

---

# Evolutionary conservation of minor U12-type spliceosome between plants and humans

---

ZDRAVKO J. LORKOVIĆ, REINHARD LEHNER,<sup>1</sup> CHRISTINA FORSTNER, and ANDREA BARTA

Max F. Perutz Laboratories, University Departments at the Vienna Biocenter, Department of Biochemistry, Medical University of Vienna, A-1030 Vienna, Austria

## ABSTRACT

Splicing of rare, U12-type or AT-AC introns is mediated by a distinct spliceosome that assembles from U11, U12, U4atac, U6atac, and U5 snRNPs. Although in human cells the protein composition of minor and major snRNPs is similar, differences, particularly in U11 and U12 snRNPs, have been recently described. We have identified an *Arabidopsis* U11 snRNP-specific 35K protein as an interacting partner of an RS-domain-containing cyclophilin. By using a transient expression system in *Arabidopsis* protoplasts, we show that the 35K protein incorporates into snRNP. Oligo affinity selection and glycerol gradient centrifugation revealed that the *Arabidopsis* 35K protein is present in monomeric U11 snRNP and in U11/U12-di snRNP. The interaction of the 35K protein with *Arabidopsis* SR proteins together with its strong sequence similarity to U1-70K suggests that its function in splicing of minor introns is analogous to that of U1-70K. Analysis of *Arabidopsis* and *Oryza sativa* genome sequences revealed that all U11/U12-di-snRNP-specific proteins are conserved in dicot and monocot plants. In addition, we have identified an *Arabidopsis* gene encoding the homolog of U4atac snRNA and a second *Arabidopsis* gene encoding U6atac snRNA. Secondary structure predictions indicate that the *Arabidopsis* U4atac is able to form dimeric complexes with both *Arabidopsis* U6atac snRNAs. As revealed by RNaseA/T1 protection assay, the U4atac snRNA gene is expressed as an ~160-nt RNA, whereas the second U6atac snRNA gene seems to be a pseudogene. Taken together, our data indicate that recognition and splicing of minor, AT-AC introns in plants is highly similar to that in humans.

**Keywords:** pre-mRNA splicing; U11/U12 snRNP; U12-dependent spliceosome; SR proteins; *Arabidopsis thaliana*; *Oryza sativa*

## INTRODUCTION

Most eukaryotic nucleus-encoded genes are interrupted by introns, which have to be removed from pre-mRNAs in order to generate functional, export-competent mRNA. Pre-mRNA splicing takes place in the spliceosome, an RNP complex that assembles from five small nuclear ribonucleoprotein particles (snRNP) and numerous additional proteins. Many dynamic RNA–RNA, RNA–protein, and protein–protein interactions are required for proper spliceosome assembly. These interactions involve universally conserved sequences at the 5' splice site (5'ss), 3' splice site (3'ss), branch point sequence (BPS), and the polypyrimidine tract (Moore et al. 1993; Krämer 1996; Burge et al. 1999).

A rare class of introns, called AT-AC or U12-dependent introns, was identified a decade ago (Jackson 1991; Hall and Padgett 1994). These introns contain distinct conserved sequences at the 5'ss, 3'ss, and BPS, and are spliced by a minor U12-type spliceosome. Like the major spliceosome, U12 spliceosome is formed from five snRNPs, of which only U5 is common to both types. During assembly of the minor spliceosome, the U11 snRNP forms base pairs with the 5'ss, whereas the U12 snRNP base-pairs with the BPS. Thus, they are functional analogs of the U1 and U2 snRNPs of the major spliceosome (Tarn and Steitz 1997; Yu and Steitz 1997; Patel and Steitz 2003). However, in contrast to U1 and U2 snRNPs, U11 and U12 snRNPs exist not only as monoparticles but also as an 18S U11/U12 di-snRNP complex in HeLa extracts, which simultaneously binds the 5'ss and BPS (Wassarman and Steitz 1992; Tarn and Steitz 1997; Frilander and Steitz 1999). Finally, by the association of the U4atac/U6atac•U5 tri-snRNP (where U4atac/U6atac snRNPs are functional analogs of U4/U6 snRNPs) (Tarn and Steitz 1996), the assembly of the minor spliceosome is completed (Yu and Steitz 1997; Frilander and Steitz 2001; Patel and Steitz 2003).

---

<sup>1</sup>Present address: Center for Brain Research, Division of Biochemistry and Molecular Biology, Medical University of Vienna, Spitalgasse 4, A-1090, Vienna, Austria

**Reprint requests to:** Zdravko J. Lorković, Max F. Perutz Laboratories, University Departments at the Vienna Biocenter, Department of Biochemistry, Medical University of Vienna, A-1030 Vienna, Austria; e-mail: zdravko.lorkovic@univie.ac.at; fax: 43 1 4277 9616.

Article and publication are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.2440305>.

Spliceosomal snRNPs are composed of snRNAs containing trimethylguanosine Cap ( $m_3G$ ) at their 5' ends (except U6 snRNA), which are associated with seven Sm (B, D1, D2, D3, E, F, and G) proteins. Sm proteins are present in all spliceosomal snRNPs, except in U6 snRNP, which contains Sm-like proteins (LSM). In addition to these core proteins, each snRNP contains a set of snRNP-specific proteins (Moore et al. 1993; Krämer 1996; Burge et al. 1999). Several recent studies revealed a similar protein composition of major and minor snRNPs, although differences have been observed as well (Will et al. 1999, 2001, 2004; Schneider et al. 2002). For example, purification of the human U11/U12 di-snRNP complex revealed that U1A, U1C, U1-70K, U2A', U2B'', and all subunits of the SF3a were absent from U11/U12 di-snRNP preparations, whereas all subunits of SF3b have been shown to be shared by both 17S U2 snRNP and 18S U11/U12 di-snRNPs (Will et al. 1999, 2001, 2004). More interestingly, several proteins have been found only in U11/U12 di-snRNP preparation, among them a 35K protein, which turned out to be specifically associated with the U11 snRNP (Will et al. 2004). Based on its domain organization and sequence homology with U1-70K, it has been proposed that this protein, named U11-35K, analogous to U1-70K, facilitates 5'ss recognition on AT-AC introns (Will et al. 1999). In addition, proteins denoted 65K, 59K, 48K, 31K, 25K, and 20K were also found to be specifically associated with U11/U12 di-snRNP in human cells. Of all these proteins, 59K, 48K, 35K, and 25K specifically associate with U11 snRNP, whereas 65K, 31K, and 20K are components either of 18S U11/U12 di-snRNP or of the 15S U12 snRNP (Will et al. 2004).

U12-type introns have been found in vertebrates, most invertebrates, and in higher plants (Wu et al. 1996; Burge et al. 1998; Zhu and Brendel 2003), and very often their positions in a particular gene are conserved among species (Burge et al. 1998; Wu and Krainer 1999). In *Arabidopsis*, the splice sites and BPS of most U12-type introns are highly similar to the respective animal sequences (Zhu and Brendel 2003). Like U2-type introns, plant U12 introns are also UA-rich (Zhu and Brendel 2003; Lewandowska et al. 2004) and this feature is important for their efficient splicing (Lewandowska et al. 2004). The existence of U12-type introns in plants implies the existence of snRNAs involved in their splicing. Indeed, *Arabidopsis* U12 and U6atac snRNAs have been identified (Shukla and Padgett 1999). The sequence conservation of these snRNAs between human and *Arabidopsis* minor snRNAs is much lower than that of U1 and U6 snRNAs. However, secondary structural elements, presumably required for function, are conserved in *Arabidopsis* U12 and U6atac snRNAs (Shukla and Padgett 1999). Likewise, the U11 snRNAs from *Arabidopsis* and rice are structurally similar to the human one, although primary sequence similarity did not allow its identification (Schneider et al. 2004). High conservation

of the U11 snRNP-specific proteins between humans and *Arabidopsis* (Will et al. 1999, 2004; this work) and the existence of U11, U12, and U6atac snRNAs (Shukla and Padgett 1999; Schneider et al. 2004) strongly support the argument for the existence of U4atac snRNAs in plants as well.

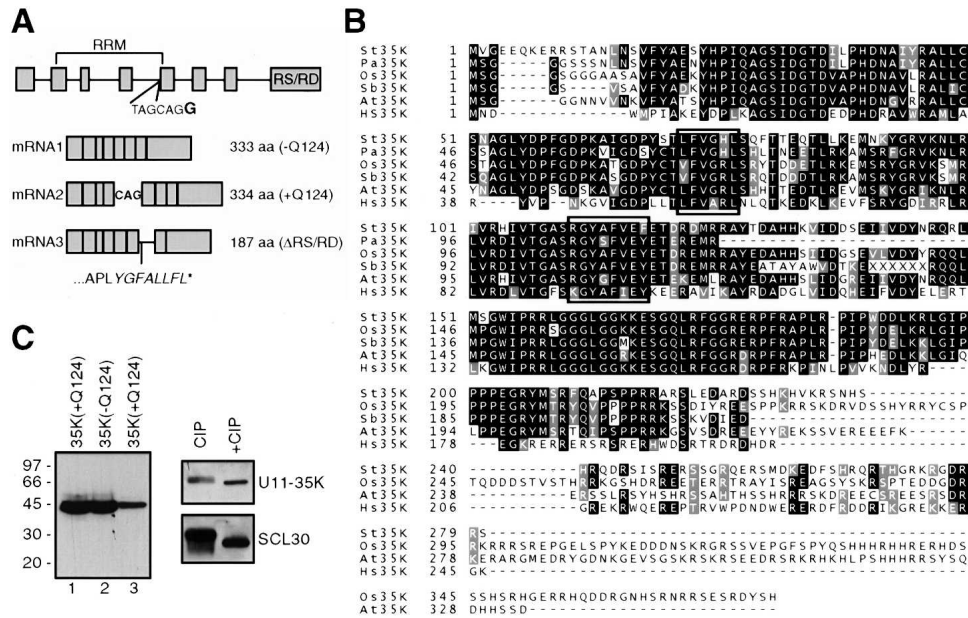
Recently, we identified an *Arabidopsis* U11-35K protein as an interacting partner of an RS-domain-containing cyclophilin (Lorković et al. 2004b). Here, we show by transient expression of U11-35K protein that *Arabidopsis* U11 snRNP also exists in two forms: the free monomeric form and in di-snRNP complex with U12 snRNP. In addition, U11-35K protein interacts with SR proteins in vitro, suggesting a function analogous to that of U1-70K. Survey of *Arabidopsis* and rice genomes, as well as EST databases, revealed that the majority of U11/U12 di-snRNP specific proteins are expressed in higher plants. Finally, we have identified an *Arabidopsis* U4atac snRNA gene, which expresses an snRNA of ~160 nt.

## RESULTS AND DISCUSSION

### Characterization of *Arabidopsis* gene encoding U11 snRNP-specific 35K protein

In a yeast two-hybrid screen with *Arabidopsis* RS-domain-containing cyclophilin, CypRS92 (Lorković et al. 2004b), a partial cDNA clone encoding the C-terminal half of a protein homologous to human U11 snRNP-specific 35K (Will et al. 1999), was isolated. To further characterize this protein and to get insight into the function of plant snRNPs involved in splicing of minor AT-AC introns, we cloned a full-length cDNA encoding the *Arabidopsis* U11-35K protein from a cDNA library. Sequencing of the amplified cDNAs revealed two forms encoding full-length proteins differing in only one amino acid. This difference results from usage of an alternative 3' splice site in intron 4, which contains two consecutive YAG (TAGCAGG) consensus 3' splice sites. The utilization of the proximal CAG results in a protein of 333 amino acids that is lacking glutamine 124 (Fig. 1A). Compared to the protein encoded by the longer cDNA, the lack of Q124 produces an ungapped alignment with human and other plant U11-35K proteins (Fig. 1B). In addition to these two cDNA forms encoding the U11-35K protein, an additional alternatively spliced cDNA containing unspliced intron 6 was cloned. This splicing event results in an mRNA encoding a truncated protein product of 187 amino acids lacking the RS/RD domain (Fig. 1A). The functional significance of these alternative splicing events is not clear yet (see below). As determined by RNase A/T1 protection assay, all three mRNAs encoding U11-35K are expressed at low levels in most *Arabidopsis* organs (data not shown).

Like its human ortholog, *Arabidopsis* and other plant U11-35K proteins contain an RRM, followed by a



**FIGURE 1.** Sequence and expression analysis of *Arabidopsis* U11–35K protein. (A) Schematic representation of exon–intron structure of the gene encoding U11–35K. Exons are represented by gray boxes, and introns by solid lines. Positions of the RRM and RS/RD domains are indicated. The two consecutive YAG consensus 3' splice sites in intron 4 are indicated (TAGCAG). Below are schematic representations of the three different mRNAs resulting from alternative splicing, with sizes of encoded proteins indicated in amino acids (aa). The amino acid sequence printed in italics is encoded by the included intron 6 of mRNA3. The asterisk denotes the stop codon. (B) Sequence alignment of *Arabidopsis thaliana* (At2g43370), *Oryza sativa* (Os; gi:32982022), *Solanum tuberosum* (St; gi:39823686), *Sorghum bicolor* (Sb; gi:4596518), *Persea americana* (Pa; gi:42641885), and *Homo sapiens* (Hs; gi:5902144) U11–35K proteins. The sequences of *S. tuberosum*, *S. bicolor*, and *P. americana* were deduced from EST clones; note that the C termini of these sequences are not complete. Xs in the *S. bicolor* sequence denote amino acids of uncertain identity. The RNP1 and RNP2 motifs of RRM are in boxes. Sequences were aligned by using the ClustalW program and shaded on the BOXSHADE server. The order of sequences is as it appeared in the ClustalW output. Amino acids identical or similar in >50% of the analyzed sequences are shown on black or gray backgrounds, respectively. The accession numbers of plant and human proteins refer to their DNA sequences at the NCBI. (C) The aberrant behavior of *Arabidopsis* U11–35K protein on SDS-PAGE is not due to extensive phosphorylation. Protein extracts from tobacco protoplasts (lanes 1,2) or from yeast cells (lane 3) expressing HA-tagged U11–35K were analyzed by Western blotting with rat monoclonal antibody (Roche) against the HA-tag. A single protein band was detected at ~45 kDa (left panel). Molecular mass standards in kilodaltons are indicated on the left. Protein extracts from tobacco protoplasts (as above) were incubated without (lane –CIP) or with calf intestinal phosphatase (lane +CIP) for 1 h at room temperature and subsequently analyzed by Western blotting with anti-HA antibody (two right panels). SCL30 (lower panel), an SR protein from *Arabidopsis* that is highly phosphorylated as evident from the shift after CIP treatment, was used as a control (lower panel). U11–35K does not show a significant shift after CIP treatment (upper panel).

glycine-hinge (Fig. 1B), which is reminiscent of both plant and metazoan U1–70K proteins (Query et al. 1989; Golovkin and Reddy 1996; Will et al. 1999). The C-terminal half of the protein contains several RS/RD/RE-dipeptide repeats (Fig. 1A,B). Sequence analysis revealed that the *Arabidopsis* U11–35K protein is more similar to its human ortholog than to *Arabidopsis* U1–70K. The RRM domain of *Arabidopsis* U11–35K protein is 54% identical and 71% similar to the corresponding region of human U11–35K (Will et al. 1999), and 41% identical and 59% similar to the RRM of *Arabidopsis* U1–70K protein. Orthologs of U11–35K have been identified in different dicot plant species as well as in the monocot *Oryza sativa* (Fig. 1B), indicating strong conservation and an important function for this protein in splicing of minor U12-type introns in plants, consistent with the fact that U12-type introns are present in land plants (Wu et al. 1996; Burge et al. 1999; Zhu and Brendel 2003).

The *Arabidopsis* U11–35K protein is 333 (or 334 including glutamine 124) amino acids long with a predicted molecular mass of 37 kDa. However, HA-tagged 35K

protein expressed in yeast cells or in plant protoplasts, as well as in vitro translated protein, migrates as a 45-kDa band (Fig. 1C, left panel; data not shown). This aberrant behavior of U11–35K cannot be attributed to extensive phosphorylation, as in contrast to the *Arabidopsis* SR protein SCL30, treatment of protein extracts with calf intestinal phosphatase resulted in a very small shift (~2 kDa) on SDS-PAGE (Fig. 1C, right two panels).

### ***Arabidopsis* U11–35K protein interacts with SR proteins in vitro and localizes to nuclear speckles**

In splicing of major U2-type introns, SR proteins interact with the U1 snRNP protein U1–70K at the 5' splice site (Wu and Maniatis 1993; Kohtz et al. 1994), and these interactions are important for stabilization of U1 snRNP at the 5' splice site. As the U11–35K protein seems to perform the function of U1–70K protein in splicing of minor introns (Will et al. 1999) and as splicing of minor introns can

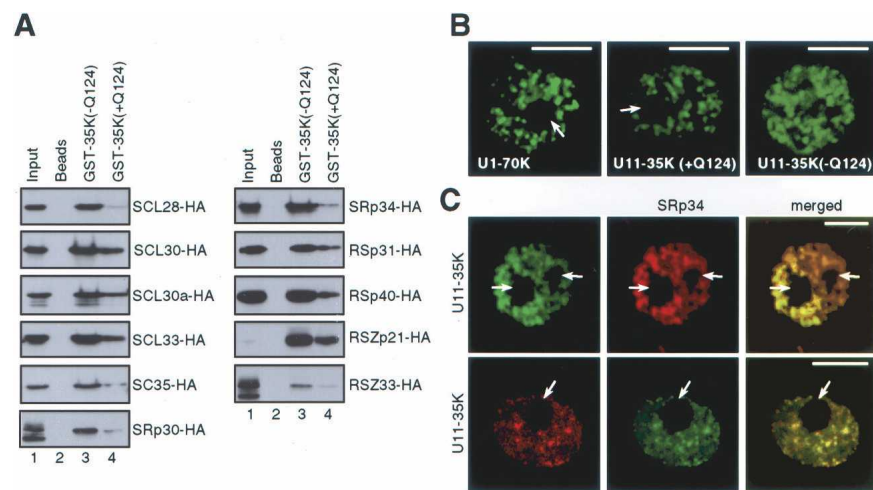
be activated by SR proteins in vitro (Hastings and Krainer 2001), we asked whether U11–35K protein is also able to interact with plant SR proteins. In order to prevent indirect RNA-mediated interactions (SR proteins and U11–35K are RNA-binding proteins), protein extracts were treated with RNase A. In pull-down assays with recombinant GST-tagged U11–35K protein interaction with most *Arabidopsis* SR proteins expressed in protoplasts was found (Fig. 2A, lanes 3). We also asked whether alternative splicing of U11–35K, which results in inclusion of Q124, affects its ability to interact with SR proteins. As shown in Figure 2A (lanes 4), all *Arabidopsis* SR proteins tested interacted with U11–35K(+Q124) protein, but with fairly reduced affinity compared to U11–35K(–Q124). For both U11–35K versions we used the same amount of the recombinant GST-fusion protein in pull-down assays, indicating that this difference is, indeed, reflecting intrinsic binding capacity of the two proteins. Interestingly, the presence of Q124 does not seem to influence assembly of the protein into snRNP in vivo (see below, Fig. 3A,B). However, RNA-binding studies revealed that U11–35K(+Q124), in contrast to U11–35K(–Q124), does not bind U11 snRNA in vitro (data not shown). Although these two results seem to be contradictory, it is possible that the assembly of U11–35K into snRNP in vivo is assisted by other U11 snRNP-specific proteins, or by an as-yet-undefined assembly factor. Alternatively, Q124, which is located in the RRM, could induce a conformational change thereby preventing binding to U11 snRNA and interaction with SR proteins in vitro. In addition, reduced affinity of U11–35K for SR proteins indicates that intactness of the C-terminal part of U11–35K RRM is required for protein–protein interaction.

Consistent with the above described in vitro interactions with SR proteins and with the role in pre-mRNA splicing, U11–35K protein localized to the nucleus (Fig. 2B). Coexpression of U11–35K protein with SRp34, an *Arabidopsis* ortholog of the human SF2/ASF (Lopato et al. 1999) and a good marker for speckles in plant nuclei (Lorkovic et al. 2004a) revealed colocalization in speckles in both tobacco and *Arabidopsis* cell nuclei (Fig. 2C). In addition to the speckled nuclear pattern, we repeatedly observed nucleolar localization of 35K-GFP fusion (data not shown). This most probably reflects the nuclear maturation pathway of the U11

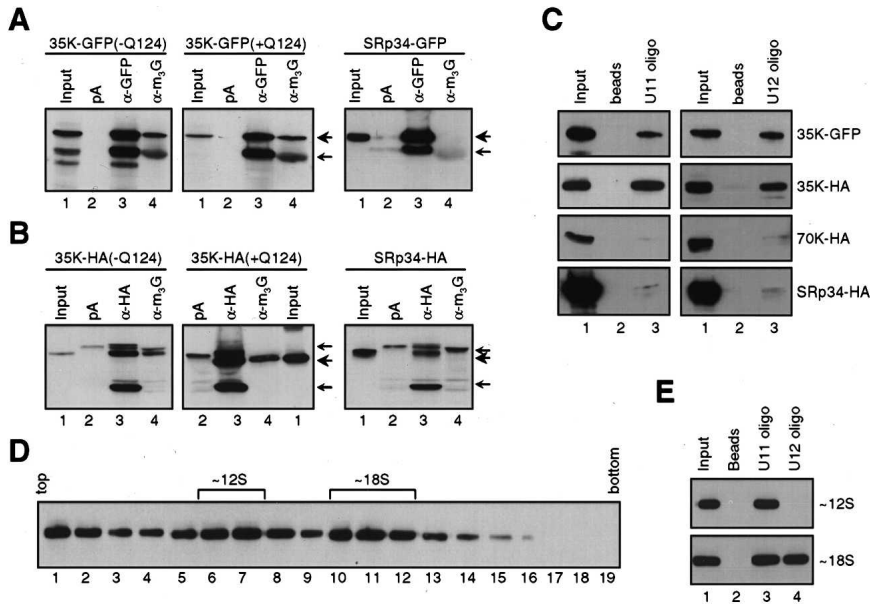
snRNP, similar to that described in human cells for major snRNPs (Sleeman and Lamond 1999).

### *Arabidopsis* U11 and U12 snRNPs exist in a stable U11/U12 di-snRNP complex

We have shown previously that transient expression of U1 and U2 snRNP-specific proteins U1–70K, U2A', and U2B'' in *Arabidopsis* cell suspension protoplasts results in their correct assembly into mature snRNPs (Lorkovic et al. 2004a). To find out whether transiently expressed U11–35K proteins (–Q124 and +Q124) associate with U snRNPs, we expressed GFP and HA-tagged U11–35K proteins in *Arabidopsis* protoplasts and performed immunoprecipitation with anti-m<sub>3</sub>G antibody which recognizes the trimethyl-guanosine Cap structure at the 5' end of U snRNAs (Lührmann et al. 1982). In addition, protein extracts were also immunoprecipitated with either anti-GFP (Fig. 3A) or anti-HA (Fig. 3B) antibodies. Protein extracts from cells expressing GFP- or HA-tagged SRp34 (an *Arabidopsis* SF2/ASF ortholog) were used as controls. Figure 3 demonstrates that both U11–35K proteins (–Q124 and +Q124) are efficiently precipitated with anti-GFP (Fig. 3A, lane 3) and



**FIGURE 2.** (A) *Arabidopsis* U11–35K protein interacts with SR proteins in vitro. Protein extracts from plant protoplasts transformed with plasmids expressing HA-tagged SCL28, SCL30, SCL30a, SCL33, SC35, SRp34, RSp31, and RSp40 proteins or from yeast cells expressing HA-tagged RSZp21, RSZ33, and SRp30 were incubated with glutathione-Sepharose beads coated with *Arabidopsis* U11–35K proteins. (Lanes 1) One-tenth of the input protein extracts used in pull-downs; (lanes 2) protein extracts incubated with glutathione-Sepharose beads only; (lanes 3) pull-down with U11–35K(–Q124); (lanes 4) pull-down with U11–35K(+Q124). Western blotting was performed with anti-HA antibody. (B) *Arabidopsis* U11–35K protein localizes to the nucleus. Tobacco protoplasts were transiently transformed with plasmids expressing GFP-tagged U11–35K proteins and analyzed by laser scanning confocal microscopy. As control, protoplasts were transformed with plasmids expressing GFP-tagged *Arabidopsis* U1–70K protein (the leftmost panel). (C) *Arabidopsis* U11–35K protein (–Q124) colocalizes with SR protein SRp34 in nuclear speckles in both tobacco (upper row; U11–35K and SRp34 were expressed as GFP and RFP fusions, respectively) and *Arabidopsis* (lower row; U11–35K and SRp34 were expressed as RFP and GFP fusions, respectively) cells. All images in B and C are single confocal sections. Merged images show superimposition of GFP and RFP signals. Arrows in B and C point to nucleoli. Bars, 6  $\mu$ m (B,C).



**FIGURE 3.** *Arabidopsis* U11 snRNP is present in a stable U11/U12 di-snRNP complex. (A,B) Transiently expressed U11–35K protein assembles into snRNP. *Arabidopsis* cell suspension protoplasts were transiently transformed with plasmids expressing (A) GFP- or (B) HA-tagged U11–35K (both –Q124 and +Q124) and SRp34, and immunoprecipitations were performed with anti-GFP and anti-m<sub>3</sub>G antibodies (A), or with anti-HA and anti-m<sub>3</sub>G antibodies (B), as described in Materials and Methods. Western blotting was performed with anti-GFP (A) or anti-HA (B) antibodies. Lanes 1 (“input”) in A and B were loaded with 1/10 of the protein extracts used for immunoprecipitations. (Lanes pA) Protein extracts incubated with protein-A Sepharose only; (lanes 3) (α-GFP in A and α-HA in B) immunoprecipitations performed with anti-GFP (A) or anti-HA (B) antibodies; (lanes 4) (α-m<sub>3</sub>G) immunoprecipitations performed with anti-m<sub>3</sub>G cap antibody. Small arrows point to immunoglobulin heavy and light chains. Large arrows point to immunoprecipitated GFP- or HA-tagged proteins. Note that SRp34 is precipitated with tag-specific antibodies but not with anti-m<sub>3</sub>G antibody. (C) *Arabidopsis* U11–35K protein is specifically affinity-selected with U11 snRNA (left panels) and U12 snRNA (right panels) directed antisense oligonucleotides. Protein extracts prepared from *Arabidopsis* cell suspension protoplasts expressing GFP- or HA-tagged U11–35K(–Q124) or HA-tagged U1–70K and SRp34 were incubated with biotinylated oligonucleotides complementary to either U11 or U12 snRNA (lanes 3) or were mock incubated (lanes 2) and then bound to streptavidin agarose (see Materials and Methods). Proteins pulled down were analyzed by Western blotting with anti-GFP and anti-HA antibodies. Lanes 1 (input) were loaded with 1/20 of the protein extract used for affinity selection. (D) Sedimentation behavior of U11–35K protein on glycerol gradient. Protein extracts from *Arabidopsis* protoplasts expressing HA-tagged U11–35K(–Q124) were analyzed by 10%–30% glycerol gradient centrifugation. Gradients were fractionated into 20 0.5-mL fractions that were subsequently analyzed by Western blotting with anti-HA antibody. (E) Oligonucleotide affinity selection of HA-tagged U11–35K protein from ~12S (pooled fractions 6 and 7; upper panel) and ~18S (pooled fractions 10–12; lower panel) glycerol gradient peaks. Lanes 1 were loaded with 1/10 of the extract used for affinity selection in lanes 2, 3, and 4. (Lanes 2) Extract incubated with streptavidin beads only; (lanes 3,4) extract incubated with U11 and U12 antisense oligonucleotides, respectively. Western blotting was performed with anti-HA antibody.

anti-HA (Fig. 3B, lane 3) antibodies. More importantly, GFP- and HA-tagged U11–35K proteins were also precipitated with anti-m<sub>3</sub>G antibody (Fig. 3A,B, lanes 4). As SRp34, which is like U11–35K an RNA-binding protein (Lopato et al. 1999), was not precipitated with anti-m<sub>3</sub>G antibody (Fig. 3A,B, lanes 4 in the rightmost panels), we conclude that precipitation of U11–35K is not due to nonspecific binding to RNA and that transient expression of both versions of U11–35K protein in protoplasts results in incorporation into mature snRNPs.

It has been shown previously that U11 and U12 snRNPs in HeLa nuclear extracts are present not only as individual monoparticles, but also as a stable 18S U11/U12 di-snRNP complex (Wassarman and Steitz 1992), and that such preformed complex interacts directly with pre-mRNA (Frilander and Steitz 1999). To find out whether U11 and U12 snRNPs form a stable complex in plant cell extracts we performed affinity selection using a biotinylated 2'-O-methyl oligonucleotide complementary to nucleotides 11 to 28 of the *Arabidopsis* U12 snRNA (Shukla and Padgett 1999; Will et al. 1999). Affinity selection was performed with protoplast extracts transiently expressing GFP- and HA-tagged U11–35K protein. Figure 3C demonstrates efficient coselection of both GFP- and HA-tagged U11–35K protein with an oligonucleotide directed against U12 snRNA (right panels, lane 3). To exclude the possibility of nonspecific protein–RNA interaction (U11–35K contains an RRM), control experiments with SRp34-HA (contains two RRMs) and with U1–70K (contains one RRM) tagged to HA were performed, but none of the two proteins was coselected with U12 oligonucleotide (Fig. 3C, right panels, lane 3). Originally, this approach was designed to isolate and clone *Arabidopsis* U11 snRNA. However, while this work was in progress, *Arabidopsis* U11 snRNA has been identified (Schneider et al. 2004). Therefore, to prove results obtained with U12 oligonucleotide, affinity selection with an U11 oligonucleotide complementary to nucleotides 2 to 20 of the *Arabidopsis* U11 snRNA was performed. Figure 3C (left panels) demonstrates efficient coselection of both GFP- and HA-tagged U11–35K proteins with the U11 oligonucleotide. In contrast, SRp34 and U1–70K, as in a case of U12 oligonucleotide, were not coselected with U11 oligonucleotide. The dimeric nature of *Arabidopsis* U11 and U12 snRNPs was also addressed by glycerol gradient centrifugation. Transiently expressed, HA-tagged U11–35K protein sedimented in two major peaks, one at ~12S and the second at ~18S (Fig. 3D). Oligonucleotide affinity selection with the two U11–35K peaks from glycerol gradients was performed.

U11 oligonucleotide coselected U11–35K from both glycerol peaks (Fig. 3E, lanes 3), whereas the U12 oligonucleotide coselected U11–35K protein only from 18S peak (Fig. 3E, lanes 4), thus confirming the dimeric nature of U11 and U12 snRNPs in plant cells.

Taken together, we conclude that coselection of U11–35K protein with the U11 and U12 oligonucleotides is specific and that *Arabidopsis* U11 and U12 snRNPs, like their human (Wassarman and Steitz 1992; Tarn and Steitz 1997; Frilander and Steitz 1999) and *Drosophila* (Schneider et al. 2004) counterparts, exist not only as monoparticles but also as a preformed U11/U12 di-snRNP complex.

### Other components of the U11/U12 di-snRNP complex in plants

In HeLa cells the U11/U12 di-snRNP complex contains several other proteins that were not found in U1 and U2 snRNP preparations (Will et al. 1999, 2004). Therefore, we searched public databases for similar proteins in plant species and found that, in addition to the U11–35K protein (Fig. 1B), the 65K, 59K, 48K, 31K, and 25K are also conserved in both dicot and monocot plants (Table 1).

The highest degree of sequence conservation was observed between 65K proteins. This is mainly because of the high sequence conservation in the two RRM (Fig. 4A, overlined), although a proline-rich (Fig. 4A, blue overlined) and an acidic/serine-rich (Fig. 4A, green box) region are also strongly conserved between plant and human 65K proteins. Both *Arabidopsis* and rice 65K proteins are shorter than the human protein (442 and 467 vs. 517 amino acids). This difference is mainly due to deletion of two segments in plant proteins (Fig. 4A, two gaps in the alignment). The 59K proteins from plants show high similarity between themselves, but less so when compared to the human protein. The highest sequence conservation was observed in the middle region comprising ~190 amino acids (Fig. 4B, red box) and in a C-terminal region of ~60 amino acids (Fig. 4B, blue box). The 31K and 25K

are also highly conserved between plants and human (Will et al. 2004; Table 1). Alignments of plant orthologs of 31K and 25K have recently been published (Will et al. 2004) and are therefore not shown here. The lowest sequence homology was observed between the 48K proteins in which only a short region in the middle shows more extensive sequence conservation (Fig. 4C, red box). However, conserved amino acids are found throughout the entire length of the protein, indicating that they are true orthologs. We have identified a rice sequence (included in the alignment) originating from Chromosome 4 having strong similarity with *Arabidopsis* 48K. However, sequencing of this chromosomal region is still in progress, and it is not possible to deduce the full sequence of rice 48K protein (the N terminus is missing). The *Arabidopsis* 48K protein is much longer (and most probably also the rice protein) (see Fig. 4C) than the human one (712 vs. 339 amino acids). This difference is mainly due to a 260-amino-acid extension at the N terminus of the *Arabidopsis* protein. However, the sequence of *Arabidopsis* 48K protein seems to be correct as it is deduced from the full-length cDNA. Despite the extensive BLAST searches, we were unable to unambiguously assign plant orthologs of the human 20K protein. The best *Arabidopsis* score was always U1C protein, which has a domain organization similar to that of 20K protein (Will et al. 2004). It is possible that this protein is functioning in both spliceosomes in plants (see below).

In light of the strong conservation of the spliceosomal machinery between humans and plants (Lorković et al. 2000; Reddy 2001; Brown et al. 2002; Wang and Brendel 2004; Z.J. Lorković, R. Lehner, C. Forstner, and A. Barta, unpubl.), most subunits of the U1 snRNP and all subunits of the U2 snRNPs have been identified in *Arabidopsis* (Table 2). To further characterize the protein composition of *Arabidopsis* U11/U12 di-snRNP, several genes encoding U1 and U2 snRNP-specific proteins have been expressed as GFP and HA fusions and analyzed in our protoplast transient expression system for incorporation into U11 and U12 snRNPs. U2A' and U2B'', which are core components of the U2 snRNP, are efficiently incorporated into U2 snRNP (Lorković et al. 2004a; Fig. 5B, lanes 5,6). However, as shown in Figure 5A (lanes 3), they were not coselected with the U11 and U12 oligonucleotides. The same is true for the SF3a60, a component of the trimeric complex associated with the U2 snRNP (Fig. 5A, lanes 3). In contrast, HA-tagged SF3b49 was clearly found in affinity-selected U11/U12 complex with both U11 and U12 snRNA-directed oligonucleotides (Fig. 5A, lanes 3). Thus, as in HeLa cells, U2A', U2B'', and SF3a are not, whereas SF3b is shared by two different spliceosomes in *Arabidopsis*.

In contrast to human cells, we were not able to demonstrate association of p14, another subunit of the SF3b, with *Arabidopsis* minor snRNPs even at 150 mM KCl concentration (Fig. 5A; data not shown). In light of efficient coselection of SF3b49 with both U11 and U12 snRNA-directed oli-

**TABLE 1.** Plant U11/U12 snRNP proteins

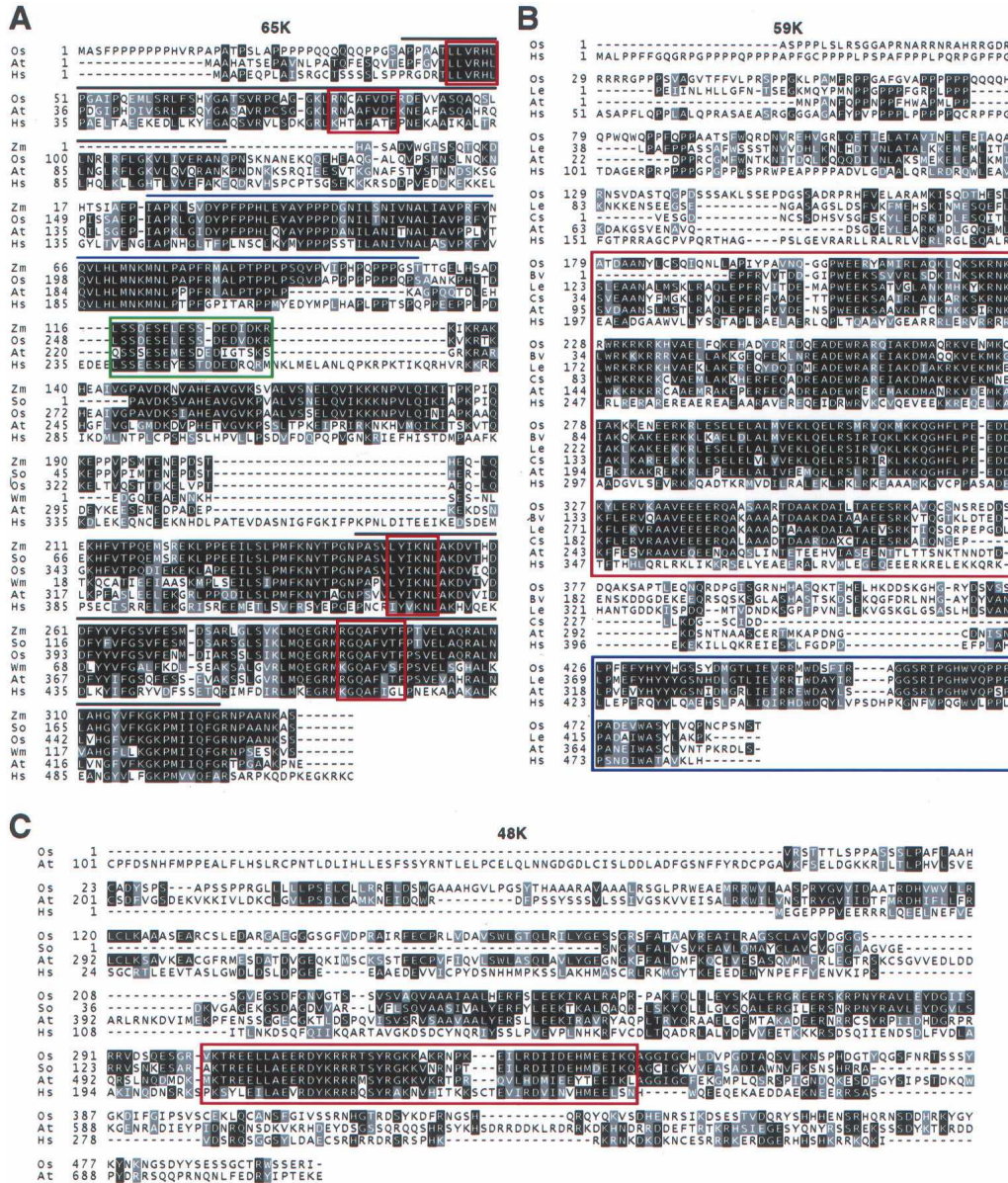
Protein	<i>Arabidopsis</i> accession no. <sup>a</sup>	Rice accession no. <sup>a</sup>
U11–59K	At2g46200 (382)	gi:32976260 <sup>b</sup>
U11–48K	At3g04160 (712)	gi:54325546 <sup>b</sup>
U11–35K	At2g43370 (333)	gi:32982022 (375)
U11/U12–65K	At1g09230 (442)	gi:32992841 (467)
U11/U12–31K	At3g10400 (261)	gi:47522602 <sup>c</sup>
U11/U12–25K	At3g07860 (165)	gi:32489031 <sup>c</sup>
U11/U12–20K	?	?

<sup>a</sup> Numbers in parentheses are protein lengths in amino acids derived from full-length cDNA sequences. The accession numbers refer to DNA sequences at the NCBI.

<sup>b</sup> Not complete at the 5'-end.

<sup>c</sup> Genomic sequences available only.





**FIGURE 4.** Sequence analysis of plant U11/U12 di-snRNP specific proteins. (A) Sequence alignment of U11/U12 di-snRNP 65K proteins from *Arabidopsis thaliana* (At; At1g09230), *Oryza sativa* (Os; gi:32992841), *Saccharum officinarum* (So; gi:35975718), *Zea mays* (Zm; gi:21206676), *Welwitschia mirabilis* (Wm; Sgi:42651093), and *Homo sapiens* (Hs; gi:16553747). The sequences of *S. officinarum*, *Z. mays*, and *W. mirabilis* are deduced from EST clones and are incomplete at their 5' ends. The two RRM motifs are overlined, and the RNP1 and RNP2 motifs of RRM motifs are in red boxes. The proline-rich region is blue overlined, and a short acidic and serine-rich region is in a green box. (B) Sequence alignment of U11 snRNP 59K proteins from *A. thaliana* (At; At2g46200), *O. sativa* (Os; gi:32976260), *Lycopersicon esculentum* (Le; gi:47104949), *Citrus sinensis* (Cs; gi:46213273), *Beta vulgaris* (Bv; gi:34892003), and *H. sapiens* (Hs; gi:22027541). The sequences of *L. esculentum*, *C. sinensis*, and *B. vulgaris* are deduced from EST clones, and they are incomplete at their 5' and 3' ends. The sequence of *O. sativa* is incomplete at the 5' end. The most conserved domains are in red and blue boxes. (C) Sequence alignment of U11 snRNP 48K proteins from *A. thaliana* (At; At3g04160), *S. officinarum* (So; gi:34923260), *O. sativa* (Os; gi:54325546), and *H. sapiens* (Hs; gi:33457355). The sequences of *S. officinarum* and *O. sativa* are deduced from the EST and genomic sequences, respectively, and they are incomplete at both 5' and 3' ends. The N-terminal extension of the *A. thaliana* sequence is not included in the alignment. The most conserved region of the protein is in a red box. Sequences in A–C were aligned by using the ClustalW program and shaded on the BOXSHADE server. The order of sequences is as it appeared in the ClustalW output. Amino acids identical or similar in more than 50% of the analyzed sequences are shown on black or gray background, respectively. The accession numbers of plant proteins refer to their DNA sequences at the NCBI.

gonucleotides, this finding was somewhat surprising. In HeLa cells p14 is located in the central cavity of SF3b (Golas et al. 2003) and is associated with U2 snRNP even at 250 mM NaCl (Will et al. 2001). As in HeLa cells extracts SF3b49 was

found to be incorporated into snRNP at 200 mM KCl, dissociation of p14 from the complex in *Arabidopsis* protoplasts extracts does not seem to be a plausible explanation for this negative result. Furthermore, sequence conservation between

**TABLE 2.** *Arabidopsis* U1 and U2 snRNP proteins

Protein	<i>Arabidopsis</i> gene acc. no.	Identity/similarity (%) with human homolog <sup>a</sup>	Evidence for snRNP incorporation
U1 snRNP			
U1-70K	At3g50670	55/74 <sup>b</sup>	Lorkovic et al. 2004a
U1A	At2g47580	41/51	This work
U1C	At4g03120	70/85 <sup>c</sup>	This work
U2 snRNP			
U2A'	At2g09760	38/49	Lorkovic et al. 2004a
U2B''	At2g30260	49/63	Lorkovic et al. 2004a
	At1g06960	47/61	Sequence similarity
SF3a120	At1g14640	26/41	Sequence similarity
	At1g14650	29/44	Sequence similarity
SF3a66	At2g32600	66/81	Sequence similarity
SF3a60	At5g06160	38/55	This work
SF3b155	At5g64270	65/74	Sequence similarity
SF2b145	At4g21660	45/56	Sequence similarity
SF3b130	At3g55220	57/75	Sequence similarity
SF3b49	At2g18510	72/83 <sup>d</sup>	This work
	At2g14550	68/81 <sup>d</sup>	Sequence similarity
SF3b14a(p14)	At5g12190	63/80	Sequence similarity
SF3b14b	At2g30000	92/95	Sequence similarity
	At1g07170	92/95	Sequence similarity
SF3b10	At3g23325	61/81	Sequence similarity
	At4g14342	60/80	Sequence similarity

<sup>a</sup> Determined by BLAST 2 SEQUENCES at the NCBI.

<sup>b</sup> Alignment of only 154 amino acids containing RRM.

<sup>c</sup> Alignment of N-terminal 51 amino acids only.

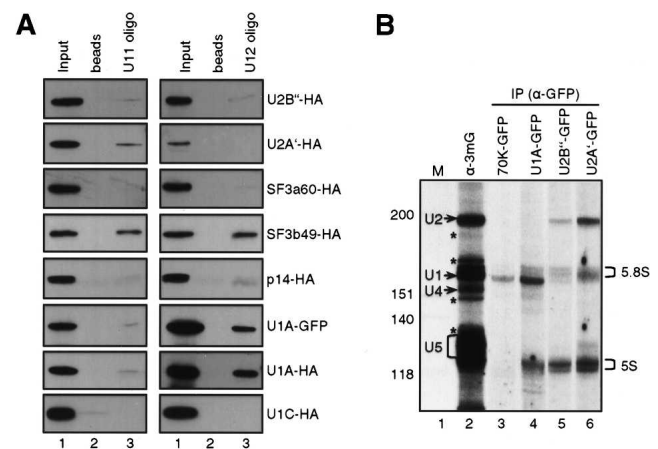
<sup>d</sup> Alignment of two RRMs only.

human and plant p14 is very high (Will et al. 2001; Table 2), indicating that we used the correct protein in our experiments. Interestingly, transiently expressed GFP- and HA-tagged *Arabidopsis* p14 could also not be immunoprecipitated with m<sub>3</sub>G antibody, although we do find p14-GFP in the nucleus (data not shown). Thus, the most likely explanation for this discrepancy between human and plant systems is that the C-terminal tags impair the assembly of this small protein into the SF3b complex.

As we were not able to identify an *Arabidopsis* ortholog of the U11/U12 di-snRNP-specific 20K protein, we asked

whether U1 snRNP-specific U1C protein, which has a domain organization similar to that of 20K (Will et al. 2004), is shared by the two different spliceosomes in plants. However, transiently expressed HA-tagged U1C protein, although being efficiently incorporated into snRNP as evident from immunoprecipitation with anti m<sub>3</sub>G antibody (data not shown), was not found in affinity-selected material with either U11 or U12 oligonucleotide (Fig. 5A). Thus, U1C most probably does not take over the function of 20K protein in plant U11/U12 di-snRNP. It is therefore possible that plants do not have a 20K protein. Alternatively, this protein might exist but available bioinformatics tools do not allow its identification owing to the low sequence conservation. Similarly, high sequence conservation between human and plant U1C protein is restricted to only ~50 amino acids at the N terminus (Table 2).

Interestingly, we repeatedly found that the U1 snRNP-specific U1A protein copurifies with the U12 oligonucleotide but not with the U11 oligonucleotide-selected material (Fig. 5A). As *Arabidopsis* and other plant species do encode a homolog of the human 65K protein (see above), it is unlikely that U1A protein, although being structurally similar to 65K, is a component of both U1 snRNP and U11/U12 di-snRNP. The significance of this finding, if any, is unclear at the moment. Analysis of RNAs coprecipitated with U1A-GFP revealed that, as with 70K-GFP (Fig. 5B,



**FIGURE 5.** Other components of the U11/U12 di-snRNP complex in *Arabidopsis*. (A) Oligonucleotide affinity selection was performed with protein extracts prepared from protoplasts expressing U1 snRNP-specific proteins U1A and U1C, U2 snRNP-specific proteins U2A', U2B'', and SF3a60, U2- and U11/U12-specific proteins SF3b49 and p14 tagged with HA, or with GFP and HA tags (U1A). (Left panels) Affinity selection with U11 oligonucleotide. (Right panels) Affinity selection with U12 oligonucleotide. Details are as in Figure 3C. (B) Transiently expressed GFP-tagged U1-70K (lane 3), U1A (lane 4), U2B'' (lane 5), and U2A' (lane 6) proteins were immunoprecipitated with anti-GFP antibodies and immunoprecipitates were analyzed for the presence of snRNAs by [<sup>32</sup>P]pCp ligation. Four spliceosomal snRNAs (immunoprecipitated with anti-m<sub>3</sub>G cap antibody; lane 2) and sizes of markers in nucleotides (lane 1) are indicated on the left. 5S and 5.8S RNAs, which appear as background, are indicated on the right. Bands marked with asterisks are of unknown identity.



lane 3), U1A-GFP coprecipitated only U1 snRNA (Fig. 5B, lane 4). In addition, U2 snRNP-specific proteins U2B'' and U2A' coprecipitated only U2 snRNA (Fig. 5B, lanes 5,6), indicating correct assembly of transiently expressed snRNP-specific proteins as reported previously (Lorkovic et al. 2004a). Thus, the most likely explanation for copurification of U1A with U12 oligonucleotide would be unspecific RNA binding of free protein to the U12 oligonucleotide in vitro.

Glycerol gradient behavior of *Arabidopsis* U11 snRNP and U11/U12 di-snRNP as well as bioinformatics identification of most U11/U12 di-snRNP proteins in *Arabidopsis* and rice support the argument for highly similar or even possibly the same protein composition of U11/U12 di-snRNP between human and plants. However, the complete comparison will await biochemical purification of plant minor snRNPs, which should also unambiguously show whether p14 and U1A are components of U11/U12 di-snRNP, and whether plants contain a homolog of 20K protein.

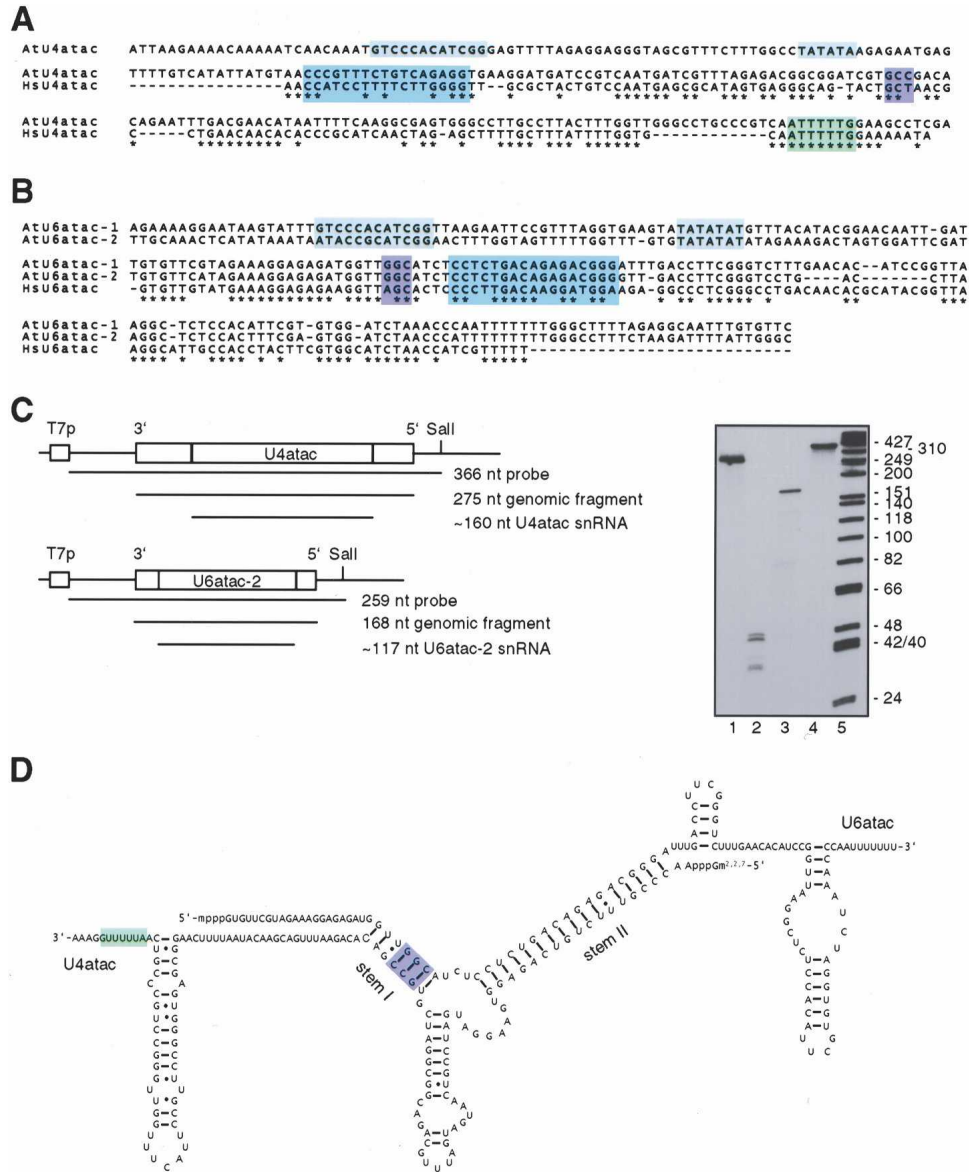
### Identification of the *Arabidopsis* U4atac snRNA

With the exception of U4atac snRNA, all other minor spliceosome-associated snRNAs have been identified in *Arabidopsis* (Shukla and Padgett 1999; Schneider et al. 2004). Although components of the minor spliceosome seem to be highly conserved between plants and human (Shukla and Padgett 1999; Schneider et al. 2004; Will et al. 2004; this work), searches of plant genomes for sequences similar to human U4atac snRNA did not reveal any significant match. Therefore, we used the complementary sequence of *Arabidopsis* U6atac snRNA, which is supposed to make base-pairing (stem II) (Shukla and Padgett 1999) with the U4atac snRNA. These searches revealed several hits; however, more detailed inspection uncovered that only a single hit on Chromosome 4 located in the intergenic region between the predicted genes At4g16060 and At4g16070 could encode a small RNA. Analysis of the sequence upstream of the putative 5' end revealed the presence of promoter elements characteristic for plant U snRNAs (Waibel and Filipowicz 1990). An upstream sequence element (USE) is present at position -76 to -65 (GTCCCACATCGG), and a TATA-like box (TATATA) at position -32 to -27 (Fig. 6A, gray boxes). Alignment of the identified sequence with the human U4atac showed 50% identity with the highest degree of conservation around the Sm-binding site (Fig. 6A, green box). Interestingly, *Arabidopsis* U4atac snRNA contains a 12-nt insertion just before the Sm-binding site. As revealed by RNaseA/T1 protection assay performed with total RNA from *Arabidopsis* seedlings, the identified *Arabidopsis* U4atac gene is expressed as an snRNA of ~160 nt (Fig. 6C, lane 3), which is ~30 nt longer than human U4atac snRNA. In addition, we identified an *Arabidopsis*

gene coding for a second U6atac snRNA (U6atac-2). The predicted snRNA is 90% identical to the previously identified U6atac snRNA (hereafter U6atac-1) (Shukla and Padgett 1999), and it would be 12 nt shorter than U6atac-1 (Fig. 6B). In the RNase A/T1 protection assay we did not observe a protected band of the expected size (~117 nt) (Fig. 6C, lane 2) for U6atac-2 snRNA; therefore, the U6atac-2 seems to be a pseudogene. The short protected bands (Fig. 6C, lane 2) correspond to U6atac-1 snRNA fragments having 100% identity with the U6atac-2 snRNA probe. Surprisingly, the U4atac snRNA sequence that base-pairs with U6atac snRNA (stem II) (Fig. 6A, blue background) is not conserved between human and *Arabidopsis* snRNAs (Fig. 6A). However, the predicted stem II of *Arabidopsis* U4atac/U6atac snRNAs contains only one G•U base pair (Fig. 6D) in contrast to five bases of human U6atac that do not make base pairs with U4atac (Tarn and Steitz 1996). The AGC triad sequence, present in human U6 and U6atac snRNAs, is in both *Arabidopsis* U6atac snRNAs represented by a GGC sequence (Shukla and Padgett 1999; Fig. 6B). However, compensatory mutation that occurred in *Arabidopsis* U4atac snRNA (CCG) allows base-pairing between U4atac and U6atac and formation of stem I (Fig. 6A,D). While mutation at this position in yeast U6 snRNA leads to defects in splicing (Fabrizio and Abelson 1990; Madhani et al. 1990), it has been shown that the A-to-G mutation in the triad of human U6 and U6atac snRNAs is fully compatible with pre-mRNA splicing in vivo (Datta and Weiner 1993; Shukla and Padgett 1999). Also, G at the first triad position of human and *Arabidopsis* U6atac snRNAs does not affect base-pairing with U12 snRNA and the formation of stem Ib, which is important for formation of the catalytically active spliceosome (Shukla and Padgett 1999).

### Concluding remarks

Although plants do possess specific requirements for intron recognition and excision (Lorkovic et al. 2000; Reddy 2001; Brown et al. 2002), protein composition of snRNPs involved in splicing of major (GT-AG) introns, as revealed by bioinformatics analysis (Wang and Brendel 2004; Z.J. Lorković, R. Lehner, C. Forstner, and A. Barta, unpubl.), seems to be highly similar (or even the same) to that in humans. Here, we show that the machinery for excision of minor (AT-AC) introns is likewise highly conserved between plants and humans. Plants express all four minor snRNAs with similar size and predicted secondary structures as well as base-pairing interactions required for formation of the catalytically active spliceosome (Shukla and Padgett 1999; Schneider et al. 2004; Wang and Brendel 2004; this work). Furthermore, we provide experimental evidence that proteins specifically associated with minor snRNPs, in particular those present in U11 and U12 snRNPs, are conserved in both monocot and dicot plants. This indicates that recognition of 5'ss and BPS of



**FIGURE 6.** Analysis of *Arabidopsis* U4atac and U6atac-2 snRNAs. (A) Sequence alignment of *Arabidopsis* and human U4atac snRNAs. *Arabidopsis* sequence includes a 5' end upstream promoter region with two conserved regulatory elements (USE and TATA-box) (Waibel and Filipowicz 1990) printed on a gray background. The Sm-binding site is on a green background. Sequence complementary to the GGC triad of U6atac snRNA is printed on a lilac background. Sequence that base-pairs with U6atac in stem II is printed on a blue background. (B) Sequence alignment of human and two *Arabidopsis* U6atac snRNAs. *Arabidopsis* sequences include a 5' upstream promoter sequence with USE and TATA-boxes printed on a gray background. The GGC triad and the region that base-pairs with U4atac snRNA in stem II are printed on lilac and blue backgrounds, respectively. (C) Expression analysis of the *Arabidopsis* U4atac and U6atac-2 snRNAs as determined by RNaseA/T1 protection assay. For this, 10  $\mu$ g of total RNA from *Arabidopsis* seedlings was hybridized with the antisense U4atac and U6atac-2 probes. After RNaseA/T1 protection assay the RNA was run on a 6% denaturing polyacrylamide gel. (Lane 1) U6atac-2 snRNA probe; (lane 2) U6atac-2 snRNA protection assay; (lane 3) U4atac snRNA protection assay; (lane 4) U4atac snRNA probe. The sizes of the  $^{32}$ P-labeled DNA markers (lane 5) in nucleotides are indicated on the right. On the right side are schematic representations of U4atac and U6atac-2 genomic fragments in pGEM T-easy. (D) Base-pairing interactions between *Arabidopsis* U4atac and U6atac snRNAs. Potential RNA-RNA base pairs (Watson-Crick) are indicated by dashes and G•U interactions by dots. Sm-binding site of U4atac snRNA is green shaded, and the GGC triad of the U6atac snRNA in stem I is printed on a lilac background.

minor introns and bridging of the reactive groups of the pre-mRNA within the minor spliceosome are also similar or even the same in higher plants and human. In human cells the majority of the proteins found in the major U4/U6•U5 tri-snRNP are also present in the minor

U4atac/U6atac•U5 tri-snRNP and are also highly conserved in plants. Together, these data strongly support the argument that minor introns and the entire machinery for their excision existed in a eukaryotic ancestor before divergence of plants and metazoans.

## MATERIALS AND METHODS

### Cloning of cDNAs encoding the *Arabidopsis* U11 snRNP-specific protein 35K

cDNA clones encoding *Arabidopsis* U11–35K protein were PCR-amplified from a yeast two-hybrid cDNA library (Clontech). Oligonucleotides, ATGAGCGGCGGAGGAAACAACGTC and GTCGGAAGAATGATGATCTTGAC, were derived from the genomic sequence (At2g43370) and from the partial cDNA sequence obtained in a yeast two-hybrid screen with *Arabidopsis* cyclophilin CypSR92 (Lorkovic et al. 2004b). Two different PCR products were cloned and sequenced (see Results).

### Plasmid constructions

Plasmids for overexpression of U11–35K proteins were obtained by cloning corresponding PCR products into pGEX-4T-1 (Amersham Pharmacia Biotech). cDNAs were amplified by using the oligonucleotides TATTGCGAATTCATGAGCGGCGGAGGAAACAACGT and GTCAGCGTCGACTTAGTCGGAA GAATGATGATCTTGAC, which introduce an EcoRI (in bold) site in front of the ATG codon and a Sall site after the stop codon, respectively, and ligated into corresponding sites of pGEX-4T-1, resulting in pGST-35K(–Q124) and pGST-35K(+Q124). To generate plant expression plasmids encoding HA-, GFP-, and RFP-tagged U11–35K protein, the coding region of U11–35K was amplified by using oligonucleotides TGCGGGTGCACAATAAACCATGAGCGGCGGAGGAAACAACGT and TATTGCGGATCCGTCGGAAGAATGATGATCTTGAC, which introduce a Sall (bold) site and plant translation consensus sequence (italics) in front of the ATG codon and a BamHI site in place of the stop codon, respectively. The PCR products were cut with Sall and BamHI and ligated into corresponding sites of the plant expression vectors pDEDH-HA, pDEDH-GFP (Lambermon et al. 2000) and pDEDH-RFP (Lorkovic et al. 2004a), resulting in p35K-HA, p35K-GFP, and p35K-RFP. For expression of HA-tagged U11–35K in yeast, the cDNA was PCR-amplified (the oligonucleotides used were GTATCACCATGGGCGGCGGAGGAAACAACG and GTCGA CCCGGGTCGGAAGAATGATGATCTTGA) and ligated into the NcoI/XmaI-opened pYX242 (Novagen), resulting in pYX-35K. For expression of GFP- and HA-tagged U1A, U1C, SF3a60, SF3b49, and p14 in plant cells, the coding regions were PCR-amplified from a yeast two-hybrid library (Clontech) with the following oligonucleotides:

U1A 5'-primer, ACTAGCCCGGGAATAAACCATGGAGATGC AAGAGGCTAATC, and 3'-primer, AGCATGGATCCCTTCTTG GCATACGTGATG;

U1C 5'-primer, GATCGGTGCACAATAAACCATGCCGAGGTA TTACTGTG, and 3'-primer, AGCTGGATCCCTCATTGCTTTC AGGAGATTC;

SF3a60 5'-primer, GATCGCCCGGGAATAAACCATGTCGTCGA CTCTCCTCGAG, and 3'-primer, AGCATGGATCCGATGAGA CCTTGGCGTTG;

SF3b49 5'-primer, ACTAGGTCGACAATAAACCATGACGACTC GAATCGCTC, and 3'-primer, AGCATGGATCCACCTTGGTG TGGTGGTGGTG; and

p14 5'-primer, GATCGCCCGGGAATAAACCATGACAACAATC

AGTCTTCG, and 3'-primer, AGCATGGATCCCTTATCTTTG GTAGAGAC.

The 5'-primers introduce an Sall (U1C and SF3b49) or XmaI (U1A, SF3a60, and p14) (bold) site and plant translation consensus sequence (italics) in front of the ATG codon and BamHI site (bold) in place of the stop codon, respectively. PCR products were ligated into Sall/BamHI- or XmaI/BamHI-opened pDEDH-HA and pDEDH-GFP, resulting in pU1A-HA, pU1A-GFP, pU1C-HA, pU1C-GFP, pSF3a60-HA, pSF3a60-GFP, pSF3b49-HA, pSF3b49-GFP, p14-HA, and p14-GFP. Plant expression plasmids p70K-GFP, p70K-RFP, p70K-HA, pSRp34-GFP, pSRp34-RFP, pSRp34-HA, pU2A'-GFP, pU2A'-HA, pU2B''-GFP, and pU2B''-HA encoding GFP-, RFP-, and HA-tagged *Arabidopsis* U1–70K, SRp34, U2A', and U2B'' proteins have been described (Lorkovic et al. 2004a). Plant expression plasmids pSCL28-HA, pSCL30-HA, pSCL30a-HA, pSCL33-HA, pSCL35-HA, pSRp30-HA, pSRp31-HA, and pSRp40-HA for the expression of HA-tagged SR proteins have been described (Lorkovic et al. 2004a). Yeast expression plasmids pRSZp21-HA and pRSZ33-HA have been described earlier (Lopato et al. 2002).

### Overexpression and purification of GST-fusion proteins

The plasmids pGST-35K(+Q124) and pGST-35K(–Q124) were transformed into *Escherichia coli* strain BL21 (DE3) CodonPlus-RIL (Stratagene). Overnight cultures grown at 37°C in the presence of 100 µg/mL of ampicillin and 40 µg/mL of chloramphenicol were diluted 100 times and grown further to an OD<sub>600</sub> of 1.0. Protein synthesis was induced with 1 mM IPTG at 28°C for 1.5 h. Cell pellets from 200 mL of culture were resuspended in 20 mL of lysis buffer (20 mM Tris-HCl at pH 7.5, 1 M NaCl, 0.2 mM EDTA, 1 mM DTT, 1% Triton X-100, and EDTA-free protease inhibitor cocktail from Roche) and disrupted with a French press (1000 p.s.i.). Cleared lysates were mixed with 300 µL of glutathione-Sepharose beads (Amersham Pharmacia Biotech) and mixed for 20 min at 4°C. Beads were washed with 100 mL of lysis buffer. For pull-down experiments, beads with bound proteins were stored at 4°C in protoplast extraction buffer (PEB150; see below). For each pull-down, 50 µL of beads was used (~1 µg of bound protein).

### Preparation and transient transformation of tobacco and *Arabidopsis* protoplasts

Tobacco leaf mesophyll protoplasts were isolated and transformed with 20 µg of plasmid DNA per 10<sup>6</sup> protoplasts by the polyethylene glycol method (Koop et al. 1996). *Arabidopsis* cell suspension protoplasts were isolated and transformed as described (Lorkovic et al. 2004a). Transformed protoplasts were collected 24 h after transformation and stored at –80°C or were analyzed by a laser scanning confocal microscope (Leica).

### Preparation of whole cell extracts from protoplasts and from yeast cells, immunoprecipitation, pull-down assay, and glycerol gradient centrifugation

Protoplasts were collected by centrifugation 24 h after transformation (15 min, 70 g), frozen in liquid nitrogen, and resuspended in protoplast extraction buffer (PEB400; 50 mM HEPES-KOH at

pH 7.9, 400 mM KCl, 2.5 mM MgCl<sub>2</sub>, 1 mM EDTA, 1 mM DTT, 0.1% Triton X-100), supplemented with EDTA-free protease inhibitor cocktail (Roche), and further processed as described in Lorković et al. (2004a). After 15 min of centrifugation in an Eppendorf centrifuge at maximum speed at 4°C, the supernatant was mixed with PEB without KCl to adjust KCl concentration to 150 mM (PEB150; for GST pull-down assay) or to 250 mM (PEB250; for immunoprecipitation and oligonucleotide affinity selection). Immunoprecipitations and pCp labeling of snRNAs were performed as described in Lorković et al. (2004a). Preparation of protein extracts from yeast cells expressing RSZp21-HA, RSZ33-HA, and SRp30-HA were as described in Lopato et al. (2002). For pull-down assays protein extracts from transformed protoplasts or from yeast cells were mixed with glutathione Sepharose beads (AmershamPharmacia Biotech) coated with recombinant U11–35K proteins or with beads alone and incubated on a rotary shaker for 2 h at 4°C. After four washings with PEB150, the beads were resuspended in 50 µL of loading buffer, and 25 µL was analyzed by SDS-PAGE.

For fractionation of snRNPs on 10%–30% (w/w) glycerol gradients prepared in PEB150 total protein extracts (KCl concentration adjusted to 150 mM) from eight transformations [HA-tagged U11–35K(–Q124)] were concentrated and loaded onto two gradients. Gradients were centrifuged at 28,000 rpm for 17 h at 4°C in a Beckman SW28 rotor. Each gradient was fractionated into 20 0.5-mL fractions that were pooled, and proteins were subsequently precipitated with acetone. Precipitated proteins were resuspended in 100 µL of loading buffer, and 25 µL of each fraction was analyzed by SDS-PAGE and Western blotting.

### Affinity selection with U11 and U12 oligonucleotides

For oligonucleotide affinity selection, protein extracts from transformed protoplasts were incubated overnight at 4°C with biotinylated 2'-O-methyl oligonucleotide complementary to nucleotides 11 to 28 of the *Arabidopsis* U12 snRNA or to nucleotides 2 to 20 of the *Arabidopsis* U11 snRNA (U12, GUUUUCCU UACUCAUUAGTTTT; U11, AUAGCCAUCCGACCCUUUUTTT; where A, C, G, and Us are 2'-O-methyl ribonucleotides and four Ts are biotinylated deoxythymidines). Streptavidin agarose beads (Pierce) were added to extracts and incubated an additional 2 h at 4°C. Beads were washed four times with PEB250, and finally resuspended in 40 µL of SDS-PAGE loading buffer.

### Analysis of *Arabidopsis* U4atac and U6atac-2 snRNAs

DNA fragments encoding U4atac and U6atac-2 were PCR-amplified from the *Arabidopsis* genomic DNA by using oligonucleotides U4atac-5', GAGGGTAGCGTTTCTTTGGCCT; U4atac-3', CAG TTTAATTGGTTTGATCCTG; U6atac-5', GAAAGACTAGTG GATTCCG; and U6atac-2-3', AGCGAGAGGCCCAATAA. PCR fragments of 275 nt (U4atac) and 168 nt (U6atac-2) were cloned into pGEM T-easy (Promega) and sequenced. For expression analysis, plasmids containing cloned U4atac and U6atac-2 were linearized with Sall, and antisense RNA probes were transcribed with T7 RNA polymerase (Promega) in the presence of [<sup>32</sup>P]UTP (800 Ci/mmol). Probe purification, RNaseA/T1 protection assay, and denaturing PAA gel electrophoresis were as described in Goodal et al. (1990). Total RNA from *Arabidopsis*

seedlings was isolated with the RNeasy Plant Mini Kit (QIAGEN). The φX174 HinfI dephosphorylated DNA markers were labeled according to the procedure provided by the manufacturer (Promega).

### SDS-PAGE and Western blotting

12% SDS-PAGE was done according to standard procedure. Proteins were transferred onto a PVDF membrane (Millipore), and Western blotting was performed according to standard procedure. Rat anti-HA monoclonal (3F10; Roche) and mouse anti-GFP (Roche) monoclonal antibodies were used at 1:1000 dilution. Secondary antibodies, goat anti-rat (Sigma), and goat anti-mouse (Biorad) IgG-conjugated with horseradish peroxidase were used at 1:10,000 dilutions. The blots were developed using an enhanced chemoluminescence kit (AmershamPharmacia Biotech).

### ACKNOWLEDGMENTS

We thank Reinhard Lührman for anti-m<sub>3</sub>G antibody and Maria Kalyna for critical comments on the manuscript. This work was supported by grants from the Austrian Science Foundation (SFB-F017C11 and SFB-F017C12) to A.B.

Received March 2, 2005; accepted April 7, 2005.

### REFERENCES

- Brown, J.W., Simpson, C.G., Thow, G., Clark, G.P., Jennings, S.N., Medina-Escobar, N., Haupt, S., Chapman, S.C., and Oparka, K.J. 2002. Splicing signals and factors in plant intron removal. *Biochem. Soc. Trans.* **30**: 146–149.
- Burge, C.B., Padgett, R.A., and Sharp, P.A. 1998. Evolutionary fates and origins of U12-type introns. *Mol. Cell* **2**: 773–785.
- Burge, C.B., Tuschl, T., and Sharp, P.A. 1999. Splicing of precursors to mRNAs by the spliceosome. In *The RNA world* (eds. R.F. Gesteland et al.), pp. 525–560. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Datta, B. and Weiner, A.M. 1993. The phylogenetically invariant ACA GAGA and AGC sequences of U6 small nuclear RNA are more tolerant of mutations in human cells than in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **13**: 5377–5382.
- Fabrizio, P. and Abelson, J. 1990. Two domains of yeast U6 small nuclear RNA required for both steps of nuclear precursor messenger RNA splicing. *Science* **250**: 404–409.
- Frilander, M.J. and Steitz, J.A. 1999. Initial recognition of U12-dependent introns requires both U11/5' splice-site and U12/branchpoint interactions. *Genes & Dev.* **13**: 851–863.
- . 2001. Dynamic exchange of RNA interactions leading to catalytic core formation in the U12-dependent spliceosome. *Mol. Cell* **7**: 217–226.
- Golas, M.M., Sander, B., Will, C.L., Luhrmann, R., and Stark, H. 2003. Molecular architecture of the multiprotein splicing factor SF3b. *Science* **300**: 980–984.
- Golovkin, M. and Reddy, A.S. 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.
- Goodal, G.J., Wiebauer, K., and Filipowicz, W. 1990. Analysis of pre-mRNA processing in transfected plant protoplasts. *Methods Enzymol.* **181**: 148–161.
- Hall, S.L. and Padgett, R.A. 1994. Conserved sequences in a class of rare eukaryotic introns with non-consensus splice sites. *J. Mol. Biol.* **239**: 357–365.

- Hastings, M.L. and Krainer, A.R. 2001. Functions of SR proteins in the U12-dependent AT-AC pre-mRNA splicing pathway. *RNA* **7**: 471–482.
- Jackson, I.J. 1991. A reappraisal of non-consensus mRNA splice sites. *Nucleic Acids Res.* **19**: 3795–3798.
- Kohtz, J.D., Jamison, S.F., Will, C.L., Zuo, P., Lührmann, R., Garcia-Blanco, M.A., and Manley, J.L. 1994. Protein–protein interactions and 5′-splice site recognition in mammalian mRNA precursors. *Nature* **368**: 119–124.
- Koop, H.U., Steinmüller, K., Wagner, H., Rossler, C., Eibl, C., and Sacher, L. 1996. Integration of foreign sequences into the tobacco plastome via polyethylene glycol-mediated protoplast transformation. *Planta* **199**: 193–201.
- Krämer, A. 1996. The structure and function of proteins involved in mammalian pre-mRNA splicing. *Annu. Rev. Biochem.* **65**: 367–409.
- Lambermon, M.H., Simpson, G.G., Wiczonek Kirk, D.A., Hemmings-Mieszczak, M., Klahre, U., and Filipowicz, W. 2000. UBPI, a novel hnRNP-like protein that functions at multiple steps of higher plant nuclear pre-mRNA maturation. *EMBO J.* **19**: 1638–1649.
- Lewandowska, D., Simpson, C.G., Clark, G.P., Jennings, N.S., Barciszewska-Pacak, M., Chiao-Feng, L., Makalowski, W., Brown, J.W.S., and Jarmolowski, A. 2004. Determinants of plant U12-dependent intron splicing efficiency. *Plant Cell* **16**: 1340–1352.
- Lopato, S., Kalyna, M., Dorner, S., Kobayashi, R., Krainer, A.R., and Barta, A. 1999. atRSp30, one of two SF2/ASF-like proteins from *Arabidopsis thaliana*, regulates splicing of specific plant genes. *Genes & Dev.* **13**: 987–1001.
- Lopato, S., Forstner, C., Kalyna, M., Hilscher, J., Langhammer, U., Indrapichate, K., Lorkovic, Z.J., and 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.
- Lorković, Z.J., Wiczonek Kirk, D.A., Lambermon, M.H., and Filipowicz, W. 2000. Pre-mRNA splicing in higher plants. *Trends Plant Sci.* **5**: 160–167.
- Lorković, Z.J., Hilscher, J., and Barta, A. 2004a. Use of the fluorescent protein tags to study nuclear organization of the spliceosomal machinery in transiently transformed living plant cells. *Mol. Biol. Cell* **15**: 3233–3243.
- Lorković, Z.J., Lopato, S., Pexa, M., Lehner, R., and Barta, A. 2004b. 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.* **33890**–33898.
- Lührmann, R., Appel, B., Bringmann, P., Rinke, J., Rothe, S., and Bald, R. 1982. Isolation and characterisation of a rabbit anti-m3 2,2,7 G antibodies. *Nucleic Acids Res.* **10**: 7103–7113.
- Madhani, H., Bordonne, R., and Guthrie, C. 1990. Multiple roles for U6 snRNA in the splicing pathway. *Genes & Dev.* **4**: 2264–2277.
- Moore, M.J., Query, C.C., and Sharp, P.A. 1993. Splicing of precursors to mRNA by the spliceosome. In *The RNA world* (eds. R.F. Gesteland et al.), pp. 303–357. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Patel, A.A. and Steitz, J.A. 2003. Splicing double: Insights from the second spliceosome. *Nat. Rev. Mol. Cell. Biol.* **4**: 960–970.
- Query, C.C., Bentley, R.C., and Keene, J.D. 1989. A common RNA recognition motif identified within a defined U1 RNA binding domain of the 70K U1 snRNP protein. *Cell* **57**: 89–101.
- Reddy, A.S.N. 2001. Nuclear pre-mRNA splicing in plants. *CRC Crit. Rev. Plant Sci.* **20**: 523–571.
- Schneider, C., Will, C.L., Makarova, O.V., Makarov, E.M., and Lührmann, R. 2002. Human U4/U6•U5 and U4atac/U6atac•U5 tri-snRNPs exhibit similar protein composition. *Mol. Cell. Biol.* **22**: 3219–3229.
- Schneider, C., Will, C.L., Brosius, J., Frilander, M.J., and Lührmann, R. 2004. Identification of an evolutionarily divergent U11 small nuclear ribonucleoprotein particle in *Drosophila*. *Proc. Natl. Acad. Sci.* **101**: 9584–9589.
- 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.
- Sleeman, J.E. and Lamond, A.I. 1999. Nuclear organization of pre-mRNA splicing factors. *Curr. Opin. Cell Biol.* **11**: 372–377.
- Tarn, W.Y. and Steitz, J.A. 1996. Highly diverged U4 and U6 small nuclear RNAs required for splicing rare AT-AC introns. *Science* **273**: 1824–1832.
- . 1997. Pre-mRNA splicing: The discovery of a new spliceosome doubles the challenge. *Trends Biochem. Sci.* **22**: 132–137.
- Waibel, F. and 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 U1-snRNA genes. *Nucleic Acids Res.* **18**: 3451–3458.
- Wang, B.B. and Brendel, V. 2004. The ASRG database: Identification and survey of *Arabidopsis thaliana* genes involved in pre-mRNA splicing. *Genome Biol.* **5**: R102.
- Wassarman, K.M. and Steitz, J.A. 1992. The low-abundance U11 and U12 small nuclear ribonucleoproteins (snRNPs) interact to form a two-snRNP complex. *Mol. Cell. Biol.* **12**: 1276–1285.
- Will, C.L., Schneider, C., Reed, R., and Lührmann, R. 1999. Identification of both shared and distinct proteins in the major and minor spliceosomes. *Science* **284**: 2003–2005.
- Will, C.L., Schneider, C., MacMillan, A.M., Katopodis, N.F., Neubauer, G., Wilm, M., Lührmann, R., and Query, C.C. 2001. A novel U2 and U11/U12 snRNP protein that associates with the pre-mRNA branch site. *EMBO J.* **20**: 4536–4546.
- Will, C.L., Schneider, C., Hossbach, M., Urlaub, H., Rauhut, R., Elbashir, S., Tuschl, T., and 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.
- Wu, Q. and Krainer, A.R. 1999. AT-AC pre-mRNA splicing mechanisms and conservation of minor introns in voltage-gated ion channel genes. *Mol. Cell. Biol.* **19**: 3225–3236.
- Wu, J.Y. and Maniatis, T. 1993. Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell* **75**: 1061–1070.
- Wu, H.-J., Gaubier-Comella, P., Delseny, M., Gellet, F., Van Montagu, M., and Rouze, P. 1996. Non-canonical introns are at least 10<sup>9</sup> years old. *Nat. Genet.* **14**: 383–384.
- Yu, Y.T. and Steitz, J.A. 1997. Site-specific crosslinking of mammalian U11 and U6atac to the 5′ splice site of an AT-AC intron. *Proc. Natl. Acad. Sci.* **94**: 6030–6035.
- Zhu, W. and Brendel, V. 2003. Identification, characterisation and molecular phylogeny of U12-dependent introns in the *Arabidopsis thaliana* genome. *Nucleic Acids Res.* **31**: 4561–4572.