Network of Interactions of a Novel Plant-specific Arg/Ser-rich Protein, atRSZ33, with atSC35-like Splicing Factors*

Sergiy Lopato‡, Christina Forstner, Maria Kalyna, Julia Hilscher, Ulrike Langhammer, Korakod Indrapichate†, Zdravko J. Lorkovic, and Andrea Bartā¶

From the Institut für Medizinische Biochemie, Universität Wien, Vienna BioCenter, Dr. Bohrgasse 9/3, Wien A-1030, Austria

Arg/Ser-rich (RS) proteins play a crucial role in splicing and are implicated in splice site selection in metazoans. In plants, intron recognition seems to differ from the one in animals due to specific factor requirements. Here we describe a new plant-specific RS-rich protein, atRSZ33, with a unique domain structure consisting of an RNA recognition motif (RRM), two zinc knuckles embedded in a basic RS region, and an acidic C-terminal domain. atRSZ33 was found to be a phosphoprotein that concentrates in nuclear speckles and is predominantly present in roots and flowers. In a yeast two-hybrid screen, atRSZ33 interacted with splicing factors atSRp34/SR1, an Arabidopsis ortholog of human SF2/ASF; atRSZp21 and atRSZp22, which are similar to the human 9G8; and three novel SC35-like splicing factors termed atSCL28, atSCL30, and atSCL33/SR33. Two further members of the SCL family, namely SCL30a and the ortholog of mammalian SC35, atSC35, were also found to interact with atRSZ33. These interactions were verified by in vitro binding assays; furthermore, the transcriptional activity of atRSZ33 was found to overlap with the ones of its interacting partners. These specific interactions coupled with the many similarities of atRSZ33 to SR proteins suggest that its main activity is in spliceosome assembly. Mapping of regions necessary for protein-protein interaction between atRSZ33 and atSCL33/SR33 revealed that both zinc knuckles together with a small part of the RS and the RRM domain are required for efficient binding. However, the interacting domain is relatively small, allowing binding of additional proteins, a feature that is consistent with the proposed role of atRSZ33 in spliceosome assembly.

Eukaryotes have enlarged their protein-coding potential by the combinatorial removal of intervening sequences during the splicing process. Nuclear pre-mRNA processing takes place in the spliceosome, a large complex, consisting of five small nuclear ribonucleoproteins (U1, U2, U4/U6 and U5 snRNPs) and numerous non-snRNPs proteins. The precise excision of an intron is accomplished by complementary binding of U1 and U2 snRNPs to the 5′ and 3′ splice sites, respectively. A network of proteins then assembles and brings the two splice sites into close proximity (reviewed in Refs. 1 and 2). The consecutive transesterification steps then generate the mature mRNA. The rearrangements necessary for these two steps make the spliceosome a dynamic particle with varying composition of structural proteins and enzymes (reviewed in Ref. 3).

Investigation of proteins involved in the splicing process has shown that many of them have a modular structure consisting of one or two RNA binding domains, which could be the RNA recognition motif (RRM), K homology domain, zinc knuckle domain, or RGG box, and auxiliary domains, which include Arg/Ser (RS)-rich regions. There is a growing family of RS proteins; the first ones identified were a group of proteins termed SR (serine/arginine) proteins, which contain one or two RRM domains, and usually a C-terminal RS domain with many SR dipeptides (reviewed in Refs. 4 and 5). The SR proteins range in size from 20 to 75 kDa, and they are phosphoproteins that can be precipitated by high magnesium ion concentrations. The best characterized members of this group are SF2/ASF, SC35, and SRp20. Other proteins containing RS domains include splicing factors like U2AF65, U2AF35, tra, tra2, U1–70K, and SmRm160, which have a modular protein structure distinct from the SR proteins but are otherwise functionally related (reviewed in Refs. 4 and 6). By binding to splicing enhancer regions SR proteins determine the selection of alternative splice sites and spliceosome assembly (reviewed in Refs. 6–8). For example, SF2/ASF helps to define the 5′ splice site by binding to the pre-mRNA and to the U1 snRNP protein U1–70K (9, 10). The large subunit of the U2 auxiliary factor, U2AF65, binds to the polypyrimidine tract between the 3′ splice site and the branch point, whereas the small subunit U2AF35 has recently been shown to bind to the highly conserved AG nucleotides at the 3′-end of the intron (11–15). Interactions between members of the SR protein family are crucial for constitutive and enhancer-dependent splicing (9, 16). It has been shown that SC35 and SF2/ASF interact with both U1–70K and U2AF35 to bridge 5′ and 3′ splice sites (9, 17, 18). In general, SR proteins are thought to recruit the factors necessary for spliceosome assembly through either protein-protein or RNA-protein (enhancer) interactions. Interestingly, some of the RS proteins have been found to bind to the C-terminal

* This work was supported by grants (P13634-GEN and SFB-F017/10/11) from the Österreichischer Forschungsfonds zur Förderung der wissenschaftl. Forschung (to A. B.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† Present address: Dept. of Plant Science, Waite Campus, PMB 1, Glen Osmond, South Australia 5064, Australia.
‡ Present address: School of Biology, Institute of Science, Suranaree University of Technology, Nakhon Ratchasima 30000, Thailand.
¶ To whom correspondence should be addressed. Tel.: 43-1-4277-6164; Fax: 43-1-4277-9616; E-mail: Andrea@bch.univie.ac.at.

1 The abbreviations used are: snRNP, small nuclear ribonucleoprotein; RS, arginine/serine; SR, serine/arginine; RRM, RNA recognition motif; GST, glutathione S-transferase; GUS, β-glucuronidase; GFP, green fluorescent protein; UTR, untranslated region; EST, expressed sequence tag; HA, hemagglutinin; RT, reverse transcriptase.
domain of RNA polymerase II, suggesting they may have a function in the coordination of transcription and splicing (reviewed in Refs. 19 and 20).

Protein-protein interactions have been shown to be mediated by the RS domain, but this domain also has a role in RNA binding and the modulation of RNA-RNA interactions (6, 7, 21). Phosphorylation/dephosphorylation cycles of the RS domain by SR protein kinase or Ctk/Sty kinase play a critical role in splicing (reviewed in Ref. 22). The phosphorylation status of SR proteins influences their binding interactions (23), distribution within the nucleus, and recruitment to transcription initiation sites (24, 25).

In plants, little is known about their splicing factors (reviewed in Ref. 26). However, the basic splicing mechanism appears to be the same in both plants and animals (reviewed in Refs. 26 and 27), although plants are not able to recognize animal introns (28). Therefore, there seems to be a difference in intron/exon definition in plants, and recent studies have focused on factors that might be important for intron recognition.

Generally, plant introns have a higher AU content than exons, and the presence of short U-rich stretches is required for efficient intron removal (29; reviewed in Refs. 26 and 27). Therefore, some effort has gone into identification and characterization of proteins, which bind U-rich pre-mRNA sequences (29–31). On the other hand, SR proteins are known to be crucial for splice site selection in metazoan, and it was therefore important to show that this protein family also exists in plants and had similar functions (33, 34). It was also demonstrated that Arabidopsis has at least two SF2/ASF-like factors termed atSRp30 and atSRp34/SR1, with different expression patterns (35). In human, a family of zinc knuckle-containing SR proteins, similar in their activities to human 9G8 and SRp20, termed atSRp31 family (including atSRp31, atSRp40, and atSRp41) were shown to act as splicing factors (38).

To isolate further plant SR proteins we identified a gene from Arabidopsis with high sequence similarity to human SRp55 (39) in the RRm region, but differing considerably in other parts of the protein. The encoded protein contains two zinc knuckles inserted into an RS domain followed by an SP domain at the C-terminal end. The protein, termed atRSZ33, is a nuclear phosphoprotein and has no known homolog in metazoan.

The protein was found to interact with known plant splicing factors and with a new family of SC35-like proteins suggesting a role in plant splicing assembly.

**EXPERIMENTAL PROCEDURES**

**Cloning of atRSZ33 and Sequence Analysis—Two Arabidopsis EST clones (GenBank™ accession numbers T42174 and T75762) were found by a data base search to have sequence similarity to the RNA recognition motif (RRM) of the human SRp55 and SRp75 splicing factors (39, 40). Sequencing of the entire cDNAs revealed that they were identical, encoding a protein, which we named atRSZ33. A genomic clone encoding atRSZ33 was PCR-amplified from genomic DNA as the template with primers derived from the ends of the cDNA. 2.7 kb of the promoter region was isolated using a GenomeWalker kit (Clontech).

The nucleotide sequences obtained were verified by comparison with the genomic sequence, which in the meantime appeared in the Arabidopsis genome database corresponding to genes At2g137340 and At2g137350. No sequence differences were detected, but the predictions of intron/exon regions in the Arabidopsis genome data base were wrong.

The cDNA of alternatively spliced mRNA was cloned by RT-PCR using mRNA from two-day-old seedlings. EST clones with the same alternative splicing pattern were found in the EST data base (GenBank™, accession numbers H36789 and N57114). Intron-exon junction sequences were determined by comparison of genomic and cDNA sequences.
into the E. coli expression strain BL21-CodonPlus-RIL (Stratagene). Overnight cultures grown at 37 °C in LB, in the presence of 100 μg/ml ampicillin and 40 μg/ml chloramphenicol, were diluted 100 times and further grown at 37 °C for 3 h. Protein expression was induced with 1 mM isopropyl-β-D-thiogalactopyranoside at an A600nm of 1.0 for 2 h at 30 °C. Cells were harvested in 20 ml of ice-cold lysis buffer (20 mM Tris, pH 8.0, 150 mM NaCl, 0.2 mM EDTA, 1 mM dithiothreitol, 0.5% Triton X-100, protease inhibitors) and broken with a French press (1,500 p.s.i.). After centrifugation (15,000 × g, 20 min at 4 °C) the supernatant was mixed with glutathione-Sepharose 4B beads (Amersham Biosciences) and incubated for 15 min at 4 °C. The mixture was transferred to disposable columns and washed three times with 10 ml of ice-cold lysis buffer. Finally, the buffer was exchanged for yeast extraction buffer without NaCl.

Growing of Yeast Cells and Preparation of Yeast Protein Extracts—SR protein constructs in pYX242 were transformed into the yeast strain FY7C and plated on yeast-selective media lacking leucine. A few colonies of each construct were inoculated into 25 ml of SD media lacking leucine, and the culture was grown for 2 days at 30 °C. The cells were harvested by centrifugation (5 min, 2,000 rpm) and washed once with ice-cold water. Three cell volumes of extraction buffer (50 mM Tris-HCl, pH 8.0, 300 mM NaCl, 5 mM EDTA, 1 mM dithiothreitol, 1% Triton X-100, protease inhibitors) and 3 cell volumes of glass beads were added to the pellet, and lysis was performed by vortexing four times. The supernatant was transferred to a new tube and five volumes of extraction buffer were added to the glass beads, and the mixture was vortexed once more. The supernatants were mixed together and centrifuged for 20 min at 4 °C 15,000 × g. The supernatant from one protein extraction (1 ml) was used for three pull-down assays with the GST fusion proteins bound to glutathione-Sepharose 4B.

Pull-down Assay—Glutathione-Sepharose 4B beads were coated with the GST fusion proteins as described above. Two-hundred fifty microliters of extraction buffer without NaCl was added to 50 μl (15-μl bead volume) of the Sepharose beads coated with GST fusion proteins and mixed with 300 μl of yeast extract. The beads were incubated on a rotary shaker for 2 h at 4 °C and then washed three times with 1 ml of ice-cold extraction buffer, containing 150 mM NaCl. Bound proteins were denatured in 50 μl of Laemmli buffer, and 20 μl was separated on a 12% SDS-polyacrylamide gel. The proteins were blotted onto a polyvinylidene difluoride membrane (Millipore Inc.). Western blotting was performed according to standard procedures using mouse anti-HA antibodies (Dako, The Netherlands) as a substrate and were performed on intact seedlings or excised organs of mature plants as described in a previous study (42). The samples were sequentially treated with 70% ethanol for 2–6 h to remove chlorophyll from tissues. GUS staining was performed with plants of T2 or further generation.

RESULTS

Sequences, Domain Structure, and Expression of atRSZ33—To isolate homologs to mammalian SR proteins, the RRM of human SRp55 was used for a data base search of Arabidopsis EST libraries. Two EST clones were identified encoding proteins with significant similarity to SRp55 around the RNP1 and RNP2 submotifs of their RRM’s. Sequencing of the entire ESTs revealed that they encoded the same protein of 290 amino acids (Fig. 1A). However, comparison with mammalian SR proteins showed that the sequence similarity was restricted to the RRM domain. The N terminus of the encoded protein consists of an RRM followed by a basic RS-domain (20 Arg and 4 Lys) with multiple RS repeats and two CCHC-type zinc knuckles. (Fig. 1, A and B). In contrast, the C terminus (amino acid residues 198–290) is highly acidic (14 Asp, 5 Glu) and contains 10 SP-dipeptides (Fig. 1, A and B), SP-domain. Due to this domain structure and protein size, it was termed atRSZ33 (Arabidopsis thaliana RS-containing zinc knuckle protein with molecular mass of 33 kDa). A recent search of the Arabidopsis genome data base detected a second closely related protein (accession number At3g55500) with 80% identity to atRSZ33, which was named atRSZ32 (Fig. 1D). Although these two proteins have some structural similarity to the human 9G8 splicing factor and to members of the recently described RSZ family of Arabidopsis proteins, containing one zinc knuckle motif (36), they seem to be unique plant proteins with no obvious homologs in metazoans.

The gene encoding atRSZ33 is situated on the second chromosome and has six introns (Fig. 1C). The second intron (626 bp) was found to be alternatively spliced during the early stages of development. As a result of the usage of a 3’ alternative splice site, part of the second intron is included in the sequence of mRNA2 and a new stop codon is generated. Because of the new stop codon, a truncated protein can be deduced from mRNA2, which contains part of the RRM (RNP2) and some novel amino acids encoded by the included intron sequence (Fig. 1C). mRNA2 can be detected by RT-PCR in seedlings immediately after germination (data not shown), and cDNA clones with this alternative splicing pattern were found in the Arabidopsis EST data base. The distribution of the atRSZ33 mRNA was determined by Northern blot analysis of poly(A)+ mRNA isolated from various tissues of wild-type Arabidopsis plants (Fig. 2A). atRSZ33 RNA was highly abundant in roots and flowers but almost absent in leaves and stems (see also below). The alternatively spliced mRNA2 could not be detected in this assay, indicating that it is a minor isoform only detectable by RT-PCR.

Nuclear Localization of atRSZ33—The rather unique domain structure of atRSZ33 raised the question of which part of the RNA metabolism might be influenced by this protein. It was therefore important to determine the intracellular distribution of this protein. Consequently, atRSZ33 was transiently expressed in Nicotiana tabacum protoplasts as a fusion with GFP. Fluorescence microscopy analysis showed that atRSZ33 is localized in the nucleus, in contrast to GFP alone, which is present in both the cytoplasm and the nucleus (Fig. 2B). Interestingly, the nuclear localization of atRSZ33 showed a speckled pattern that is characteristic for SR proteins and other splicing
factors, indicating that it may be involved in splicing.

AtRSZ33 Is a Phosphoprotein—Polyclonal antibodies were raised against purified truncated recombinant protein, containing the RRM and both zinc knuckles. The specificity of the antibody was tested by immunoblot analysis of an SR protein preparation from aerial tissues of 3-week-old flowering Arabidopsis plants (Fig. 3A). A single band with a molecular mass of about 36 kDa was detected. However, an SR protein preparation from rapidly growing Arabidopsis cell suspension culture resulted in two close migrating bands with molecular masses of about 36 and 38 kDa (Fig. 3B, lane 1). To investigate whether the lower band represented a dephosphorylated form of atRSZ33 or its close homolog atRSZ32, the protein preparation was treated with alkaline phosphatase. This treatment resulted in the disappearance of the upper band, suggesting that it is the phosphorylated form of atRSZ33 (Fig. 3B, lanes 2–4).

AtRSZ33 Interacts with Splicing Factors in a Yeast Two-hybrid Assay and In Vitro—atRSZ33 has many structural features of an SR protein, but its domain composition is unique and not found in any SR protein characterized to date. We were interested in determining the function of this protein by identifying possible interacting partners using a yeast two-hybrid system. We encountered problems expressing the full-length protein, so we expressed truncated versions of the protein. We found that atRSZ33 interacted with splicing factors in a yeast two-hybrid assay and in vitro.
protein in both yeast and bacteria. Therefore, the truncated protein lacking the SP domain was used as the bait (pBD-RS33ΔSP). Eleven His" and LacZ" clones were obtained from a total of 1.2 × 10⁶ transformants. Sequence analysis of the positive clones established three groups of sequences. The first group consisted of one and two clones of the previously described splicing factors atRSZp21 and atRSZp22, respectively (36, 37). The second group contained two clones of atSRp34/ SR1, an ortholog of the human SF2/ASF splicing factor (34, 35). The third group consisted of three clones of the recently identified protein SR33, which interacts with the U1 snRNP 70K protein (47), and one and two clones encoding proteins of 28 and 30 kDa, respectively, showing sequence similarity to the human SC35 splicing factor. Therefore, these proteins were designated as atSCL28 and atSCL30 (Arabidopsis thaliana SC35-like protein with molecular masses of 28 and 30 kDa, respectively). Sequence analysis revealed that SR33 is highly related to atSCL30 and atSCL28 and was therefore renamed to atSCL33/SR33 to indicate this fact. Overall, six unique proteins were isolated, and the reconstitution of all interactions is described splicing factors atRSZp21 and atRSZp22 in a yeast two-hybrid assay, we decided to determine whether the expression patterns of these splicing factors are overlapping. Promoter fragments of atRSZ33 (949 bp), atSRp34/SR1 (1095 bp) (35), and atRSZp22 (2119 bp) were fused to the reporter gene GUS and used to transform A. thaliana plants. Expression patterns during seedling development and in different organs and tissues of adult plants were analyzed from representative numbers of transgenic lines.

Comparison of the expression patterns of the genes tested showed a distinct staining pattern for each gene, with overlapping expression in various tissues and at certain stages of plant development. The widest range of overlapping expression was observed during flower development. In anther and pollen development, atRSZ33 colocalizes with atSRp34/SR1 and atRSZp22 within the anther in tapetal tissue (Fig. 5, A, E, and I). At later stages, transcription activity of all genes becomes confined to the mature pollen (Fig. 5, B, F, and J). In gynoecium before pollen release, atRSZ33 is co-expressed with atRSZp22 within the ovules, septum, stigmatic tissue, and style (Fig. 5, B and J). atSRp34/SR1 does not produce any staining within the gynoecium before fertilization (Fig. 5I). However, after fertilization, atSRp34/SR1 together with atRSZ33 and atRSZp22 is expressed in developing seeds (Fig. 5, C, G, and K).

Analysis of other plant organs revealed more diverse expression patterns, with the major overlapping parts in roots. In general, all genes are transcriptionally active in roots. Although atSRp34/SR1 is expressed both in the root meristem and in the elongation zone (Fig. 5H), expression of atRSZ33 and atRSZp22 is restricted to the elongation zone of the root (Fig. 5, D and L). The expression pattern of atRSZ33 was verified by in situ hybridization experiment.²

Analysis of other plant organs revealed more diverse expression patterns, with the major overlapping parts in roots. In general, all genes are transcriptionally active in roots. Although atSRp34/SR1 is expressed both in the root meristem and in the elongation zone (Fig. 5H), expression of atRSZ33 and atRSZp22 is restricted to the elongation zone of the root (Fig. 5, D and L). The expression pattern of atRSZ33 was verified by in situ hybridization experiment.²

The data presented show that atRSZ33, atSRp34/SR1, and atRSZp22 have overlapping patterns of expression, which is consistent with their interaction in a yeast two-hybrid assay and in vitro. However, the expression of these genes also has characteristic features, suggesting distinct protein complex compositions for their respective activities in different tissues.

Taken together, the co-precipitation data with full-length atRSZ33 and the co-expression data of atRSZ33 with atRSZp22 and atSRp34/SR1 in various cell types are consistent with the in vivo protein-protein interaction data. Therefore, the evi-

² M. Kalyna, S. Lopato, and A. Barta, manuscript in preparation.
dence presented in this report suggests a role for atRSZ33 in pre-mRNA splicing.

**Novel Family of SC35-like Splicing Factors in Arabidopsis**—Three SCL proteins (atSCL28, atSCL30, and atSCL33/SR33) have been isolated in a yeast two-hybrid screen using atRSZ33 as the bait. One additional member of this protein family, atSCL30a, was found in the Arabidopsis EST database. Using atSCL30a as the bait in a yeast two-hybrid screen, we identified a fifth member of the SCL family, atSC35, which seems to be an ortholog of the human SC35 splicing factor (see below). As shown in Fig. 4A, atSC35 also interacts with atRSZ33 in the yeast two-hybrid assay.

Fig. 6A shows an alignment of protein sequences of all members of the plant SCL family and the human and alfalfa SC35 splicing factors. They all share a striking sequence similarity in the single RRM (amino acids 15–72 of hSC35), but their C-terminal RS domains are of different length and composition. The N-terminal region upstream of the RRM is very short in plant and animal SC35 protein but is considerably longer in the plant SCL proteins. This region can be considered as a second short RS domain, because it is rich in the amino acids characteristic of this domain, namely, arginines, prolines, serines, glycines, and tyrosines (Fig. 6A). atSCL30 has a stretch of six prolines followed by a stretch of eight glycines, which suggests that this domain is highly flexible. From the alignment and the dendrogram of Arabidopsis and human SR proteins (Fig. 6, A and B), it is apparent that the four atSCL proteins are more related to each other than to the Arabidopsis and mammalian SC35. Moreover, atSCL33/SR33 and atSCL30a represent a...
pair of close homologs with nearly identical N-terminal RS domains and RRMs together with a high level of identity in the C terminus. With the exception of the genuine Arabidopsis SC35 ortholog, none of the SCL proteins has been identified in mammals, which strengthens the possibility that these proteins are plant-specific splicing factors.

**Mapping the Segments of atRSZ33 Involved in Interaction with atSCL33/SR33**

It is known that the RS-domain of vertebrate SR proteins is important for protein-protein interactions (9, 17, 48, 49), but the regions involved in these interactions are not precisely known. To determine the sequence requirement for atRSZ33, which is necessary and sufficient to interact with atSCL33/SR33, we assayed binding in vivo using the yeast two-hybrid assay. Different segments of atRSZ33 fused to the binding domain of Gal4 were co-expressed with a chimeric protein containing the open reading frame of atSCL33/SR33 fused to the transcriptional activation domain of GAL4. The relative strength of the protein-protein interaction was determined by testing the ability of the yeast to grow on His(-)-selective medium and by the β-galactosidase activation assay (Fig. 7). Surprisingly, a fragment containing the RRM, the two zinc knuckles, and the small RS region between them was sufficient to produce strong binding. In contrast, a fragment containing the zinc knuckles together with the rest of the RS domain showed no β-galactosidase activity but yeast cells were able to grow on His(-) plates, indicating weak protein-protein interaction. This suggests that a part of the RRM region is necessary for efficient binding. However, the RS and

---

**Fig. 6. A**: Alignment of SC35 protein sequences from alfalfa (ms), human (hs), and Arabidopsis (at) with the SC35-like proteins (SCL-proteins) from Arabidopsis. Alignment was generated using the ClustalW and shaded using the BOXSHADE program (www.ch.embnet.org/software/BOX_form.html). Amino acids identical or similar in more than 50% of the analyzed sequences are shown by black or gray background, respectively. The conserved RNP1 and RNP2 submotifs are overlined. **B**: Dendrogram of 18 Arabidopsis and 9 human SR proteins. Dendrogram and the corresponding multiple sequence alignment were generated using the PileUP program.
We have noted that many of the homologs of atRSZ33 in the unique and not yet found in metazoans. We have identified a close homolog from Arabidopsis (53, 54). Many features of atRSZ33 suggest its involvement in splicing—arRSZ33 has many characteristics of an SR protein: (i) it is a phosphoprotein and quantitatively precipitated by high concentrations of magnesium ions, (ii) it is present in the same ammonium sulfate fractions as the other SR proteins, (iii) it is a nuclear protein concentrated in speckles, and (iv) it interacts with other SR proteins.

SP domain without the zinc knuckles, or the two zinc knuckles only, did not interact with atSCL33/SR33. Thus, a small part of the RS domain together with both zinc knuckles and part of the RRM of atRSZ33 interacts with atSCL33/SR33.

Taken together, we have shown that the region of interaction included RNA binding domains, and that part of these domains are necessary for specific binding. It is conceivable that both RNA binding domains indirectly contribute to the interaction by assisting the RS domain to adopt the conformation required for binding. However, the specific protein interaction region is small enough to allow other proteins to interact with the rest of the RS domain, a feature important for complex formation.

DISCUSSION

The correct recognition of splice sites and the modulation of alternative splicing in higher eukaryotes is brought about mainly by a group of RS-rich proteins. RS proteins, which include the SR protein family, have multiple functions in pre-mRNA processing both in constitutive splicing as well as in the regulation of alternative splicing (reviewed in Refs. 6, 8, and 50). It was proposed that they act by binding to pre-mRNA and recruiting other splicing factors to the spliceosome. In addition, SR proteins are thought to mediate interactions between splicing factors already bound near the 5’ and 3’ splice sites. SR proteins are nuclear proteins, although some of them shuttle between the nucleus and the cytoplasm (51). Their subnuclear localization is in speckles (which may function as storage sites for certain splicing factors), from which they are thought to migrate to active transcription sites (20, 52). The localization of SR proteins as well as their interaction abilities can be strongly influenced by their phosphorylation status (23, 53, 54).

Here we describe the characterization of a novel RS protein from Arabidopsis, atRSZ33, which has some sequence similarity to mammalian SR proteins, but its domain organization is unique and not yet found in metazoans. We have identified a close homolog of atRSZ33 in the Arabidopsis database and named it atRSZ32, but we have no experimental data about this protein. We have noted that many of the Arabidopsis splicing factors have close homologs, a situation different to that in animals (55). This has already been noted for other large protein groups in Arabidopsis; indeed, about 44% of such cases result from large intra- and interchromosomal duplications (56, 57).

Many Features of atRSZ33 Suggest Its Involvement in Splicing—atRSZ33 has many characteristics of an SR protein: (i) it is a phosphoprotein and quantitatively precipitated by high concentrations of magnesium ions, (ii) it is present in the same ammonium sulfate fractions as the other SR proteins, (iii) it is a nuclear protein concentrated in speckles, and (iv) it interacts with other SR proteins.

The biochemical analysis of plant splicing factors has been hampered by the lack of a plant in vitro splicing extract, which would have allowed depletion and complementation assays. Therefore, to gain insight into the functions of atRSZ33, we used a yeast two-hybrid screen to identify possible interaction partners. By doing so we have isolated six different proteins, atSRp34/SR1, atRSZp21, atRSZp22, atSCL28, atSCL30, and atSCL33/SR33, that all were either known or putative splicing factors. Two of the identified proteins, atSCL28 and atSCL30, have not been described previously. The majority of these interactions have been verified by in vitro pull-down assay data. However, because our bait in the two-hybrid screen did not contain the SP domain we might have missed proteins that bind to this region. Indeed, C. Koncz and coworkers3 have found atRSZ33 complexed with AKN10 and AKN11 kinases, an interaction which might involve the SP region. It is interesting to note that AKN10 and AKN11 are negatively regulated by PRL1, a protein known to be associated with the spliceosome (58).

One of the proteins found in screening with atRSZ33 was atSRp34/SR1, one of four SP2/ASF-like splicing factors from Arabidopsis (55). Human SF2/ASF has been shown to be an essential protein important for the recognition of the 5’ splice sites by interacting with the U1–70K protein and for modulating alternative splice site selection by binding to enhancer sequences (reviewed in Refs. 4 and 50). In Arabidopsis, analysis of two of the SF2/ASF-like splicing factors has revealed that atSRp34/SR1 acts as a general splicing factor, whereas the related atSRp30 functions as a specific splicing modulator (34, 35).

The SCL protein family, which is highly homologous to the well-characterized human splicing factor SC35, was also isolated by this screen, and all five members of the family interacted with atRSZ33 in a yeast two-hybrid assay. One of these five proteins, atSCL33/SR33, has recently been shown to be associated with the Arabidopsis U1–70K protein (47). Mammalian U1 snRNP protein has previously been shown to interact with splicing factors SC35 and SF2/ASF (9, 17). In the screen

3 C. Koncz, personal communication.
using atRSZ33 as the bait, we have isolated several clones of atSCL33/SR33, demonstrating the efficient binding between the two proteins. The importance of human SC35 in spliceosome assembly is discussed in the following paragraph.

The third group of interacting proteins consisted of two small proteins containing one zinc knuckle motif, atRSZp21, and atRSZp22. atRSZp22 has been shown to be a splicing factor, because it is able to efficiently complement splicing-deficient mammalian extracts, and because it possesses an RNA binding specificity similar to the human splicing factors 9G8 and SRp20 (36). Both atRSZp21 and atRSZp22 have been shown to interact with the U1–70K protein in the yeast two-hybrid assay (37).

Taken together, all the interacting partners of atRSZ33 listed above are either known or putative splicing factors. Fig. 8 summarizes the interactions of plant splicing factors as reported in this and previous papers (37, 47). Although these authors have not found interactions of atSCL33/SR33 with atRSZp21 or atRSZp22, we do observe the interaction with atRSZp21 but not with atRSZp22.4 Because one of the functions of SF2/ASF and SC35 proteins is recognition of the 5′ splice site, and both the atRSZ and atSCL proteins bind to a U1 snRNP protein, we conclude that at least one of the activities of atRSZ33 is spliceosome assembly, most probably near the 5′ splice site (Fig. 8). Association of atRSZ33 with atSRp34/SR1 also indicates that this protein may participate in splicing enhancer complexes, an activity amply demonstrated for SR proteins (50).

Northern blot analysis of mRNA isolated from various parts of the plant showed that atRSZ33 is mainly present in the roots and flowers, suggesting that it might be preferentially expressed in dividing cells. This is consistent with our observation that more protein is present in rapidly growing Arabidopsis suspension cells than in aerial tissues of adult Arabidopsis plants. Furthermore, the expression data from the promoter-GUS construct showed that atRSZ33 was expressed in elongating cells, e.g. in the roots and anthers, which suggests that the absence of the zinc knuckle domain and a part of the RRM. However, it remains to be determined whether this property holds true in other binding assays. Because the protein interaction sites are close to the RNA binding region, it is conceivable that specific protein-protein interactions could in turn modulate RNA binding specificity or the strength of interaction. Therefore, SR protein-protein interactions can modulate specificity of RNA interaction and/or increase the strength of binding. This may explain our failure to map specific RNA sequences interacting with several of our SR proteins in a yeast two-hybrid screen, two SCL proteins selected three unknown proteins containing CCHC-type zinc knuckle domains but with no obvious RNA sequence. This indicates that SCL proteins possess the ability to recognize sequences around zinc knuckles.

It is interesting that a part of the RRM region is necessary for efficient protein binding in the yeast two-hybrid assay. However, it remains to be determined whether this property holds true in other binding assays. Because the interaction domains are sites close to the RNA binding region, it is conceivable that specific protein-protein interactions could in turn modulate RNA binding specificity or the strength of interaction. Therefore, SR protein-protein interactions can modulate specificity of RNA interaction and/or increase the strength of binding. This may explain our failure to map specific RNA sequences interacting with several of our SR proteins in a yeast three-hybrid system, because only one protein was expressed in this assay. Furthermore, an interesting observation is the fact that the interaction domain is relatively small, which provides enough space for additional interactions with either other RS or non-RS proteins. This is consistent with the proposed function of RS proteins as bridging proteins in both the assembly of spliceosomes and interaction with the transcription apparatus.

Acknowledgments—We are thankful to Dr. Helen Healy for critical reading of the manuscript. The EST clones were kindly provided by the Arabidopsis Biological Resource Center (Ohio State University, Columbus, OH).

REFERENCES

1. Reed, R., and Paldandjian, L. (1997) in Spliceosome Assembly: Eukaryotic mRNA Processing (Krainer, A. R., ed) pp. 103–129, IRL Press, Oxford
2. Kramer, A. (1996) Annu. Rev. Biochem. 65, 367–409
3. Staley, J. P., and Guthrie, C. (1998) Cell 92, 315–326
4. Fu, X.-D. (1995) RNA (N. Y.) 1, 663–680
5. Manley, J. L., and Tacke, R. (1996) Genes Dev. 10, 1569–1579
6. Blencowe, B. J., Bowman, J. A., McCracken, S., and Rounsma, E. (1999) Biochem. Cell Biol. 77, 277–291
7. Valcarcel, J., and Green, M. R. (1996) Trends Biochem. Sci. 21, 296–301
8. Graveley, B. R. (2000) RNA (N. Y.) 6, 1197–1211
9. Wu, J. Y., and Maniatis, T. (1993) Cell 75, 1061–1070
10. Zuo, P., and Manley, J. L. (1994) Proc. Natl. Acad. Sci. U. S. A. 91, 3363–3367
11. Guth, S., Martinez, C., Gaur, R. K., and Valcarcel, J. (1999) Mol. Cell. Biol. 19, 8263–8271
12. Merendino, L., Guth, S., Bilbao, D., Martinez, C., and Valcarcel, J. (1999) Nature 402, 838–841
13. Wu, S., Romfo, C. M., Nilsen, T. W., and Green, M. R. (1999) Nature 402, 838–841
