase is another type of splicing regulator (see the chapter by R. Fluhr, this volume). By phosphorylating SR proteins, SR protein kinase can modulate splicing in both constitutive and alternative splicing (reviewed in Stojdl and Bell 1999). Sequences of SR protein kinases are highly conserved in plants, suggesting the critical function of these proteins. Eight homologs corresponding to three mammalian SR protein kinases, including Lammer/CLK kinases, SRPK1, and SPRK2, were identified in Arabidopsis (Wang and Brendel 2004). Lammer/CLK kinases were proven to phosphorylate SR proteins *in vitro* in plants (Golovkin and Reddy 1999; Savaldi-Goldstein et al. 2000). Overexpression of one tobacco Lammer/CLK kinase homolog, *PK12*, causes alternative splicing pattern changes in several genes, including U1-70K and two SR proteins, atSRp30 and atSR1/atSRp34 (Savaldi-Goldstein et al. 2003). Connections between splicing and other SR protein kinase are yet to be established.

In addition to proteins, many other factors may affect splicing, either directly or indirectly. For instance, *cis*-elements in pre-mRNAs are critical for the spliceosome to define and remove introns correctly. Hormones, such as ABA and IAA, can change the alternative splicing pattern in three SR genes in Arabidopsis (Palusa et al. 2007). Ethylene can induce the expression of SR protein kinase *PK12* in tobacco (Savaldi-Goldstein et al. 2000). Environmental factors, such as heat, cold, and salt, have dramatic effects on the alternative splicing pattern of several SR genes (Palusa et al. 2007). Light can also regulate the expression of some splicing-related genes (Heintzen et al. 1994). Detailed discussion of these regulators is beyond the scope of this chapter. It is clear that plants developed specific regulatory mechanisms to control the largely conserved splicing machinery when adapting to their environments.

Conclusion and Perspective

The complex interactions of pre-mRNA, snRNAs, and spliceosomal proteins in pre-mRNA processing are critical to the accurate expression of a eukaryotic genome. It has become clear that while the core molecular entities involved in these processes are conserved throughout the animal and plant kingdoms, there are also distinct proteins that participate in splicing in plants. Much less clear is the extent to which these proteins are conserved among different phylogenetic groups of plants, and very little is known about how distinct proteins influence processing of particular pre-mRNAs and determine alternative splicing choices, for example. A recent study identified 84 putative exonic splicing enhancer motifs, of which 35 were experimentally shown to promote splicing (Pertea et al. 2007). In the years ahead it will be fascinating to correlate such motifs with specific (SR) proteins that recognize them and to seek understanding of how these proteins interact with the core constituents of the spliceosome. Proteomics studies of isolated plant spliceosomes may be key to progress toward these goals.

Acknowledgements Y. R. and V. B. were supported in part by National Science Foundation Grant DBI-0606909. B.-B. W. was supported by National Science Foundation Grants DBI-0321460 and DBI-0606966.

References

Ali GS, Palusa SG, Golovkin M, Prasad J, Manley JL, Reddy ASN (2007) Regulation of plant developmental processes by a novel splicing factor. PLoS ONE 2:e471

Bezerra IC, Michaels SD, Schomburg FM, Amasino RM (2004) Lesions in the mRNA cap-binding gene ABA HYPERSENSITIVE 1 suppress FRIGIDA-mediated delayed flowering in Arabidopsis. Plant J 40:112–119

Burge CB, Tuschl T, Sharp PA (1999) Splicing of precursors to mRNAs by the spliceosome. In: Gesteland RF, Cech TR, Atkins JF (eds) The RNA World: the nature of modern RNA suggests a prebiotic RNA. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, pp 525–560

Campbell MA, Haas BJ, Hamilton JP, Mount SM, Buell CR (2006) Comprehensive analysis of alternative splicing in rice and comparative analyses with Arabidopsis. BMC Genomics 7:327

Caspary F, Shevchenko A, Wilm M, Seraphin B (1999) Partial purification of the yeast U2 snRNP reveals a novel yeast pre-mRNA splicing factor required for pre-spliceosome assembly. EMBO J 18:3463–3474

Colot HV, Stutz F, Rosbash M (1996) The yeast splicing factor Mud13p is a commitment complex component and corresponds to CBP20, the small subunit of the nuclear cap-binding complex. Genes Dev 10:1699–1708

Connelly S, Filipowicz W (1993) Activity of chimeric U small nuclear RNA (snRNA)/mRNA genes in transfected protoplasts of *Nicotiana plumbaginifolia*: U snRNA 3′-end formation and transcription initiation can occur independently in plants. Mol Cell Biol 13:6403–6415

Domon C, Lorković ZJ, Valcarcel J, Filipowicz W (1998) Multiple forms of the U2 small nuclear ribonucleoprotein auxiliary factor U2AF subunits expressed in higher plants. J Biol Chem 273:34603–34610

Fabrizio P, Esser S, Kastner B, Lührmann R (1994) Isolation of *S. cerevisiae* snRNPs: comparison of U1 and U4/U6.U5 to their human counterparts. Science 264:261–265

Gao H, Gordon-Kamm WJ, Lyznik LA (2004) ASF/SF2-like maize pre-mRNA splicing factors affect splice site utilization and their transcripts are alternatively spliced. Gene 339:25–37

Golovkin M, Reddy ASN (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

Golovkin M, Reddy ASN (1998) The plant U1 small nuclear ribonucleoprotein particle 70K protein interacts with two novel serine/arginine-rich proteins. Plant Cell 10:1637–1648

Golovkin M, Reddy ASN (1999) An SC35-like protein and a novel serine/arginine-rich protein interact with Arabidopsis U1-70K protein. J Biol Chem 274:36428–36438

Golovkin M, Reddy ASN (2003) Expression of U1 small nuclear ribonucleoprotein 70K antisense transcript using APETALA3 promoter suppresses the development of sepals and petals. Plant Physiol 132:1884–1891

Gottschalk A, Neubauer G, Banroques J, Mann M, Lührmann R, Fabrizio P (1999) Identification by mass spectrometry and functional analysis of novel proteins of the yeast [U4/U6.U5] tri-snRNP. EMBO J 18:4535–4548

Gottschalk A, Tang J, Puig O, Salgado J, Neubauer G, Colot HV, Mann M, Seraphin B, Rosbash M, Lührmann R, Fabrizio P (1998) A comprehensive biochemical and genetic analysis of the yeast U1 snRNP reveals five novel proteins. RNA 4:374–393

Gupta S, Ciungu A, Jameson N, Lal SK (2006) Alternative splicing expression of U1 snRNP 70K gene is evolutionary conserved between different plant species. DNA Seq 17:254–261

Hanley BA, Schuler MA (1991) cDNA cloning of U1, U2, U4 and U5 snRNA families expressed in pea nuclei. Nucleic Acids Res 19:1861–1869

Heintzen C, Melzer S, Fischer R, Kappeler S, Apel K, Staiger D (1994) A light- and temperature-entrained circadian clock controls expression of transcripts encoding nuclear proteins with homology to RNA-binding proteins in meristematic tissue. Plant J 5:799–813

Hofmann CJ, Marshallsay C, Waibel F, Filipowicz W (1992) Characterization of the genes encoding U4 small nuclear RNAs in *Arabidopsis thaliana*. Mol Biol Rep 17:21–28

Hugouvieux V, Kwak JM, Schroeder JI (2001) An mRNA cap binding protein, ABH1, modulates early abscisic acid signal transduction in Arabidopsis. Cell 106:477–487

Ibrahim AF, Watters JA, Brown JW (2001) Differential expression of potato U1A spliceosomal protein genes: a rapid method for expression profiling of multigene families. Plant Mol Biol 45:449–460

Iida K, Go M (2006) Survey of conserved alternative splicing events of mRNAs encoding SR proteins in land plants. Mol Biol Evol 23:1085–1094

Isshiki M, Tsumoto A, Shimamoto K (2006) The serine/arginine-rich protein family in rice plays important roles in constitutive and alternative splicing of pre-mRNA. Plant Cell 18:146–158

Kalyna M, Barta A (2004) A plethora of plant serine/arginine-rich proteins: redundancy or evolution of novel gene functions? Biochem Soc Trans 32:561–564

Kalyna M, Lopato S, Voronin V, Barta A (2006) Evolutionary conservation and regulation of particular alternative splicing events in plant SR proteins. Nucleic Acids Res 34:4395–4405

Kmieciak M, Simpson CG, Lewandowska D, Brown JW, Jarmolowski A (2002) Cloning and characterization of two subunits of *Arabidopsis thaliana* nuclear cap-binding complex. Gene 283:171–183

Krämer A, Gruter P, Groning K, Kastner B (1999) Combined biochemical and electron microscopic analyses reveal the architecture of the mammalian U2 snRNP. J Cell Biol 145:1355–1368

Kuhn JM, Breton G, Schroeder JI (2007) mRNA metabolism of flowering-time regulators in wild-type Arabidopsis revealed by a nuclear cap binding protein mutant, abh1. Plant J 50:1049–1062

Lambermon MH, Fu Y, Wieczorek Kirk DA, Dupasquier M, Filipowicz W, Lorković ZJ (2002) UBA1 and UBA2, two proteins that interact with UBP1, a multifunctional effector of pre-mRNA maturation in plants. Mol Cell Biol 22:4346–4357

Lambermon MH, Simpson GG, Wieczorek Kirk DA, Hemmings-Mieszczak M, Klahre U, Filipowicz W (2000) UBP1, a novel hnRNP-like protein that functions at multiple steps of higher plant nuclear pre-mRNA maturation. EMBO J 19:1638–1649

Landsberger M, Lorković ZJ, Oelmuller R (2002) Molecular characterization of nucleus-localized RNA-binding proteins from higher plants. Plant Mol Biol 48:413–421

Lazar G, Goodman HM (2000) The Arabidopsis splicing factor SR1 is regulated by alternative splicing. Plant Mol Biol 42:571–581

Lazar G, Schaal T, Maniatis T, Goodman HM (1995) Identification of a plant serine-arginine-rich protein similar to the mammalian splicing factor SF2/ASF. Proc Natl Acad Sci USA 92:7672–7676

Leader D, Connelly S, Filipowicz W, Waugh R, Brown JW (1993) Differential expression of U5snRNA gene variants in maize (*Zea mays*) protoplasts. Plant Mol Biol 21:133–143

Lewis JD, Gorlich D, Mattaj IW (1996a) A yeast cap binding protein complex (yCBC) acts at an early step in pre-mRNA splicing. Nucleic Acids Res 24:3332–3336

Lewis JD, Izaurralde E, Jarmolowski A, McGuigan C, Mattaj IW (1996b) A nuclear cap-binding complex facilitates association of U1 snRNP with the cap-proximal 5′ splice site. Genes Dev 10:1683–1698

Lin CH, Patton JG (1995) Regulation of alternative 3′ splice site selection by constitutive splicing factors. RNA 1:234–245

Lopato S, Borisjuk L, Milligan AS, Shirley N, Bazanova N, Langridge P (2006) Systematic identification of factors involved in post-transcriptional processes in wheat grain. Plant Mol Biol 62:637–653

Lopato S, Forstner C, Kalyna M, Hilscher J, Langhammer U, Indrapichate K, Lorković ZJ, Barta A (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

Lopato S, Gattoni R, Fabini G, Stevenin J, Barta A (1999a) 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

Lopato S, Kalyna M, Dorner S, Kobayashi R, Krainer AR, Barta A (1999b) atSRp30, one of two SF2/ASF-like proteins from *Arabidopsis thaliana*, regulates splicing of specific plant genes. Genes Dev 13:987–1001

Lopato S, Mayeda A, Krainer AR, Barta A (1996a) Pre-mRNA splicing in plants: characterization of Ser/Arg splicing factors. Proc Natl Acad Sci USA 93:3074–3079

Lopato S, Waigmann E, Barta A (1996b) Characterization of a novel arginine/serine-rich splicing factor in Arabidopsis. Plant Cell 8:2255–2264

Lorković ZJ, 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

Lorković ZJ, Hilscher J, Barta A (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

Lorković ZJ, Lehner R, Forstner C, Barta A (2005) Evolutionary conservation of minor U12-type spliceosome between plants and humans. RNA 11:1095–1107

Lorković ZJ, Wieczorek Kirk DA, Klahre U, Hemmings-Mieszczak M, Filipowicz W (2000a) RBP45 and RBP47, two oligouridylate-specific hnRNP-like proteins interacting with poly(A)+ RNA in nuclei of plant cells. RNA 6:1610–1624

Lorković ZJ, Wieczorek Kirk DA, Lambermon MH, Filipowicz W (2000b) Pre-mRNA splicing in higher plants. Trends Plant Sci 5:160–167

Merendino L, Guth S, Bilbao D, Martinez C, Valcarcel J (1999) Inhibition of msl-2 splicing by Sex-lethal reveals interaction between U2AF35 and the 3′ splice site AG. Nature 402:838–841

Musci MA, Egeland DB, Schuler MA (1992) Molecular comparison of monocot and dicot U1 and U2 snRNAs. Plant J 2:589–599

Palusa SG, Ali GS, Reddy ASN (2007) Alternative splicing of pre-mRNAs of Arabidopsis serine/arginine-rich proteins: regulation by hormones and stresses. Plant J 49:1091–1107

Papp I, Mur LA, Dalmadi A, Dulai S, Koncz C (2004) A mutation in the Cap Binding Protein 20 gene confers drought tolerance to Arabidopsis. Plant Mol Biol 55:679–686

Pertea M, Mount SM, Salzberg SL (2007) A computational survey of candidate exonic splicing enhancer motifs in the model plant *Arabidopsis thaliana*. BMC Bioinformatics 8:159

Polycarpou-Schwarz M, Gunderson SI, Kandels-Lewis S, Seraphin B, Mattaj IW (1996) *Drosophila* SNF/D25 combines the functions of the two snRNP proteins U1A and U2B' that are encoded separately in human, potato, and yeast. RNA 2:11–23

Reddy ASN (2001) Nuclear pre-mRNA splicing in plants. Critical Rev Plant Sci 20:523–571

Reddy ASN (2004) Plant serine/arginine-rich proteins and their role in pre-mRNA splicing. Trends Plant Sci 9:541–547

Reddy ASN (2007) Alternative splicing of pre-messenger RNAs in plants in the genomic era. Annu Rev Plant Biol 58:267–294

Ruskin B, Zamore PD, Green MR (1988) A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly. Cell 52:207–219

Saldi T, Wilusz C, MacMorris M, Blumenthal T (2007) Functional redundancy of worm spliceosomal proteins U1A and U2B. Proc Natl Acad Sci USA 104:9753–9757

Savaldi-Goldstein S, Aviv D, Davydov O, Fluhr R (2003) Alternative splicing modulation by a LAMMER kinase impinges on developmental and transcriptome expression. Plant Cell 15:926–938

Savaldi-Goldstein S, Sessa G, Fluhr R (2000) The ethylene-inducible PK12 kinase mediates the phosphorylation of SR splicing factors. Plant J 21:91–96

Shukla GC, Padgett RA (1999) Conservation of functional features of U6atac and U12 snRNAs between vertebrates and higher plants. RNA 5:525–538

Simpson CG, Jennings SN, Clark GP, Thow G, Brown JW (2004) Dual functionality of a plant U-rich intronic sequence element. Plant J 37:82–91

Simpson GG, Clark GP, Rothnie HM, Boelens W, van Venrooij W, Brown JW (1995) Molecular characterization of the spliceosomal proteins U1A and U2B from higher plants. EMBO J 14:4540–4550

Staiger D, Zecca L, Wieczorek Kirk DA, Apel K, Eckstein L (2003) The circadian clock regulated RNA-binding protein AtGRP7 autoregulates its expression by influencing alternative splicing of its own pre-mRNA. Plant J 33:361–371

Stevens SW, Abelson J (1999) Purification of the yeast U4/U6.U5 small nuclear ribonucleoprotein particle and identification of its proteins. Proc Natl Acad Sci USA 96:7226–7231

Stevens SW, Barta I, Ge HY, Moore RE, Young MK, Lee TD, Abelson J (2001) Biochemical and genetic analyses of the U5, U6, and U4/U6 x U5 small nuclear ribonucleoproteins from *Saccharomyces cerevisiae*. RNA 7:1543–1553

Stojdl DF, Bell JC (1999) SR protein kinases: the splice of life. Biochem Cell Biol 77:293–298

Tarn WY, Steitz JA (1997) Pre-mRNA splicing: the discovery of a new spliceosome doubles the challenge. Trends Biochem Sci 22:132–137

Valadkhan S (2005) snRNAs as the catalysts of pre-mRNA splicing. Curr Opin Chem Biol 9:603–608

van Santen VL, Spritz RA (1987) Nucleotide sequence of a bean (*Phaseolus vulgaris*) U1 small nuclear RNA gene: implications for plant pre-mRNA splicing. Proc Natl Acad Sci USA 84:9094–9098

Vankan P, Edoh D, Filipowicz W (1988) Structure and expression of the U5 snRNA gene of *Arabidopsis thaliana*. Conserved upstream sequence elements in plant U-RNA genes. Nucleic Acids Res 16:10425–10440

Vankan P, Filipowicz W (1988) Structure of U2 snRNA genes of *Arabidopsis thaliana* and their expression in electroporated plant protoplasts. EMBO J 7:791–799

Vankan P, Filipowicz W (1989) A U-snRNA gene-specific upstream element and a −30 'TATA box' are required for transcription of the U2 snRNA gene of *Arabidopsis thaliana*. EMBO J 8:3875–3882

Vaux P, Guerineau F, Waugh R, Brown JW (1992) Characterization and expression of U1snRNA genes from potato. Plant Mol Biol 19:959–971

Waibel F, Filipowicz W (1990) U6 snRNA genes of Arabidopsis are transcribed by RNA polymerase III but contain the same two upstream promoter elements as RNA polymerase II-transcribed U-snRNA genes. Nucleic Acids Res 18:3451–3458

Wang BB, Brendel V (2004) The ASRG database: identification and survey of *Arabidopsis thaliana* genes involved in pre-mRNA splicing. Genome Biol 5:R102

Wang BB, Brendel V (2006a) Genomewide comparative analysis of alternative splicing in plants. Proc Natl Acad Sci USA 103:7175–7180

Wang BB, Brendel V (2006b) Molecular characterization and phylogeny of U2AF35 homologs in plants. Plant Physiol 140:624–636

Will CL, Lührmann R (2001) Spliceosomal UsnRNP biogenesis, structure and function. Curr Opin Cell Biol 13:290–301

Will CL, Schneider C, Hossbach M, Urlaub H, Rauhut R, Elbashir S, Tuschl T, Lührmann R (2004) The human 18S U11/U12 snRNP contains a set of novel proteins not found in the U2-dependent spliceosome. RNA 10:929–941

Will CL, Schneider C, Reed R, Lührmann R (1999) Identification of both shared and distinct proteins in the major and minor spliceosomes. Science 284:2003–2005

Wu S, Romfo CM, Nilsen TW, Green MR (1999) Functional recognition of the 3′ splice site AG by the splicing factor U2AF35. Nature 402:832–835

Xiong L, Gong Z, Rock CD, Subramanian S, Guo Y, Xu W, Galbraith D, Zhu JK (2001) Modulation of abscisic acid signal transduction and biosynthesis by an Sm-like protein in Arabidopsis. Dev Cell 1:771–781

Zamore PD, Patton JG, Green MR (1992) Cloning and domain structure of the mammalian splicing factor U2AF. Nature 355:609–614

Zorio DA, Blumenthal T (1999) Both subunits of U2AF recognize the 3′ splice site in *Caenorhabditis elegans*. Nature 402:835–838

http://www.springer.com/978-3-540-76775-6

Nuclear pre-mRNA Processing in Plants
Reddy, A.S.N.; Golovkin, M.V. (Eds.)
2008, XIII, 321 p., Hardcover
ISBN: 978-3-540-76775-6